Application of robust processing of magnetotelluric data using single and remote reference sites

Bimalendu B. Bhattacharya and Shalivahan

Department of Applied Geophysics, Indian School of Mines, Dhanbad 826 004, India

The estimation of transfer function from magnetotelluric (MT) sounding data is not satisfactory in many cases due to the non-stationarity of the fields. This estimate is essentially a statistical problem. The least square approach is not satisfactory in many cases. The robustness signifies intensitivity to small deviations from the assumptions. Robust processing is a reasonable procedure for rejecting the outliers and thus preventing the catastrophic effect of bad data points.

The maximum likelihood estimate (M) regression is robust. Several MT sounding data along the Dhanbad–Dalma–Singhbhum–Badampahar section have been collected. These data have been reprocessed by robust method with reference to the same site as well as a remote reference site. The robust processed result of one such for data with reference to the same site as well as a remote site situated 45 km away are presented. It is seen that the robust processing of even such noisy data improves the data quality considerably. The improvement is remarkable when the processing is carried out with a remote reference site.

THE first step in the interpretation of magnetotelluric (MT) data generally involves the estimation of a frequency-dependent relationship between the electric (E) and magnetic (H) fields expressed by the relation

$$E(\omega) = Z(\omega)H(\omega), \tag{1}$$

where $Z(\omega)$ is the impedance tensor. Equation (1) is justified if the external source for the field is a plane wave. This assumption is an approximation and there are always measurement errors. In reality, therefore, the transfer function is determined for a set of data which is a deviation from the exact. Thus the estimation of transfer function is essentially a statistical problem.

The least square (LS) method has generally been used for the estimation of transfer function. Sims *et al.*[1] computed 6 different MT impedance tensors from the measured data by using auto-power and cross-power density spectra. Each estimation satisfies a mean-square error criterion. Computations of all of the estimates provide a measure of the total amount of noise present, as indicated by a stability coefficient for the estimates. In the absence of additional information about the relative signal-to-noise ratios of E and H signals, Sims *et al.* suggested that a mean estimate be used.

LS procedures are quite simple and commonly used. The Gaussian distribution is given by

$$\frac{1}{\sigma(2\pi)^{0.5}}\exp\left[\frac{(x-\langle x\rangle)^2}{2\sigma^2}\right], \tag{2}$$

where $\langle x\rangle$ and σ are mean and variance respectively.

The failure of eq. (1) is due to the assumption that the error variance is independent of the signal power. Basically, misfit of the data to the linear model is due to the failure of the uniform source field assumption. The magnitude of the misfit is dependent on the strength of the source because the earth's response is proportional to the source magnitude. It is expected to give an appreciable amount of outlier and such a situation is poorly modelled by eq. (2). Also some related errors occur in cluster so that the uncorrelated error assumption may fail. In a nutshell, a few bad points dominate the estimate in Gaussian distribution.

Robust method

According to Huber[2], 'robustness signifies insensitivity to small deviations from the assumptions'. Basically one is concerned as the shape of the true underlying distribution deviates slightly from the assumed model defined usually by Gaussian law. Any reasonable procedure for rejecting outliers will prevent the catastrophic effect of bad data points. Therefore, in such cases, a two-step approach should suffice: (1) cleaning of data by applying some rule for outlier rejection; (2) using estimation and testing procedures on the remainder. These steps cannot perform the same job in a simpler way for the following reasons[2]:

(i) It is generally not possible to separate the two steps clearly. In multiparameter regression problems, outliers are difficult to recognize unless reliable robust estimates for the parameters are available.
(ii) If the original batch of observations consists of normal observations with some gross errors, the

cleared data will not be normal. The situation is much worse when the original batch is derived from non-normal distribution. In such cases, the actual performance of such a two-step procedure is difficult to work out than a straight robust procedure.

(iii) Empirically, it has been found that the robust procedure is better than the best rejection procedures.

The robust prcedure is, therefore, expected to achieve the following:

(1) An optimal efficiency at the assumed model.
(2) It should be robust in the sense that small deviations from the actual model should not influence the performance. It should be close to the normal value of the model.
(3) The larger deviations from the model should not cause wide-scale disturbances.

Robust in MT

Neither the LS nor the standard error estimates are robust. The maximum likelihood estimate (M) regression is robust though it appears to be similar to the LS. Here, one estimates the impedance tensor which minimizes an expression of the form[3]

$$\sum_i \rho\left[\frac{E_i - \hat{Z}H_i}{\sigma}\right], \qquad (3)$$

where $\rho(r)$ is some loss function. It corresponds to L_2 minimization for small residuals and to L_1 minimization for large residuals. The L_n norm is expressed as:

$$\|e\|_n = \left[\sum_i |e_i|^n\right]^{1/n},$$

where e is the prediction error or misfit between the observed and the predicted data, also called the residual.

We have used the loss function as[2]:

$$\rho(r) = \begin{cases} \dfrac{r^2}{2} & |r| < r_0 \\ r_0|r| - \dfrac{r_0^2}{2} & |r| \geq r_0 \end{cases}, \qquad (4)$$

with $r_0 = 1.5$. For Gaussian errors, possibly contaminated by a small number of outliers, transition points of $r_0 = 1.5\,\sigma$ of the (uncontaminated) Gaussian error distribution work well.

The scale parameter (σ) with r_0 determines which residuals are considered to be large. The scale parameter is a measure of dispersion. A robust estimate of the residual standard deviation for the error scale has been used. The minimum of eq. (3) could be found by solving

$$\sum_i \psi\left[\frac{E_i - H_i\hat{Z}}{\sigma}\right]H_i = 0, \qquad (5)$$

where $\psi(r) = \rho'(r)$.

Let $w(r) = \psi(r)/r$. Starting with the LS estimates of the impedance tensor ($\hat{Z}_0$) and error scale $\hat{\sigma}_0$, compute the predicted and residual E fields

$$\hat{E}_{i0} = \hat{Z}_{i0}H_i$$

$$r_{i0} = E_i - \hat{E}_{i0},$$

and the modified observations

$$E_{i1} = \hat{E}_{i0} + w\left[\frac{r_{i0}}{\hat{\sigma}_0}\right]r_{i0}.$$

Then $\rho(r)$ given by eq. (4), reduces to:

$$\rho(r) = \begin{cases} 1 & |r| \leq r_0 \\ \dfrac{r_0}{|r|} & |r| \geq r_0 \end{cases},$$

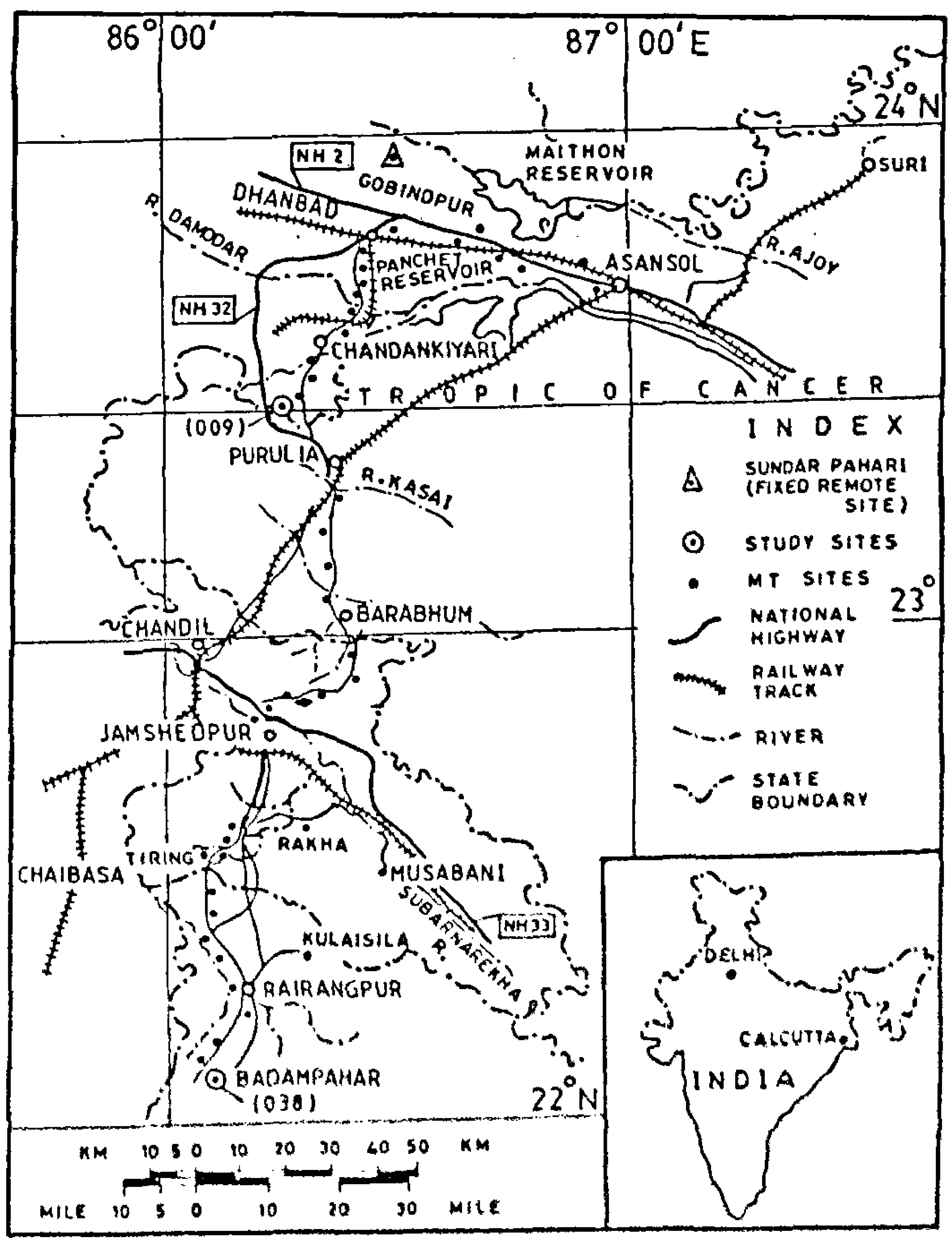

Figure 1. Magnetotelluric site locations over the transect Dhanbad-Dalma-Singhbhum-Badampahar.

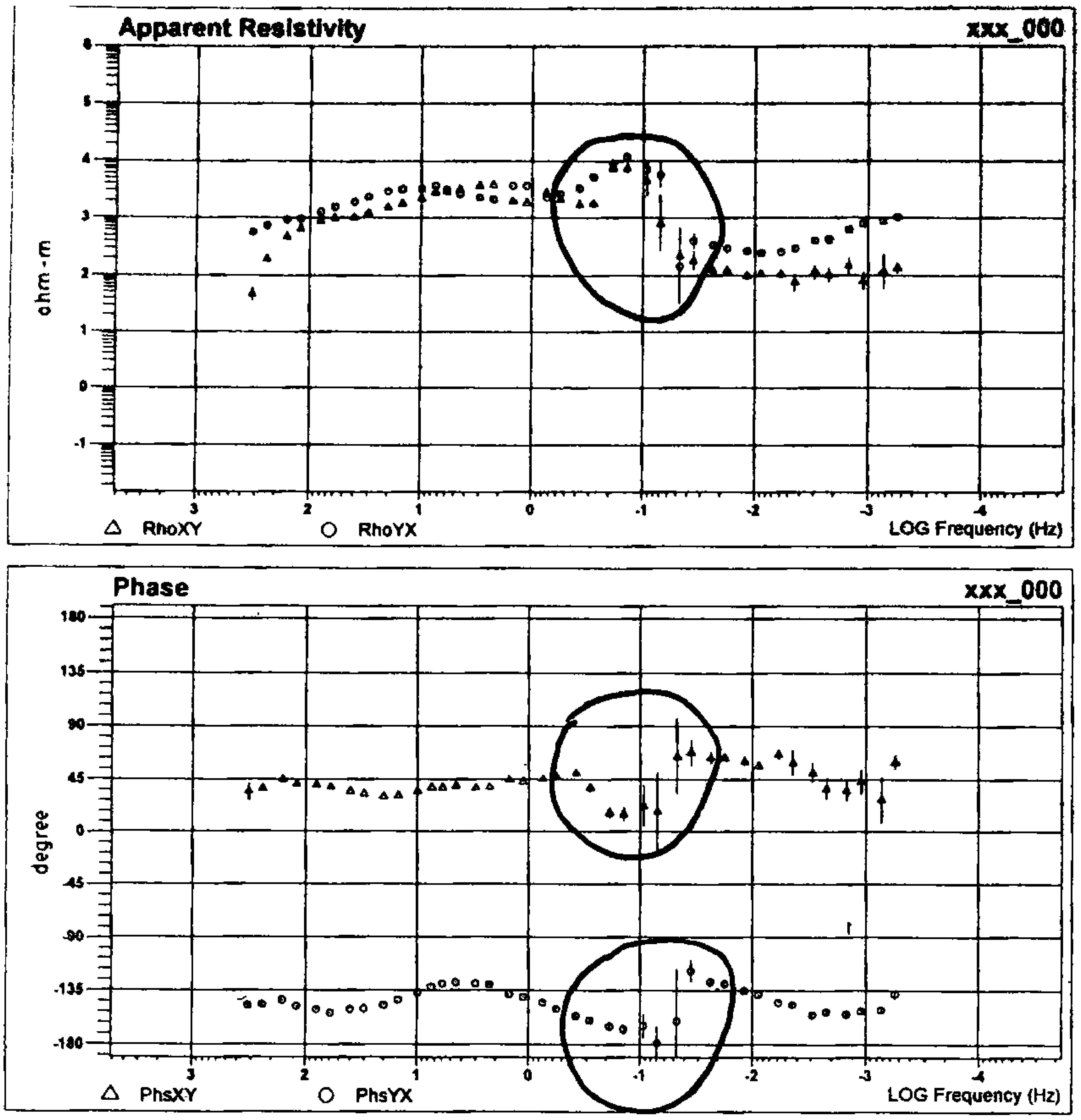

Figure 2. Acquired magnetotelluric data at Sundar Pahari. (Top) apparent resistivity; (bottom) phase. Bars represent standard deviation.

so that the modified observations are identical to the original observations for small ($\leq r_0$ standard deviations) residuals while for larger residuals the observations approach the predicted values. Now using the modified observations (E_1) in place of the originals ($E_0 \equiv E$), compute new LS estimates of the transfer functions[2]

$$Z_1 = (H^T H)^{-1} (H^T E_1),$$

together with a new error scale estimate $\hat{\sigma}_1$.

An iterative application of this procedure is used to compute the regression M-estimate. Thus, from the nth iteration new modified observations of field, impedance and error scale are computed. The procedure is repeated until convergence to the solution of eq. (5) is achieved when the following conditions are satisfied:

$$\rho(0) = 0$$

$$\rho'(r) \geq 0, 0 \leq \rho'(r) \leq 1$$

Huber[2] has described the algorithm for the computation of regression M-estimates. Egbert and Booker[3] have shown that the algorithm converges when $\rho(r)$ is convex ($\rho'' > 0$). This requires that ψ be a nondecreasing function. Ideally one would like to discard completely the very bad data points – the worst outliers. It means $\psi(r) \approx 0$ for large values of r. In other words, the terms corresponding to large residuals in eq. (5) are completely omitted. In such a situation, ψ must be non-monotone. The non-uniqueness of the solution of ψ can lead to very bad results if the starting estimate is poorly determined. Huber[2] suggested the use of non-monotone ψ for final iteration after the convergence is achieved with monotone ψ. To eliminate the very bad points completely, two final iterations are made with

$$\psi(r) = r \exp\{-\exp[r_0(|r| - r_0)]\},$$

using $r_0 = 2.8$. For small residuals the r_0 value can be 1.5.

The robust scheme completely eliminated the worst outliers by using a non-monotone ψ for final iterations.

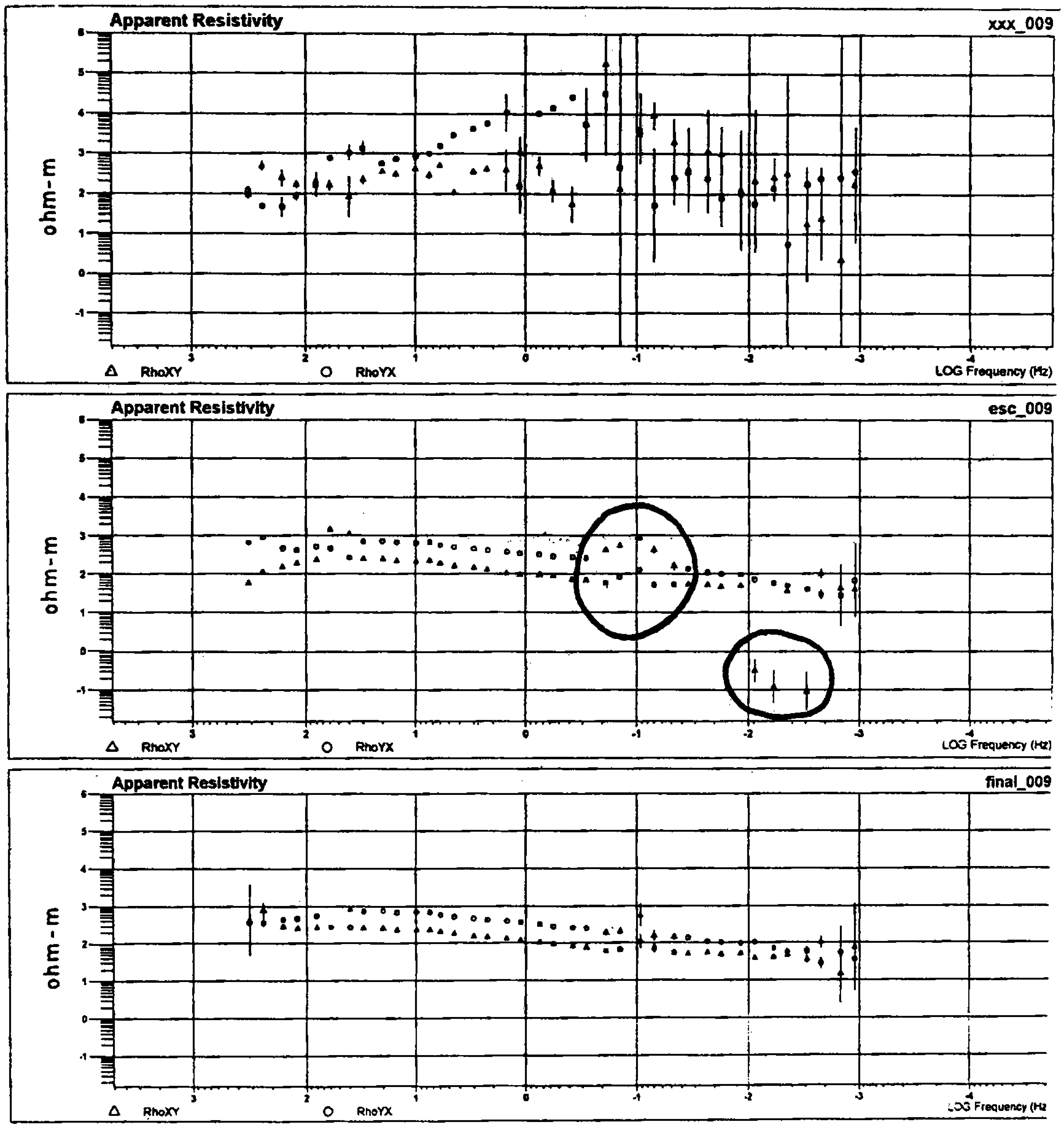

Figure 3. Apparent resistivity from magnetotelluric data. (Top) acquired data at site 009; (middle) robust processed data with the site itself; (bottom) robust processed data with remote site. Poor quality processing is shown by encircled zone. Bars represent standard deviation.

Data acquisition

A brief account of MT data acquisition and processing system of V5-16 multiple geophysical receiver of Phoenix Geophysics Limited, Canada used in field work is described here. A site with 2 tellurics and 3 magnetics with a single V5-16 was hooked up to a signal processor unit (SPV5). A remote reference station was also set up at distances varying from 20 km to 215 km. H reference has been used in all the MT sounding configurations. V5 system is a true real time acquisition system. The data acquisition in V5 system consists of two steps, firstly a pre-acquisition step in which the data about the sounding is entered and system gains are set-up and secondly data acquisition.

During pre-acquisition data entry, various parameters like powerline frequency, number of active channels, type of sensor on each channel, sounding name and number, location, elevation, coil number, telluric line lengths and electrode impedance are noted. Then a sounding calibration file is created from channel and sensor responses previously received using the V5 MT calibration program. Finally the sounding data are stored in non-volatile memory to be dumped at a future time or re-accessed if the sounding must be restarted. After all the parameters are entered, the data acquisition is started.

MT data acquisition in V5 is divided by frequency in two ranges[4]: high range, 12 frequencies ranging from

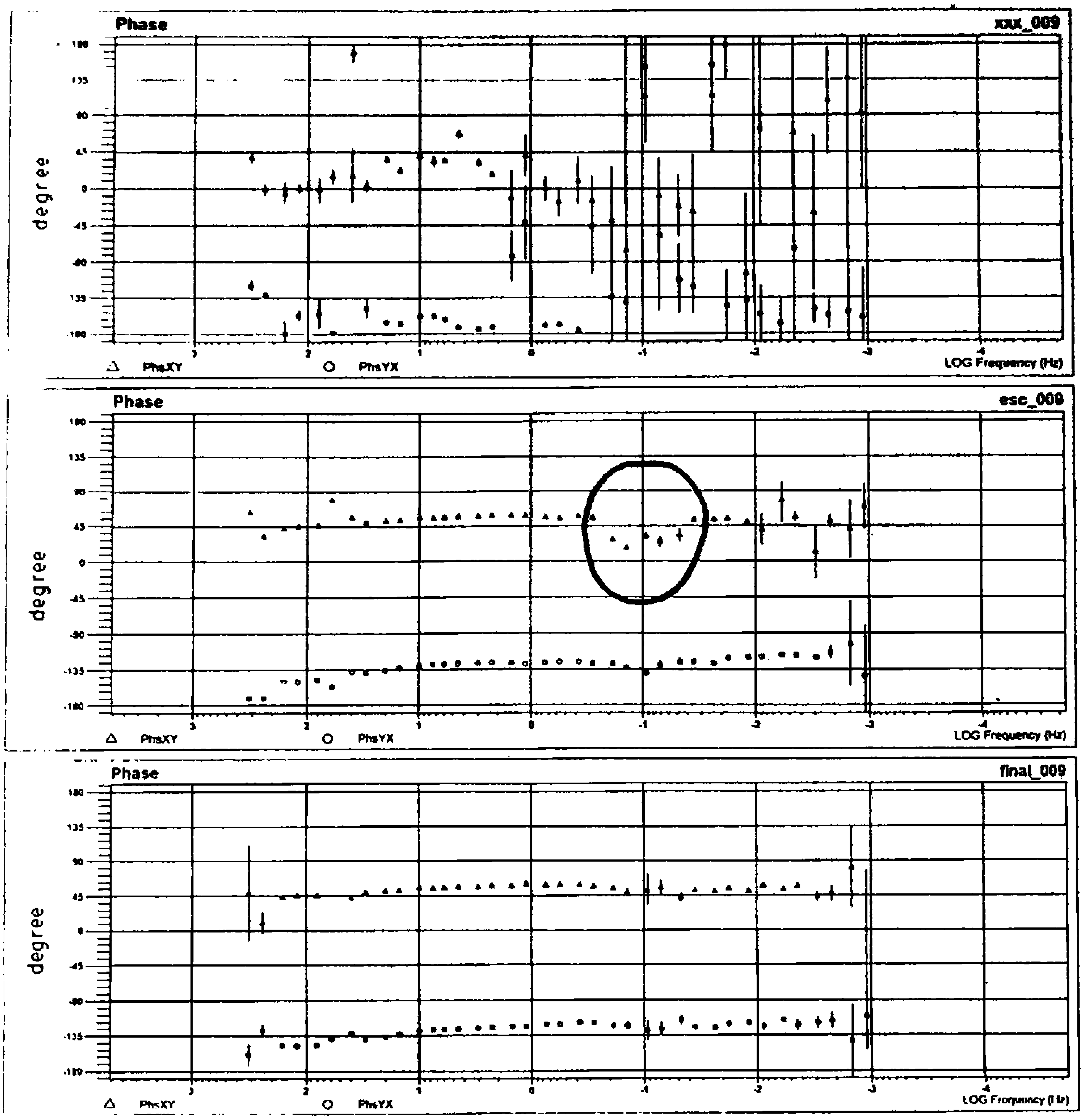

Figure 4. Phase from magnetotelluric data. (Top) acquired data at site 009; (middle) robust processed data with the site itself; (bottom) robust processed data with remote site. Bars represent standard deviation.

320 Hz to 7.5 Hz and low range, 28 frequencies ranging from 6.0 Hz to 0.0005 Hz.

Field results

For robust processing, the magnetic field is taken as the reference pair as it is generally less contaminated with noise as compared to the electric field. Robust processing uses a subroutine which requires the following: type used, coherency value and use of rho variance. The value of rho variance is to be stated if used.

For remote reference processing the type used is 7 while for single site processing it is 1. The number 7 is assigned when the combined weighting factor of the multiple coherence of R_x with the total magnetic field and the multiple coherence of R_y with the total magnetic field is used. The number assigned will become 1 when the site uses the combined weighting factor of the multiple coherence of E_x with the total magnetic field and the multiple coherence of E_y with the total magnetic field is used. Coherency value used is 0.80 while the rho-variance value is 0.75.

Forty-three pairs of MT sounding data have been collected along the Dhanbad–Dalma–Singhbhum–Badampahar section using a remote reference site with varying distances of 20 km to 215 km from the measuring site. All these data have been robust processed with respect to the same site as well as the remote site. Here we present the robust processed MT data over a single site (009) near Purulia situated along the Dhanbad–Dalma–Singhbhum–Badampahar section (Figure 1). The corresponding acquired data for the remote reference site at Sundar Pahari (23°57′46″N, 86°32′09″E; Figure 1) are shown in Figure 2. Considerable care has been taken to select the reference site at a place having less external noise. The quality of the acquired data in the remote site in this case was very good. However, a shift of the data is seen in the dead band shown by an encircled zone in Figure 2.

Figures 3 and 4 (top) show the apparent resistivity (rho*XY* and rho*YX*) and phase (phs*XY* and phs*YX*) data respectively, obtained at the site 009. The acquired MT data is highly noisy in most of the frequencies. Figures 3 and 4 (middle) are the corresponding robust processed data with the site itself. Figures 3 and 4 (bottom) are the corresponding data obtained after robust processing with a remote site situated 45 km from the station. The bars at each observed frequency give the standard deviation of the acquired data as well as that after robust processing. The robust processing with same site as well as remote reference improves the signal-to-noise ratio remarkably in almost all the frequencies. However, it is to be noted that the dead band problem, due to the weak and fluctuating signals with dominant noise in that band, does not get resolved satisfactorily by robust processing with the site itself (Figures 3 and 4 (middle)). These zones are encircled in these two figures. The processed data with the same site in the apparent resistivity curve in the frequency range 0.01 Hz to 0.001 Hz (Figure 3, middle), shown by another encircled zone, show unsatisfactory results. The robust processed data with remote reference, however, give the best result for the entire range of observations. The dead band effect has also been removed by this processing.

Conclusions

The field example presented here clearly shows that robust processing improves the data quality considerably even when the data are very noisy. Robust processing with far remote site improves the data quality further. This improvement is also seen in the dead band of MT signals. It is therefore recommended that all MT data must be processed by robust technique with remote reference site for improving their quality. The bias error of the impedance estimation is avoided by using the remote reference technique. Robust processing of acquired data with remote site helps in meaningful and better interpretation of MT data for various geological problems.

1. Sims, W. E., Bostick, F. X., Jr. and Smith, H. W., *Geophysics*, 1971, **36**, 938–942.
2. Huber, P. J., *Robust Statistics*, Wiley, New York, 1991.
3. Egbert, G. D. and Booker, J. R., *Geophys. J. R. Astron. Soc.*, 1986, **87**, 173–194.
4. Operating Manual, Phoenix Geophysics Limited, Canada, 1995.

ACKNOWLEDGEMENTS. We thank Sobhan Pathak, Indian Oil Corporation, New Delhi for assisting in collection of data at site 009. We also thank Drs R. K. Shaw and P. R. Mohanty for collection of data at the remote reference site corresponding to station 009. All the computations have been carried out in the Computer Centre of UGC SAP DSA of Applied Geophysics. The work is a part of DST sponsored project, Special Assistance Programme of UGC and AICTE project on MT.

Received 11 December 1997; revised accepted 12 March 1999