Remote reference magnetotelluric impedance estimation of wideband data using hybrid algorithm

Shalivahan,[1] Rajib K. Sinharay,[2] and Bimalendu B. Bhattacharya[3]

Received 6 February 2006; revised 16 June 2006; accepted 15 August 2006; published 11 November 2006.

[1] Precise and accurate determination of the magnetotelluric (MT) impedance is fundamental to valid interpretation. This paper deals with the results of hybrid processing schemes in conjunction with the remote reference (RR), for a typical site out of several sites collected over the wide band data ($\sim 10^3$ to 10^{-3} Hz). The standard practice of MT impedance estimation is the use of robust processing in conjunction with the remote referencing. The estimation by robust technique helps in reducing the effect of outliers in the electric field but is often not sensitive to the exceptional predictor (magnetic field) data, which are called leverage points. The data processed with robust M estimation (RME) exacerbated the bias problem in the dead band. The application of hybrid (coherence weighted estimation (CWE) + RME, rho-variance weighting + RME) and extra hybrid (CWE + rho variance + RME) approach helps in reducing the influence of both the outliers in the electric field and the leverage points. These two approaches perform considerably better than either data weighting scheme by itself.

Citation: Shalivahan, R. K. Sinharay, and B. B. Bhattacharya (2006), Remote reference magnetotelluric impedance estimation of wideband data using hybrid algorithm, *J. Geophys. Res.*, *111*, B11103, doi:10.1029/2006JB004330.

1. Introduction

[2] The interpretation of magnetotelluric (MT) data requires estimation of tensor impedance elements from the electric (**E**) and magnetic (**H**) vector field measurements assuming them to be a plane wave. There exists a linear statistical model in the frequency domain which relates the $N \times 1$ vector **E** of observations of the response (also called as dependent) variable to the NX2, rank 2 matrix **H** of predictors (explanatory variables), **Z** is the solution 2 vector, and assuming a homogenous plane wave, as

$$\mathbf{E} = \mathbf{Z}\mathbf{H} + \mathbf{M} \tag{1}$$

where **M** is an $N \times 1$ vector of random errors, and N is the number of observations (i.e., N Fourier transform of N independent data sections at a given frequency). The least squares (LS) solution for model (1) is

$$\hat{\mathbf{Z}} = \left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\left(\mathbf{H}^{\mathrm{H}}\mathbf{E}\right) \tag{2}$$

where the superscript H denotes the Hermitian (complex conjugate) transpose. The terms within parenthesis are the average estimates of autopower and cross-power spectra based on available data. The predicted values of the response from the regression are derived from the observed values by

$$\hat{\mathbf{E}} = \mathbf{P}\mathbf{E} \tag{3}$$

Where the $N \times N$ matrix **P** is called the prediction or hat matrix as it transforms the observed vector **E** into its LS using equation (3). It is given by

$$\mathbf{P} = \mathbf{H}\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\mathbf{H}^{\mathrm{H}} \tag{4}$$

or

$$\hat{\mathbf{E}} = \mathbf{H}\hat{\mathbf{Z}} \tag{5}$$

The predictor matrix **P** is idempotent (**PP** = **P**) and symmetric ($\mathbf{P}^{\mathrm{H}} = \mathbf{P}$). The regression residuals **r** is the differences between the measured values of the response variables **E** and predicted **E** (equation (5)). These serves as an estimate on the random errors. The classical Gauss-Markov theorem gives following condition for the linear regression to yield the best linear unbiased estimate:

[3] 1. The error term, on average, has no effect on the dependent variable.

[4] 2. Explanatory variable(s) are determined independently of the values of the error term (and, therefore, the dependent variable).

[5] 3. The error term has a constant variance; the observations of the error term are assumed to be drawn continually from identical distributions.

[6] 4. The observations of the error term are uncorrelated with each other.

[7] The complex tensor impedance **Z** can be estimated by Fourier transformation of time series and using the LS

[1]Department of Applied Geophysics, Indian School of Mines, Dhanbad, India.
[2]Central Water and Power Research Station, Khadakwasla, Pune, India.
[3]S. N. Bose National Centre for Basic Sciences, Salt Lake, Kolkata, India.

0148-0227/06/2006JB004330$09.00

method to obtain the possible fit to equation (1) [*Sims et al.*, 1971]. However, for noisy data, this approach fails, see *Egbert and Livelybrooks* [1996]:

[8] Equation (1) is appropriate to the case where noise is restricted to the output or the "predicted" electric field channels, while the input magnetic field is observed without error. Thus the violation of this assumption would result in LS estimated impedance to be biased downward [*Sims et al.*, 1971].

[9] It is, usually, necessary to estimate the response function **Z** from data corresponding to large residuals, which are usually called outliers. The common types of outliers are: point defects and local nonstationarity. Point defects are isolated outliers, which are independent of the assumed linear statistical model (equation (1)). Some examples of point defects are transient instrumental errors and spike noise due to natural phenomena (e.g., nearby lightning). Local nonstationarities in geophysical problems are seen in observations of the time-varying fields (E_x, E_y, H_x, and H_y): most of the time the data statistics are approximately constant i.e., coming close in satisfying equation (1), but this stationary process is interrupted sporadically by brief but intense disturbances such as magnetic storms with markedly different characteristics [*Chave et al.*, 1987], and these are non-Gaussian in nature. Even severe bias could be produced by auroral substorm source fields of a short spatial scale [*Garcia et al.*, 1997]. Some types of midlatitude sources have short and temporally variable spatial scales, which can also alter MT responses [*Egbert et al.*, 2000]. Because of marked nonstationarity, departures from the linear statistical model that produce very large residuals in the data are more likely than "normal" or "expected" and such residuals are heavy-tailed or very long tailed with a Gaussian center. Therefore the conventional LS approach is beset with problems.

[10] Thus there are basically two types of statistical errors inherent in the estimation of response function: random and bias errors due to **E** and/or **H** field noise. The statistical errors are due to all errors normal or nonnormal. The variance gives a quantitative measure of the precision of an estimate. A precise estimate may still be inaccurate because of bias error due to **E** and/or **H** field noise.

[11] To tackle the bias errors due to noise, remote reference (RR) measurements of MT was suggested about three decades back [*Goubau et al.*, 1978a, 1978b; *Gamble et al.*, 1979a, 1979b; *Goubau et al.*, 1984]. The equation of estimating RR MT impedance in terms of power spectral densities for magnetic field reference is given by [*Vozoff*, 1996]

$$Z_{xy} = \frac{\left\langle E_x R_y^* \right\rangle \left\langle H_x R_x^* \right\rangle - \left\langle E_x R_x^* \right\rangle \left\langle H_x R_y^* \right\rangle}{\left\langle H_y R_y^* \right\rangle \left\langle H_x R_y^* \right\rangle - \left\langle H_y R_x^* \right\rangle \left\langle H_y R_x^* \right\rangle} \quad (6)$$

where R_x and R_y are the remote magnetic field components. The field components in asterisk indicate complex conjugate. *Schultz et al.* [1993] and *Garcia et al.* [1997] have shown that the RR can give severe bias in the auroral substorm source fields of a short spatial scale. MT estimates even get distorted from DC electric rail system [*Junge*, 1996].

[12] A number of MT processing methods have been proposed on the basis of some sort of coherence weighted estimates (CWE) [*Stodt*, 1983; *Jones and Jodicke*, 1984] and apparent resistivity variance (rho-var) weighting estimates [*Stodt*, 1983, also J. A. Stodt, Computation of MT parameters and their error estimates, unpublished report submitted to Phoenix Geophysics Ltd., 1980, hereinafter referred to as J. A. Stodt, unpublished report, 1980]. Subsequently, the robust M estimates (RME) [*Huber*, 1981; *Rousseeuw and Leroy*, 1987] have been applied to estimate the impedance functions [*Egbert and Booker*, 1986; *Chave et al.*, 1987; *Chave and Thompson*, 1989; *Jones et al.*, 1989; *Larsen*, 1989; *Sutarno and Vozoff*, 1989, 1991; *Larsen et al.*, 1996; *Egbert and Livelybrooks*, 1996; *Egbert*, 1997; *Shalivahan and Bhattacharya*, 2002; *Smirnov*, 2003]. *Banks* [1998] has studied the effect of nonstationary noise on electromagnetic response estimates in the frequency range of 0.05–0.000167 Hz. RME processing has long been used for very low frequency (<0.1 Hz) MT and magnetovariational data due to its easy availability. *Jones et al.* [1989] have demonstrated the superiority of the RME processing for low-frequency MT data and further suggested that RR measurements, wherever possible, should be made.

[13] The RME methods tackle the noise on the electric field but are often not sensitive to exceptional predictor (magnetic field) data called as leverage point [*Chave and Thompson*, 2004]. This happens particularly in the dead band 5 Hz to 0.05 Hz and the MT impedance estimates by RME processing are severely biased [*Egbert and Livelybrooks*, 1996]. This happens because RME is sensitive to data, which produce statistically unusual residuals. Such data may or may not correspond to all of the influential data in a data set. Some suggested methodologies for tackling the outliers and leverage points have been suggested by *Egbert and Livelybrooks* [1996], *Shalivahan* [2000], *Garcia and Jones* [2002], *Jones and Spratt* [2002], and *Chave and Thompson* [2004]. *Egbert and Livelybrooks* [1996] have used for the first time, a hybrid approach, i.e., combination of CWE with RME only for single site MT impedance estimation over a wide band to tackle the problems associated with the RME particularly in the dead band (5–0.05 Hz). *Shalivahan* [2000] has used two hybrid approaches, i.e., combination of CWE with RME as well as rho-var weighting with RME in conjunction with RR for improving the data quality for the entire frequency range including the dead band. *Garcia and Jones* [2002] presorted AMT data based on the power in the magnetic field channels, selecting only those values, which exceed the known instrument noise level by a specified amount. *Jones and Spratt* [2002] preselected data segments whose vertical field power was below a threshold value to minimize auroral source field bias in high-latitude MT data.

[14] In this paper we compare different processing schemes, i.e., CWE, rho-var weighting, RME, hybrid approaches, i.e., the combinations of CWE with RME as well as rho-var weighting with RME and extra hybrid approach (CWE+rho-var weighting+RME), in the estimation of RR MT impedance over a wide band ($\sim 10^3$ Hz to 10^{-3} Hz) for a representative site out of 44 sites of Dhanbad-Badampahar transect (Figure 1).

2. Data Acquisition

[15] A brief account of MT data acquisition and processing system of V5-16 multiple geophysical receiver of Phoenix

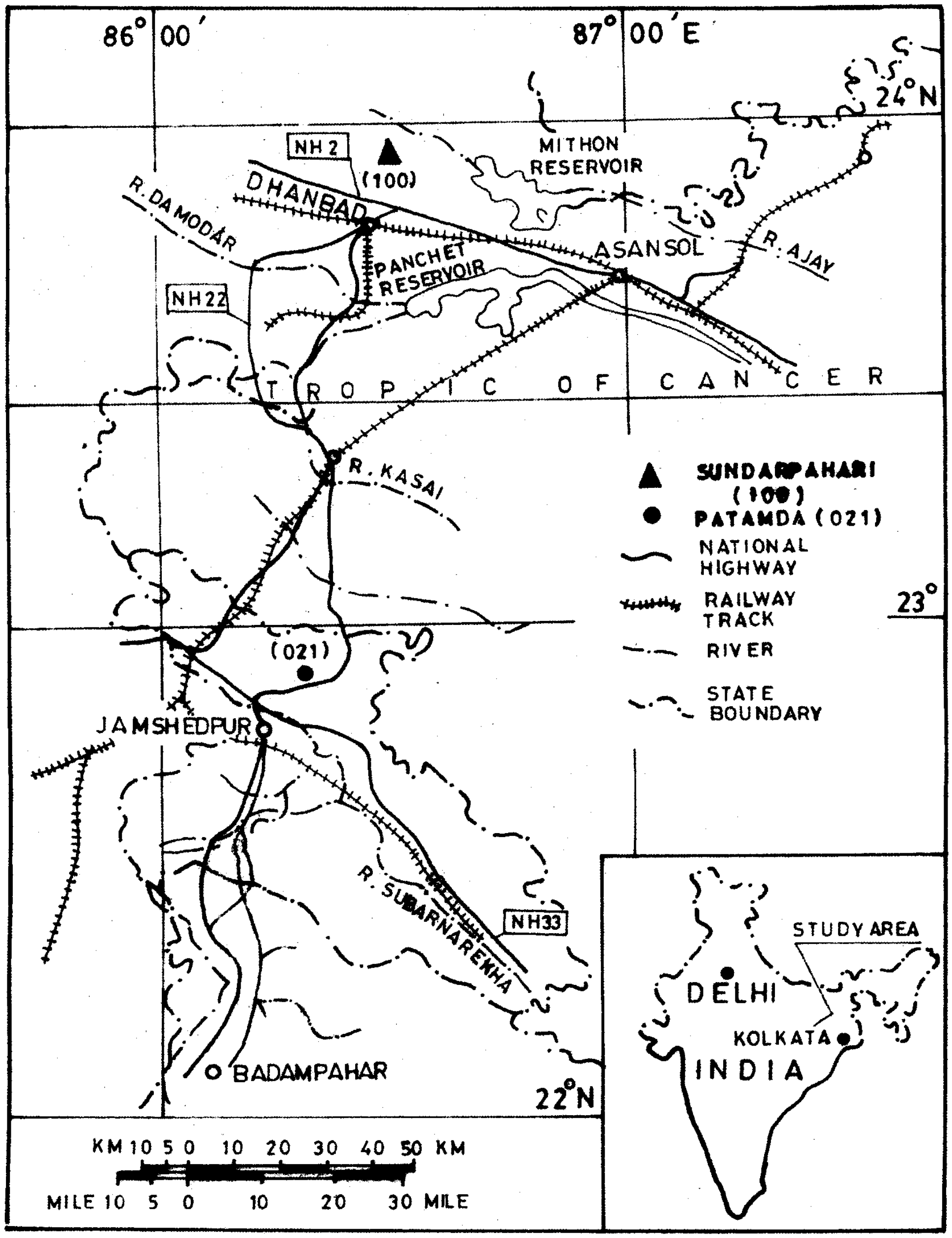

Figure 1. Location map.

Geophysics Limited, Canada used in fieldwork is described here. At each site 2 tellurics and 3 magnetics sensors were hooked up to signal processor unit (SPV5), which was connected to a receiver (V5) through a communication cable and placed at a distance of about 150 m from SPV5. For RR MT measurements two sites had the same configuration. Both the sites were synchronized to the GPS clock. Time series data was collected at the SPV5 box. V5 system acquires as well as process the data. The time series data is Fourier transformed and V5 acquires data in two frequency ranges: (1) high range, 12 frequencies ranging from 320 Hz to 7.5 Hz and (2) low range, 28 frequencies ranging from 6.0 Hz to 0.00055 Hz. Each station along with its RR site was occupied for 18 hours. Out of this, 3 hours were for high range (320 Hz to 7.5 Hz) and 15 hours for low range (6.0 Hz to 0.00055 Hz). The dipole lengths varied from 100 m to 140 m.

3. Data Analysis

[16] The source for higher frequencies is broadband radiation from lightning strikes that propagates between the earth and the ionosphere to the measuring site where as the source for low frequency is the micropulsation. There exists a dead band in a part of the spectrum where the signal levels are very low and lies between the frequency ranges of two types of sources, i.e., thunderstorm activities and micro pulsations. In this band the noise may sometimes be stronger than the signal (noise/signal >1.0). The data are often mostly noise with occasional bursts of real signal. The

data noise may also remain a problem for the low-frequency signals.

[17] Following MT processing methods has been used: (1) coherency weighted estimation (CWE), (2) apparent resistivity variance (rho-var) weighting, (3) robust M estimation (RME), (4) hybrid approach, and (5) extra hybrid approach to remove the statistical errors. The data processing begins by dividing the time series into a sequence of short data segments and each is Fourier-transformed. This when combined with frequency band averaging gives a series of L complex data vectors for a given frequency ω. The tensor impedance $\mathbf{Z}$ (ω) for that frequency is then estimated by minimizing weighted-residual sums of squares. The weighted deviation for the x component electric field is given as

$$\sum_{i=1}^{L} w_i \left| E_{xi} - \left(Z_{xx} H_{xi} + Z_{xy} H_{yi} \right) \right|^2 \tag{7}$$

The weights w_i are estimated from the data. The CWE, rho-var and RME choose weights to emphasize the best quality data.

3.1. Coherency Weighted Estimation

[18] In CWE the sequence of L Fourier coefficient vectors [*Egbert and Livelybrooks*, 1996] are divided into q temporary contiguous groups. Generally, the coefficients for group q typically correspond to all data collected in one of a series of "runs" at the measuring site. Then for each group the standard multiple coherence (γ_q^2) between the x component of electric field and the magnetic field is computed. This helps in determining the weights as a function of γ_q^2. The data with higher value of γ_q^2 is given higher weight and vice versa. For RR estimation of the tensor impedance, ordinary and multiple coherences between local and remote fields can also be computed [*Stodt*, 1983, also unpublished report, 1980] to assess correlations between sites and determine the best reference pair. For a RR the generalized multiple squared coherence between the observed electric field $\mathbf{E}$ and its predicted value $\hat{\mathbf{E}}$ with magnetic field ($\mathbf{R}$) as a reference and the local magnetic field as $\mathbf{H}$ is given as [*Chave and Thompson*, 2004]

$$\gamma_{E\hat{E}}^2 = \frac{\left| S_{E\mathbf{R}} (S_{\mathbf{RH}})^{-1} S_{\mathbf{R}E} \right|}{S_{EE} S_{\mathbf{R}E}^H (S_{\mathbf{RH}}^H)^{-1} S_{\mathbf{HH}} (S_{\mathbf{RH}}^H)^{-1} S_{\mathbf{R}E}} \tag{8}$$

where S_{xy} is a cross power between vector variables x and y. It is a complex quantity whose amplitude is analogous to the standard multiple coherence and its phase is a measure of the similarity of the local and remote reference variables. The minimum value of squared coherency (γ_q^2) is set and only those data segments meeting this minimum are used in the estimation of the impedance. The coherence weight in this instance consists of zeros and ones. J. A. Stodt (unpublished report, 1980) and *Egbert and Livelybrooks* [1996] have shown that CWE tends to increase the signal/noise ratio.

3.2. Apparent Resistivity Variance (rho-var) Weighting Estimation

[19] The expressions of variance are given by *Gamble et al.* [1979b]. The variance in each element of $\mathbf{Z}^{\mathrm{R}}$ (impedance estimation using RR) can be expressed in terms of known average powers, if it is assumed that the noise is independent of signals, and the noise is stationary. The variances decrease as the number of measurements contained in the average power increases. In this case this number has been fixed at 20 (J. A. Stodt, unpublished report, 1980). Here the variances of apparent resistivities (rho) in conjunction with the remote referencing computed from principle impedance elements have been used. These are first averaged to obtain a minimum variance estimate.

[20] Let us assume that the base field noises are uncorrelated with reference field noise so that the estimates $\mathbf{Z}^{\mathrm{R}}$ are unbiased by correlated noise powers. The aim is to obtain a weighted average of the four stable estimates, which has the property that it is the minimum variance unbiased estimate obtainable. This estimate is given by [*Gamble et al.*, 1979a, 1979b; J. A. Stodt, unpublished report, 1980]

$$\overline{Z}_{ij} = \sum_{k=1}^{M} W_k Z_{ijk} \tag{9}$$

where k indicates a sum over the weighted individual estimates and M is the four possible reference pairs, $E_x^R E_y^R$, $H_x^R H_y^R$, $E_x^R H_x^R$, and $E_y^R H_y^R$ used for the impedance estimates. The weights take the form

$$W_k = \frac{1/\mathrm{Var}\left(Z_{ijk}\right)}{\sum_{k=1}^{M} 1/\mathrm{Var} Z_{ijk}} \tag{10}$$

For the impedance estimate to be unbiased the sum of the weights should be equal to 1.0, i.e., $\sum_1^4 W_i = 1$. It is important to average the real and imaginary parts of the Z_{ij}^R, rather than their magnitudes and phase, in order to avoid introducing other bias error. If the individual Z_k are unbiased and if we have accurate estimates of the Var Z_k, then equation (10) will give an unbiased estimate with the minimum possible variance. If the Var Z_k are not estimated accurately, then too equation (10) gives an unbiased estimate but not with a minimum possible variance. The variance of the average estimate $\overline{Z}_{ij}^R$ is given as (J. A. Stodt, unpublished report, 1980)

$$\mathrm{Var}\overline{Z}_{ij}^R = \sum_{k=1}^{4} W_k^2 \mathrm{Var}\overline{Z}_{ij_k}^R + 2\sum_{k=1}^{3} \sum_{l=2}^{4} W_k W_l \mathrm{Cov}\left(Z_{ij_k}^R, Z_{ij_l}^R \right) \tag{11}$$

If the variance estimates are accurate, then $\mathrm{Var}\overline{Z}_{ij_k}^R$ will be smaller than any of the $\mathrm{Var}Z_{ij_k}$. *Gamble et al.* [1979b] defined Var Z_{ij} as

$$\mathrm{Var}Z_{ij}^R = \frac{|\overline{r}_i|^2 |\overline{A}_j|^2}{N|D|^2} \tag{12}$$

where

$$|D|^2 = \left| \overline{H_x R_x^*}\,\overline{H_y R_y^*} - \overline{H_x R_y^*}\,\overline{H_y R_x^*} \right|$$

$$\mathbf{r} = \mathbf{E} - \hat{\mathbf{E}}$$

$$A_x^* = \overline{R_x^* H_y R_y^* - R_y^* H_y R_x^*}$$

$$A_y^* = \overline{R_y^* H_x R_x^* - R_x^* H_x R_y^*}$$

$$|A_j|^2 = \overline{A_j A_j^*}$$

N is the number of independent determinations of each field. For large value of N, one can replace $|r_i|^2$ in equation (12) as

$$\overline{|r_i^P|}^2 = |\overline{E_i}|^2 - 2\,\mathrm{Re}\left[Z_{ix}^R\overline{H_xE_i^*} + Z_{iy}^R\overline{H_yE_i^*} - Z_{ix}^R Z_{iy}^{R*}\overline{H_xH_y^*}\right] + |Z_{ix}^R|^2|\overline{H_x}|^2|Z_{iy}^R|^2|\overline{H_y}|^2$$

where Re(x) is the real part of x. N is the number of averages in spectral estimates, $i = x, y$ and $j = x, y$.

[21] When the individual estimates Z_k are obtained from disjoint sets of spectra, then the variance of $\overline{Z_{ij}}$ is

$$\mathrm{Var}\overline{Z_{ij}} = \sum_{k=1}^{M} W_k^2 \mathrm{Var} Z_{ijk} \tag{13}$$

If equation (10) is substituted into equation (13), $\overline{Z_{ij}^R}$

$$\mathrm{Var}\overline{Z_{ij}^R} = \frac{1}{\left[\sum_k 1/\mathrm{Var}Z_{ijk}\right]}$$

and the variance of rho is $(0.2\mathrm{T})^2$ Var (Z_{ij}).

[22] The use of inverse variance as weights not only incorporates the signal criteria but also down weights the events for high coherence between the orthogonal components of magnetic field and down weights events for low multiple coherence between the output electric field and the input magnetic components [*Jones et al.*, 1989].

[23] Variance of $Zij^{\mathbf{R}}$ is correctly defined by the equation (13) only if (1) **R** is uncorrelated with the noise in **E** and **H**, (2) the noises in **E** and **H** are independent of the signals, and (3) the noises are stationary. The purpose of RR technique is to ensure that the first condition is satisfied. The second assumption is likely to be well satisfied if the noises are generated locally. On the other hand, if the noises arise from inhomogeneous atmospheric source, both assumptions 1 and 2 may be violated. Assumption 2 may also be violated if the measuring equipment produces errors that are proportional to the signal [*Gamble et al.*, 1979b]. The requirement of noise stationarity is not particularly restrictive. Here we do not need to assume that the signals are stationarity. Z^R and errors in Z^R involve only the ratios of average cross powers and since the electric and magnetic fields (both from local and remote) are causally related, these ratios do not depend on the statistics of the field. *Stodt* [1983] has shown that the scheme tends to increase signal/noise ratio.

3.3. Robust M Estimation

[24] Robustness signifies some level of insensitivity to a small number of outliers in the data. For MT data, robust M estimation (RME) are used [*Egbert and Booker*, 1986; *Chave et al.*, 1987; *Chave and Thompson*, 1989; *Jones et al.*, 1989; *Larsen*, 1989; *Sutarno and Vozoff*, 1989, 1991; *Egbert and Livelybrooks*, 1996; *Bhattacharya and Shalivahan*, 1999]. The impedance estimates by RME are robust against violations of distributional assumption and thus are resistant to outliers. The weights in this case are determined iteratively from the normalized residuals (r). The Huber weights as used by *Egbert and Booker* [1986], *Chave et al.* [1987], and *Egbert and Livelybrooks* [1996] are given as

$$w_i = \begin{cases} 1 & |r_i| \le 1.5 \\ 1.5/|r_i| & |r_i| > 1.5 \end{cases} \tag{14}$$

and

$$r_i = \frac{\left[E_{xi} - \left(Z_{xx}H_{xi} + Z_{xy}H_{yi}\right)\right]}{\hat{\sigma}}$$

Here $\hat{\sigma}$ is the estimate of the scale of the error in the impedance estimation and determines which of the residuals are to be regarded as large. The median absolute deviation from median (MAD) gives one of the most robust estimates of scale. The sample value of it is given as:

$$S_{\mathrm{MAD}} = |r - r'|_{(N+1)/2} \tag{15}$$

Where N is the total number of values quantity r' is the median of **r**. The theoretical MAD is the solution σ_{MAD} of

$$F(\mu' + \sigma_{\mathrm{MAD}}) - F\left(\tilde{\mu'} - \sigma_{\mathrm{MAD}}\right) = 1/2 \tag{16}$$

Where μ' is the theoretical median and F denotes the target cumulative distribution function. Robust processing proceeds as follows: the LS approach determines the initial estimate of the impedance at each frequency, and is further used to compute the residuals **r** in (1) and $\hat{\sigma}$ from the ratio of equations (15) and (16). An iterative procedure is then applied with the weights as in (14) where the residuals from the previous iteration are used to get scale and weights. This process is repeated until convergence is reached. *Huber* [1981] has proved that the weights as used in equation (14) converge.

[25] The RME can be even severely biased than the least squares when typical signal/noise ratios are low (in dead band). It is also worth noting that the generalized RME, which down weights leverage points which may actually down weight or throw away the real data with the best signal-to-noise ratio exacerbating the bias problem.

3.4. Hybrid Approach

[26] The problem of affecting estimates when working with low signal data (in dead band) have been dealt by *Park* [1991] and *Larsen et al.* [1996] by reanalyzing all of the time series points each time a new estimate is made. This way, the potential bias from the a few bad transfer functions estimates is quickly identified and eliminated.

[27] In order to overcome the problem of estimates in the dead band we apply hybrid RME. The data recorded during the periods of high signal power are weighted more heavily and then RME is applied. The weights are determined using CWE and rho-var, and subsequently, RME is applied: (1) CWE plus RME, as CWE tends to improve the S/N ratio

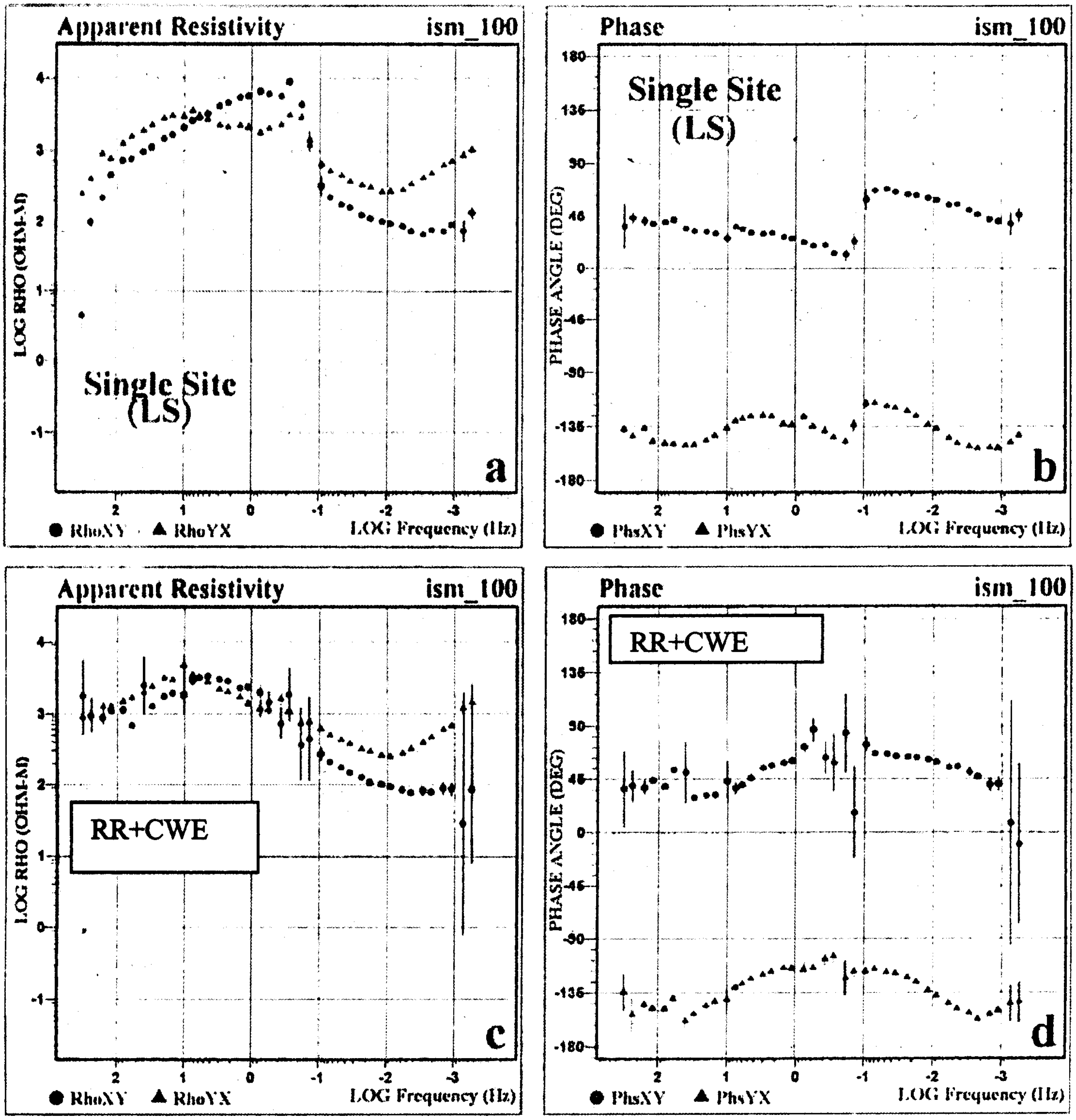

Figure 2. Apparent resistivity and phase curves, respectively, for site 100: (a) and (b) standard LS; (c) and (d) CWE. Figures 2a and 2b are single site and Figures 2c and 2d are RR site estimates. Apparent resistivity and phase curves, respectively, for site 100: (e) and (f) rho-var; (g) and (h) RME. Figures 2e to 2h are RR site estimates. Apparent resistivity and phase curves, respectively, for site 100: (i) and (j) CWE+RME; (k) and (l) rho-var weighting + RME. Figures 2i to 2l are RR site estimates. Apparent resistivity and phase curves, respectively, for site 100: (m) and (n) CWE + rho-var weighting + RME. Figures 2m and 2n are RR site estimates.

and hence can decrease the bias effect; subsequently, the RME is applied to this CWE data [*Egbert and Livelybrooks*, 1996]; and (2) rho-var weighting plus RME, as rho-var tends to improve the signal/noise ratio and hence decreases the bias effect due to the low signal data. Using rho-var, we weight more heavily the data recorded during the periods of high signal power. The RME is then applied to this rho-var weighted data (J. A. Stodt, unpublished report, 1980).

3.5. Extra Hybrid Approach (CWE Plus rho-var Weighting Plus RME)

[28] The extra hybrid scheme works in three stages. First, it weights according to the coherences of the induced

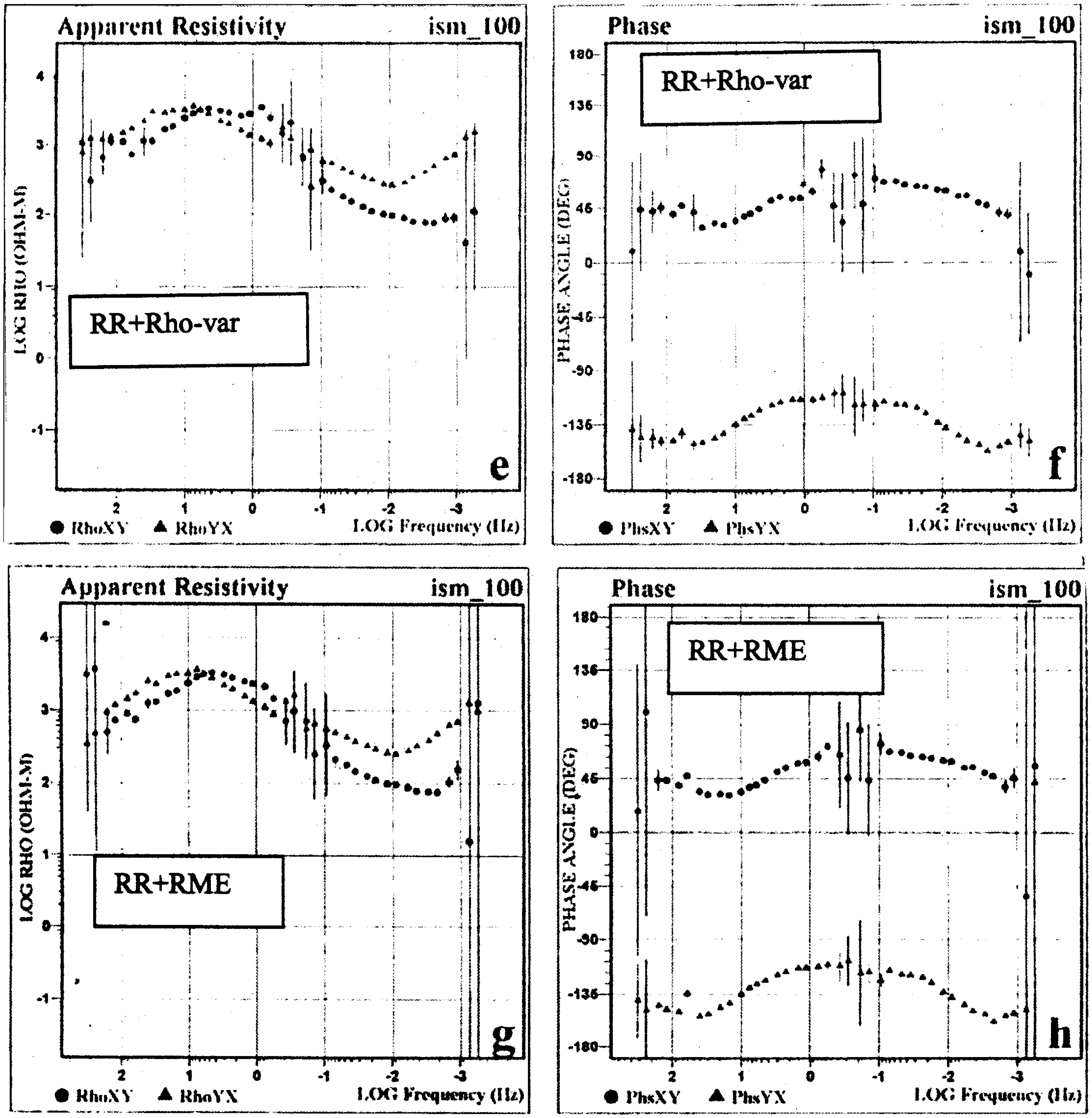

Figure 2. (continued)

electric field with the corresponding component of magnetic field and minimizes the effect of incoherent noise on electric field. This step is followed by inverse variance as weights. Finally, the iterative reweighting RME scheme is used to clean up outliers.

4. Discussion and Conclusions

[29] Forty-four RR MT data were collected over the Dhanbad-Badampahar transect (Figure 1). We present the different RR MT processing techniques with wide band data for a site Sundarpahari (100) (23°57′46″N; 86°32′05″E) whose reference site was at Patamda (021) (22°53′26″N, 86°21′41″E).

[30] For processing, the data are segmented into 1024 points per channel records. In level 3 (320–60 Hz) and level 4 (40–7.5 Hz), only two records are continuous at each minute of processing. However, in level 5 (6–0.00055 Hz) the sample rate is only 24 Hz and is sampled continuously. Each 1024-point record is divided into 32 segments of 32 point each. Several segments are processed together for each frequency band and then a weight factor is calculated for this small group. The weighted cross power is then added to accumulated cross-power sum. The cosine window is applied to each 32 points in a data record. The

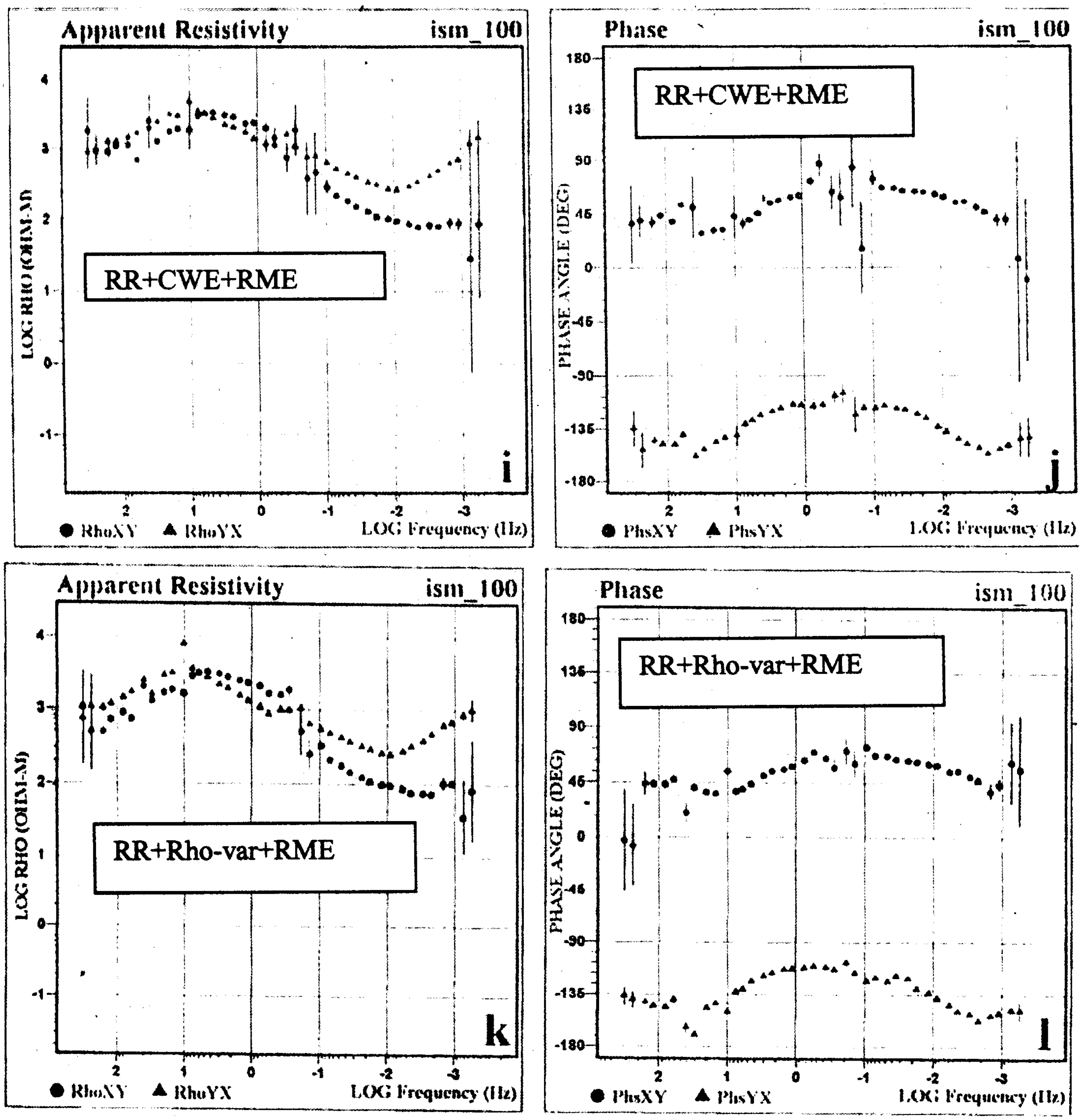

Figure 2. (continued)

14-point cascade decimation after [*Wight and Bostick*, 1980] filter is applied to the data to get the next level of time series. In level 5, this is applied 14 times to get the low period information.

[31] Figures 2a and 2b are the single site standard LS estimates showing apparent resistivity and phase curves, respectively. CWE and rho-var weighting estimates (Figures 2c, 2d, 2e, and 2f, respectively) with RR processing are smooth compared to the single site standard LS estimates. However, the estimates for some frequencies are not precise (larger errors). However, the data in the dead band are biased. RR RME (Figures 2g and 2h) show biased data and large bars in the dead band as compared to the single site standard LS estimates. RR CWE with RME (Figures 2i and 2j) does not show improvement in this band. RR rho-var weighting with RME (Figures 2k and 2l) reduces both the bias and error bar considerably in this band. For the extra hybrid estimates (i.e., CWE+rho-var weighting+RME) (Figures 2m and 2n) the error bars are dramatically reduced in this band and thus the estimates are smooth as well as precise as compared to all other RR estimates.

[32] The LS estimates (both amplitude and phase; Figures 2a and 2b) are distorted. This may be explained by the presence of coherent noise events in the original time

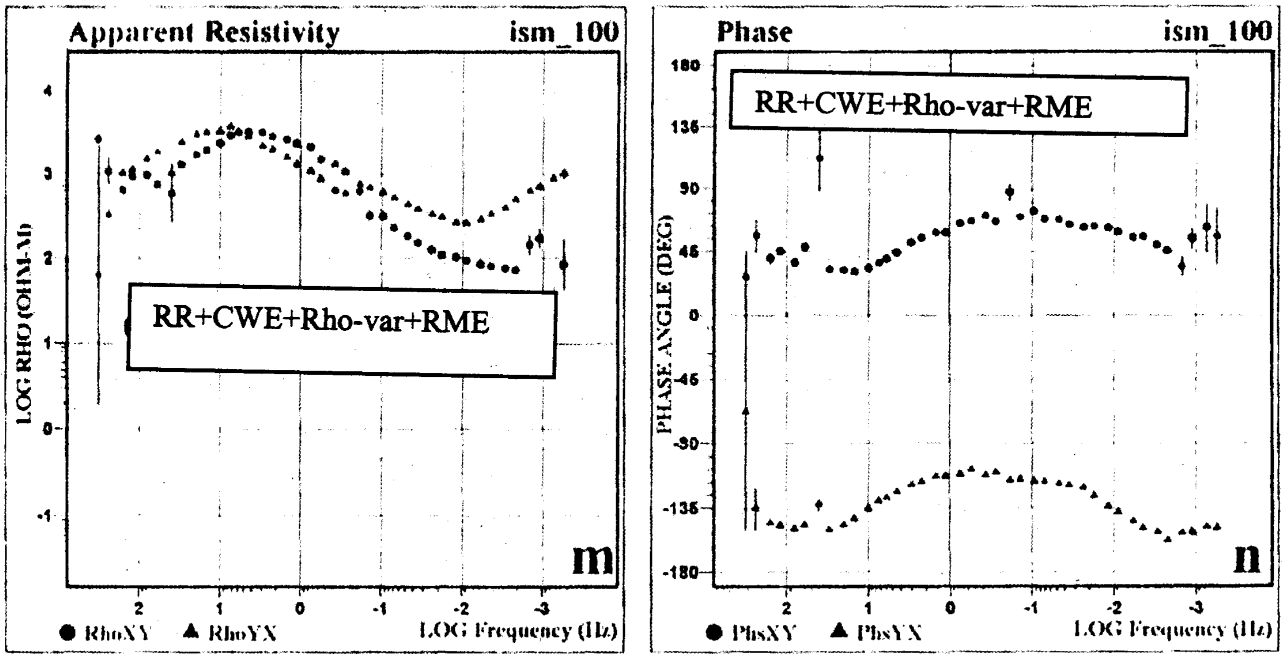

Figure 2. (continued)

series that are not eliminated either by CWE or rho-var when applied in conjunction with the remote reference. The robust estimates are also distorted. The estimates are biased in the hybrid approach (CWE+Robust) scheme indicating the presence of high coherency between the orthogonal components of the magnetic field. The estimates obtained from rho-var+RME improve estimation but still remain distorted indicating the presence of coherent events between the output electric field component and the input magnetic field components. The distortions in the amplitude of the transfer functions might be observed when significant uncorrelated noise is present in input channels and also the coherency between the orthogonal components of the magnetic field is large. However, the phases are usually not distorted. The distortion of the phase here is an indication of the presence of correlated noise. The extra hybrid approach gives estimates which are not distorted indicating that the coherent events between the output electric field and input magnetic field components as well as between the two orthogonal components of magnetic field along with the performance of RME removes the distortion. However, for the xy component of both amplitude and phase all the processing schemes resulted in the distorted estimation for the longest period indicating that either the signal strength is too low or the presence of uncorrelated noise in input channels and also the coherency between the orthogonal components of the magnetic field is large. Distorted phase in the lowest frequency indicates the presence of correlated noise.

[33] The standard practice of MT impedance estimation is the use of robust processing in conjunction with the remote referencing. The use of remote reference method helps in overcoming the bias effect from uncorrelated local magnetic field noise. The estimation by robust technique helps in reducing the effect of outliers in the electric field but is often not sensitive to the exceptional predictor (magnetic field) data, which are called leverage points. In the dead band often due to the low levels of signals and high levels of noise the signal/noise ratio is often very low. Thus the outlier data are rare time segments with useful signal. The robust technique may thus actually down weights or throw way data with the best signal-to-noise ratio and thus exacerbating the bias problem. This problem is overcome when the data recorded during higher signal power is given larger weights than that recorded during lower signal power and then subsequently applying the robust technique. In general it has been found that when the data is weighed both with the CWE and rho-variance technique and then the robust technique is applied it gives a smooth as well as precise estimates of the impedance. Thus the application of hybrid approach helps in reducing the influence of both the outliers in the electric field and the leverage points. We also processed a data very close to DC electrified railway track (the results not presented here) and found that the proposed technique fails here. Because of the presence of DC electrified railway track there is a contamination of the MT signal by nonuniform EM cultural noise sources such as electrified railway track, since this type of source is at ground level and it does not give the same field configuration that of the ionospheric sources for MT field. MT transfer functions cannot be estimated from time series highly contaminated by correlated noise signals by using a remote magnetic site free of correlated noise.

[34] **Acknowledgments.** The work has been carried out for the projects funded by the Department of Science and Technology (DST), Government of India and University Grants Commission's Special Assistance Programme (UGC SAP). Authors express thanks to Sobhan Pathak, presently in Indian Oil Corporation, New Delhi, Priya R. Mohanty, Ranjit K. Shaw, Samir K. Acharya, Department of Applied Geophysics (AGP), Indian School of Mines (ISM) Dhanbad, and Raja Mukhopadhyay presently in Central Water and Power Research Station (CWPRS), Pune, for assisting

in collecting data. All the computations were carried out in the computer centre of UGC SAP of AGP, ISM. R.K.S. also thanks CWPRS. The last author (B.B.B.) thanks Council of Scientific and Industrial Research (CSIR) and National Academy of Engineers (INAE) for their support.

References

Banks, R. J. (1998), The effects of non-stationary noise on electromagnetic response estimates, *Geophys. J. Int.*, *135*, 553–563.

Bhattacharya, B. B., and Shalivahan (1999), Application of robust processing of MT data using single and remote sites, *Current Sci.*, *76*, 1108–1113.

Chave, A. D., and D. J. Thompson (1989), Some comments on magnetotelluric response function estimation, *J. Geophys. Res.*, *94*, 14,215–14,226.

Chave, A. D., and D. J. Thompson (2004), Bounded influence magnetotelluric response function estimation, *Geophys. J. Int.*, *157*, 988–1006.

Chave, A. D., D. J. Thompson, and M. E. Ander (1987), On the robust estimation of power spectra coherences and transfer functions, *J. Geophys. Res.*, *92*, 633–648.

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

Egbert, G., and J. R. Booker (1986), Robust estimation of geomagnetic transfer functions, *Geophys. J. R. Astron. Soc.*, *87*, 173–194.

Egbert, G., and D. W. Livelybrooks (1996), Single station magnetotelluric impedance estimation: Coherence weighting and the regression M-estimate, *Geophysics*, *61*, 964–970.

Egbert, G. D., M. Eisel, O. S. Boyd, and H. F. Morrison (2000), DC trains and Pc3s: Source effects in mid-latitude geomagnetic transfer functions, *Geophys. Res. Lett.*, *27*, 25–28.

Gamble, T. D., W. M. Goubau, and J. Clarke (1979a), Magnetotellurics with a remote reference, *Geophysics*, *44*, 53–68.

Gamble, T. D., W. M. Goubau, and J. Clarke (1979b), Error analysis for remote reference magnetotellurics, *Geophysics*, *44*, 959–968.

Garcia, X., and A. G. Jones (2002), Atmospheric sources for audiomagnetotelluric (AMT) soundings, *Geophysics*, *67*, 448–458.

Garcia, X., A. D. Chave, and A. G. Jones (1997), Robust processing of magnetotelluric data from the auroral zone, *J. Geomagn. Geolectr.*, *49*, 1451–1468.

Goubau, W. M., T. D. Gamble, and J. Clarke (1978a), Magnetotellurics using loch-in signal detection, *Geophys. Res. Lett.*, *5*, 543–546.

Goubau, W. M., T. D. Gamble, and J. Clarke (1978b), Magnetotelluric data analysis. removal of bias, *Geophysics*, *43*, 1157–1166.

Goubau, W. M., P. M. Moxton, R. H. Koch, and J. Clarke (1984), Noise correlation lengths in remote reference magnetotellurics, *Geophysics*, *49*, 432–438.

Huber, P. J. (1981), *Robust Statistics*, John Wiley, Hoboken, N. J.

Jones, A. G., and H. Jodicke (1984), Magnetotelluric transfer function estimation improvement by a coherence based rejection, paper presented at 54th Annual International Meeting, Soc. of Explor. Geophys., Atlanta, Ga.

Jones, A. G., and J. Spratt (2002), A simple method for deriving the uniform field MT response in auroral zones, *Earth Planets Space*, *54*, 443–450.

Jones, A. G., A. D. Chave, G. D. Egbert, D. Auld, and K. Bahr (1989), A comparison of techniques for magnetotelluric response function estimation, *J. Geophys. Res.*, *94*, 14,201–14,214.

Junge, A. (1996), Characterization and correction for cultural noise, *Surv. Geophys.*, *17*, 361–391.

Larsen, J. C. (1989), Transfer functions: Smooth robust estimates by LSs and remote reference methods, *Geophys. J. Int.*, *99*, 655–663.

Larsen, J. C., R. L. Mackie, A. Mazella, A. Fiordelisi, and S. Rieven (1996), Robust smooth magnetotelluric transfer functions, *Geophys. J. Int.*, *124*, 801–819.

Park, S. K. (1991), Monitoring resistivity changes prior to earthquakes in Parkfield, California with telluric array, *J. Geophys. Res.*, *96*, 14,211–14,237.

Rousseeuw, P. J., and A. M. Leroy (1987), *Robust Regression and Outlier Detection*, John Wiley, Hoboken, N. J.

Schultz, A., R. D. Kurtz, A. D. Chave, and A. G. Jones (1993), Conductivity discontinuities in the upper mantle beneath a stable craton, *Geophys. Res. Lett.*, *20*, 2941–2944.

Shalivahan (2000), Nonlinear inversion of electrical and magnetotelluric data using very fast simulated annealing, Ph.D. thesis, Indian School of Mines, Dhanbad, India.

Shalivahan, and B. B. Bhattacharya (2002), How remote can the far remote reference site for magnetotelluric measurements?, *J. Geophys. Res.*, *107*(B6), 2105, doi:10.1029/2000JB000119.

Sims, W. E., F. X. Bostick, and H. W. Smith (1971), The estimation of magnetotelluric impedance tensor elements from measured data, *Geophysics*, *36*, 938–942.

Smirnov, M. Y. (2003), Magnetotelluric data processing with a robust statistical procedure having a high breakdown point, *Geophys. J. Int.*, *152*, 1–7.

Stodt, J. A. (1983), Noise analysis for conventional and remote reference magnetotellurics, Ph.D. thesis, Univ. of Utah, Salt Lake City.

Sutarno, D., and K. Vozoff (1989), Robust M-estimation of magnetotelluric impedance tensor, *Explor. Geophys.*, *20*, 383–398.

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

Vozoff, K. (1996), The magnetotelluric method, in *Electromagnetic Methods in Applied Geophysics*, edited by M. N. Nabighian, pp. 641–711, Soc. of Explor. Geophys., Tulsa, Okla.

Wight, D. E., and F. X. Bostick (1980), Cascade decimation—A technique for real time estimation of power spectrum, paper presented at IEEE International Conference on Acoustic, Speech Signal Processing, Denver, Colo., 9–11 April.

B. B. Bhattacharya, S. N. Bose National Centre for Basic Sciences, Sector-III, Block-JD, Salt Lake, Kolkata -700 098, India.

Shalivahan, Department of Applied Geophysics, Indian School of Mines, Dhanbad 826 004, India. (svismagp@yahoo.co.in)

R. K. Sinharay, Central Water and Power Research Station, Khadakwasla, Pune, 411 024, India.