Robust Processing for Removing Train Signals from Magnetotelluric Data in Central Italy

Jimmy Larsen[1], Randall Mackie[2], Adolfo Fiordelisi[3], Adele Manzella[4], Shirley Rieven[2]

Affiliations: [1] **NOAA,** Seattle, Washington, [2] MIT, Cambridge, Massachusetts, [3] ENEL/VDAG, Pisa Italy, [4] IIRG, Pisa, Italy

Key Words

Magnetotellurics, robust estimates, electric train signals

Abstract

Smooth magnetotelluric transfer functions are used to represent the frequency relationship between electric and magnetic data. This facilitates the identification and correction of individual outliers in the electric and magnetic data, and the construction of frequency and time weights needed for robust estimates of the transfer functions. Errors in the transfer function are found by jackknife estimates of the solution covariance. Band averages provide alternate estimates of the transfer function and the damping for stabilizing the smooth estimates. The method is tested for: artificial data with specified noise, electric and magnetic outliers, periodic variations, and transfer function; and data from the Larderello geothermal region in central Italy contaminated by electric train signals.

1 Introduction

Magnetotelluric (MT) transfer functions represent the frequency relationship between electric and magnetic data. They are used to construct earth conductivity models and are therefore a critical link between the observed data and the inferred conductivity models. Reliable estimates of the transfer functions are therefore essential and this depends on selecting the correct frequency relationship between the electric and magnetic data and the statistical method used to obtain unbiased estimates of the transfer function. The form of the relationship depends on whether the noise is uncorrelated or correlated and the statistical method, least squares or remote reference, non-robust or robust, depends on whether the noise is intermittent or continuous. Continuous correlated noise causes the most difficulties.

Estimating the transfer function is essentially a process of comparing the magnetic and electric data in the frequency domain. Here we assume that the transfer function is a smooth continuous function of frequency because this is consistent with transfer functions generated from models. The method proposed here uses a one-dimensional (1-D) transfer function $Z^{1D}(\omega)$ as the basic transfer function and two smooth correction functions $D(\omega)$ and $U(\omega)$ so that the 1-D transfer function times the correction functions, $U(\omega)D(\omega)Z^{1D}(\omega)$, fits the data.

The method to obtain smooth estimates is made robust because most data contain gaps, noise, electric outliers, and magnetic outliers (leverage values) that limit the duration of clean sections of data and distort estimates of the transfer functions if they are included in the analysis. We have therefore developed a method for cleaning up every data section by locating and replacing individual electric and magnetic outlier, and a method for determining frequency and time weights that downweight the remaining noise. Most robust methods (e.g., Egbert and Booker, 1986; Chave *et al.*, 1987) do not identify or remove individual outliers other than removing obviously large outliers prior to applying their robust methods. Their methods deal with outliers by downweighting entire data sections but this will fail if all sections contain outliers.

The robust method proposed here uses an iterative procedure similar to reweighted least squares for robust regression analysis (Rousseeuw and Leroy, 1987). The iterative procedure starts with a generic 1-D transfer function or a 1-D transfer function derived from other studies and estimates the correction to the 1-D transfer function by robust least squares or robust remote reference. All large spikes in the magnetic data are temporarily replaced as bias protection to provide an initial robust estimate of the smooth transfer function. For subsequent iterations, the original data are recorrected for gaps, electric and magnetic outliers, and the frequency and times weights recomputed.

Of particular interest is the application of this robust method to MT data taken in the Larderello geothermal field in Central Italy. These data are contaminated by d.c. train signals and other correlated and uncorrelated cultural noise, which render standard analyses useless. The present robust method is able to deal with these types of severe noise and yield reasonable MT estimates thus providing information about the subsurface electrical properties which are useful for geothermal exploration.

2 Previous Robust Methods

Previous robust methods have concentrated on the frequency representation of data sections and dealt with noisy data by downweighting noisy data sections (Egbert and Booker, 1986; Chave *et al.,* 1987; Chave and Thomson, 1989; Egbert *et al.,* 1992). Egbert and Booker (1986) showed, for example, that residuals often become larger when the intensity of magnetic variations increase and they therefore downweighted those data sections by weights that decrease with increasing power in the magnetic variations. They used a variant of this method in Egbert *et al.* (1992) for the 1932 to 1942 hourly Tucson data where, prior to applying their method, short gaps and isolated outliers were replaced by values predicted from magnetic data using an impulse response function. Although section weighting methods overcomes the inherent sensitivity of least squares to data sections having large residuals, it can not cope with individual electric and magnetic outliers that can occur in all data sections. The papers (Larsen, 1980, 1989) were an attempt to identify and remove individual electric outliers but did not consider magnetic outliers. Chave and Thomson (1989) and Thomson and

Chave (1991) advanced the robust methods by introducing the use of jackknife estimates of error and Chave and Thomson (1989) showed the importance of identifying and removing magnetic outliers. These were not considered in the Egbert *et al.* (1992) analysis of the Tucson data or in any of the other robust methods.

3 MT Relationship

The observations are the discrete electric data $e_k(t)$, the local magnetic data $b_k(t)$, and the remote magnetic data $b_k^R(t)$ for $k = 1$ and **2** horizontal components, $t = 1,...,2J$ time values and $i = 1,...,I$ data sections. The data are converted, by Fourier transform, into $E_k(\omega)$, $B_k(\omega)$ and $B_k^R(\omega)$ for $j = 1,..,J$ angular frequencies $\omega_j = \pi j/J$.

3.1 Single-Source

The MT frequency relationship between an electric component and the local horizontal magnetic components is given, for uncorrelated noise, by

$$E(\omega) = \sum_{k=1}^{K} Z_k(\omega)B_k(\omega) + R(\omega) \tag{1}$$

and the relationship for the remote horizontal magnetic components is given by

$$E(\omega) = \sum_{k=1}^{K} Z_k^R(\omega)B_k^R(\omega) + R^R(\omega) \tag{2}$$

for $K = 2$ horizontal magnetic components where $Z_k(\omega)$ and $Z_k^R(\omega)$ are the MT transfer functions we seek and $R(\omega)$ and $R^R(\omega)$ are the residuals containing the noise in the electric and magnetic data. The frequency residuals $R(\omega)$ and $R^R(\omega)$ are used to determine the frequency weights and the time residuals $r(t)$ and $r^R(t)$, based on the inverse Fourier transform of $R(\omega)$ and $R^R(\omega)$, are used to locate outliers and determine the time weights.

3.2 Multi-Source

For cultural (man-made) sources that produces correlated electric and magnetic noise, the MT relationship (1) must be modified to include the cultural noise sources yielding the multi-source relationship

$$E(\omega) = \sum_{k=1}^{K} Z_k^{MT}(\omega)B_k^{MT}(\omega) + \sum_{k=1}^{K} Z_k^C(\omega)B_k^C(\omega) + R(\omega) \tag{3}$$

for MT transfer function $Z_k^{MT}(\omega)$ and cultural transfer function $Z_k^C(\omega)$ that are, in general, different where MT and cultural induced magnetic variations are, respectively, $B_k^{MT}(\omega)$ and $B_k^C(\omega)$, and the uncorrelated noise is $R(\omega)$.

The MT magnetic variations at the contaminated site are estimated from the clean remote magnetic data by

$$B_k^{MT}(\omega) = \sum_{m=1}^{K} T_{km}^R(\omega)B_m^R(\omega) \tag{4}$$

using estimates of the magnetic transfer functions $T_{km}^R(\omega)$ for $k = 1,...,K$ local magnetic components and $m = 1,..,K$ remote magnetic components. Robust least squares estimates of the magnetic transfer functions and cultural magnetic noise are found using the relationship

$$B_k(\omega) = \sum_{m=1}^{K} T_{km}^R(\omega)B_m^R(\omega) + B_k^C(\omega) \tag{5}$$

by assuming the weighted cultural magnetic noise is uncorrelated with the weighted remote magnetic data. The MT magnetic data $b_k^{MT}(t)$ are then given by the inverse Fourier transform of $B_k^{MT}(\omega)$ and the cultural magnetic noise $b_k^C(t)$ are given by the inverse Fourier transform of $B_k^C(\omega) = B_k(\omega) - B_k^{MT}(\omega)$. Robust least squares estimates of $Z_k^{MT}(\omega)$ and $Z_k^C(\omega)$ are then found using equation **(3).** For non-robust methods, the remote reference method and the least squares multi-source method give the same estimate of $Z_k^{MT}(\omega)$ (Egbert, 1994, personal communication).

3.3 Statistical Methods

There are two methods for estimating the transfer functions. The method of least squares (see Sims *et al.*, 1971) minimizes the variance of the residuals, $\langle |R|^2 \rangle$, and the remote reference 'method (Gamble *et al.*, 1979; Larsen, 1989) minimizes the modulus of the covariance between the residuals, $\langle R^{R*} R \rangle$. Here $\langle\ \rangle$ represents a sum over the I data sections and the J frequencies and * represents the complex conjugate.

Unstable estimates of the transfer function will occur if the two horizontal magnetic components are correlated. This is the collinearity problem and it is reduced by rotating the magnetic components so that one magnetic component (the major one) is most coherent with the electric component. The minor magnetic component is then weakly coherent with the electric componentt. This rotation therefore minimizes the collinearity and also provides information about the dimensionality of the data.

4 Smooth Representation

We represent the smooth transfer function for the kth magnetic component by

$$Z(\omega) = U(\omega)D(\omega)Z^{1D}(\omega) \tag{6}$$

in order to improve convergence and stability where $Z^{1D}(\omega)$ is the 1-D minimum-phase transfer function that is inherently smooth, $D(\omega)$ is the nondimensional distortion function and $U(\omega)$ is the nondimensional correction function. The distortion function $D(\omega)$ and the correction function $U(\omega)$ are both represented by a sum of N Chebyshev polynomials uniformly spaced in log frequency (Larsen *et al.*, submitted).

The smooth representation of $Z(\omega)$ has several benefits: It reduces scatter in the estimates of $Z(\omega)$ caused by noise, it provides a method for predicting electric data from magnetic data, it provides a method for connecting estimates across frequencies where the signals are low, it provides a compact analytical representation, and it converges to a causal minimum-phase transfer function when this is consistent with the data excluding outliers.

5 Robust Smooth Estimates

The construction of the transfer function proceeds by successive iterations starting with a $Z(\omega)$ that is either: a generic 1-D transfer function, a 1-D transfer function from a suitable site, or a 1-D transfer function derived from a GDS transfer function (Larsen, 1980; Schultz and Larsen, 1987). We start with distortion function $D(\omega) = \mathrm{I}$, rotation $\theta = 0$, whitening $W(\omega) = 1$, section weights $s = 1$, and frequency and time weights equal to zero for known periodic variations and missing data and unity otherwise. We also start by replacing all possible magnetic outliers as bias protection. The smooth correction function $U(\omega)$ is then computed by weighted least squares or weighted remote reference and a new estimate of the smooth transfer function is given by $U(\omega)$ times the pre-

vious $Z(\omega)$. The new $Z(\omega)$ and uncorrected data are then used to compute new estimates of frequency residuals $R(\omega)$ and time residuals $r(t)$. These residuals are then used to find electric and magnetic outliers, and to find the whitening $W(\omega)$, frequency weights $F(\omega)$, time weights $g(t)$, and section weights s. The estimated $Z(\omega)$ are then used to compute a new magnetic rotation **0,** a new 1-D transfer function $Z^{1D}(\omega)$, and a new smooth distortion function $D(\omega)$ for each magnetic component so that $D(\omega)Z^{1D}(\omega)$ approximates the estimated transfer function. The process is then repeated by recomputing the correction function $U(\omega)$, the corrected data, the weights, the rotation, the 1-D transfer function $Z^{1D}(\omega)$, and the distortion function $D(\omega)$. The process is stopped when the number of new outliers are less than 0.5% of the total number of data. This usually occurs within six iterations. If $Z^{1D}(\omega)$ fits the data within the error bars then $D(\omega) \approx 1$ and $U(\omega) \approx$ **1.** If $D(\omega)Z^{1D}(\omega)$ fits the data within the error bars then $U(\omega) \approx 1$.

6 Electric Trains

Electric trains in Italy are powered by an overhead 3000 d.c. voltage line generated by a series of power substations approximately 8 km apart supplying currents up to 1500 amperes (Linington, 1974). The current flows through the overhead contact wire (5 meters above the tracks), through the electric train motors via the pantograph touching the contact wire, into the train tracks via the wheels, and then back to the power stations via the grounded train tracks. The tracks are grounded for safety reasons. Each pair of substations spanning a particular train supplies a current proportional to the distance of the train from the other substation (Linington, 1974). Signals vary in intensity and frequency with position of the trains, amount of current used by the trains and harmonics of the basic 50 Hz power grid supplying the rectified 3000 d.c. volts.

There are two modes of train contamination. One mode is due to electric currents leaking into the earth via the grounded train tracks. This causes some of the current to return to the substations via the earth. The other mode is due to a magnetic dipole effect caused by currents flowing through the overhead voltage line and returning through the tracks. **A** simple calculation shows that this latter mode rapidly decays with distance (Linington, 1974) so that sites more than 1 km from the tracks are unlikely to be affected by the magnetic dipole effect. Leakage currents are therefore the dominant mode and this is likely to be the case for most cultural noise.

A review of the extensive literature on cultural electromagnetic signals including electric train induced signals is given in Szarka (1988).

The most obvious method for dealing with electric train signals is to record the electromagnetic data when the trains are not running or running infrequently, or to ignore data sections and frequencies most contaminated by electric train signals.

We propose a different method that we call the multi-source method which is based on a suggestion by Madden (1993, personal communication) to record magnetic data at a site known to be free of electric train signals while simultaneously recording electromagnetic data at sites contaminated by train signals. The clean site is used to estimate the MT and train magnetic data and these are then used to simultaneously estimate the MT and train transfer functions. This method is necessary for extracting the MT transfer function from data collected in the Larderello geothermal area of central Italy for periods from 1s to 230s, the longest period observed.

7 Artificial Data

We tested the robust smooth method using the north hourly Tucson magnetic data **as** the magnetic data. Electric data were then generated from the magnetic data using a 1-D transfer function estimated for Tucson multiplied by a distortion D to make the transfer function distinctly non 1-D to verify that the robust estimate converges to the true non 1-D transfer function (Fig. 1). The data were subdivided into 50 sections and a 4π prolate data window was used giving 45% overlapping data sections of 3456 values. The first four Fourier terms are excluded giving 2.54 decades of frequencies. The number of iterations for convergence was five.

For our first test we used noise free electric and magnetic data. The misfit between smooth estimates and true transfer function was 9.1% for N = 8 coefficients and 0.3% for N = 12 coefficients (Fig. 1). N = 12 corresponds to 4.7 coefficients per decade of frequency. Therefore N, the number of coefficients for the correction and distortion functions, should probably be greater than $4.5\log(\omega_{max}/\omega_{min})$ but this value will be dependent, of course, on how much structure there is in the true transfer function.

To test the robust least squares, we added, to the electric data, gaussian noise that was 11%of the spectral level of the electric data but kept the magnetic data noise free.

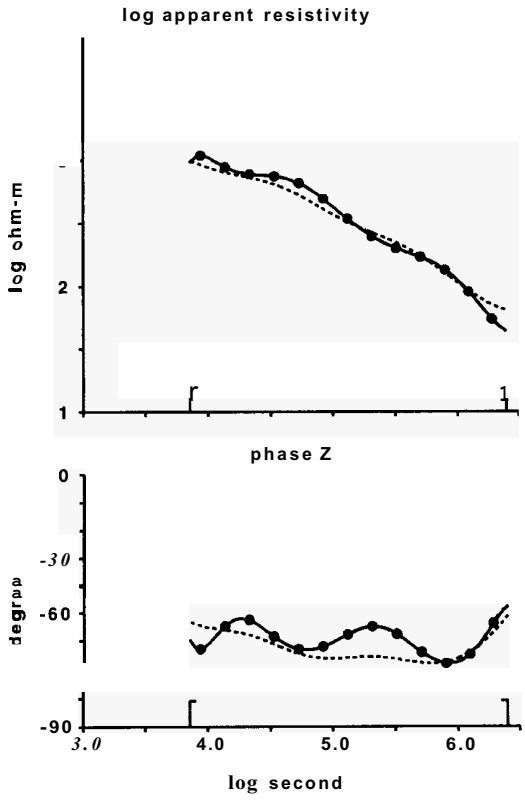

Figure 1: MT transfer functions for noise-free test data. Plotted are smooth estimates (dashed curves), band average estimates (dots), and best-fitting 1-D transfer function (dotted curves). The test electric data are derived from the Tucson magnetic data for a specified transfer function (solid curve), noise, and outliers.

This gives a theoretical coherence square of 0.91. To simulate uncorrelated outliers, we contaminated the first half of the electric data by 5% spikes at random times with random amplitudes between f5 times the standard deviation of the electric data and 5 periodic variations at random frequencies with amplitudes 5 times the background continuum. To simulate correlated outliers, we contaminated the first half of the electric and magnetic data by the same noise consisting of 5% spikes at random times with random amplitudes between f 5 times the standard deviation of the data and 5 periodic variations at randomly chosen frequencies with amplitude 5 times the background continuum.

The misfits between the least squares estimates and the true transfer function were: 2.1% for smooth estimates using weighted and corrected data; 6.9% for band averages using weighted and corrected data; 10% for band averages using unweighted but corrected data; and **36%** for band averages using unweighted and uncorrected data (Fig. 2). In these figures, the error estimates are the 95% confidence limits found by jacknife analysis of the covariance of the solution .The coherence squared for the smooth estimates was 0.90 which is close to the 0.91 theoretical limit. The robust method identified 4.6% electric outliers (5% given) and 1.8% magnetic outliers (2.5% given). The remaining outliers were mostly small. Thus the robust least squares method performed quite well in identifying and replacing outliers and producing a smooth estimate that fit the true transfer function (Fig. *2)*.

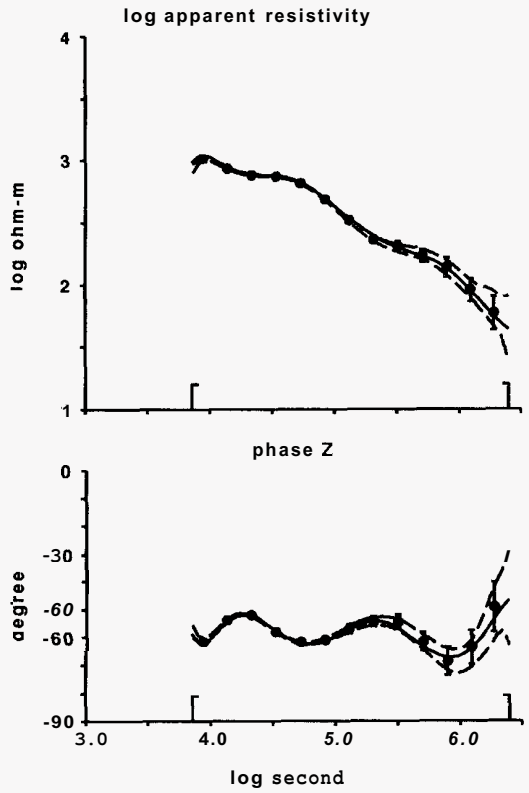

Figure 2: MT transfer functions for test data with noise as described in text. Smooth estimates (dashed curves) are 95% confidence limits and band average estimates (dots with vertical lines) are 95% confidence limits. The test electric data are derived from the Tucson magnetic data for a specified transfer function (solid curve), noise, and outliers.

8 Larderello Data

The Lardarello MT survey carried out in central Italy consisted of a north and south survey line that ran east-west (Fig. **3).** The survey area is close to several electric train tracks, in particular, one along the west coast of Italy that is west and south of the survey, one between Cecina and Volterra that is north of the survey, and one between Emploi, Siena, and further south that is east of the survey. Any particular train draws current from two substations closest to the train and this current is the source for the train induced correlated electric and magnetic signals. The remote magnetic site was situated on the offshore island, Capraia, some 55 km from the coast (Fig. **3)** and free of electric trains.

We used the NO1 site that is at the western end of the northern survey line because it is closest to the north-south train track and therefore has the strongest train induced signals. The NO1 data should therefore provide a rigorous test of the robust method used to extract the MT transfer functions from highly contaminated data. We used 100 data sections, N = 10 number of coefficients, a data window with 5% of the end values cosine tapered, and concentrated on the period range 0.17 s to 7.1 s. For this period range, the train signals were found to be dominant at the long periods and tended to occur in steps. The data were therefore first differenced to suppress the long periods and to localize the train signals in time. We found this improved the transfer functions estimates.

The MT and train transfer functions for the $85°E$ major electric data are well determined (Fig. **4)** with partial coherence squared of 0.51 for the MT signals and 0.51 for the train signals. There were 6.8% electric outliers and **3.4%** magnetic outliers. The MT transfer function for the $5°W$ minor electric data (Fig. 5) is less well determined due to the larger train signals parallel to the tracks with partial coherence squared of 0.05 for the MT signals and 0.90 for the train signals. There were 4.0% electric outliers and 1.9% magnetic outliers.

The MT transfer function for both electric components are found to be essentially minimum-phase (Figs. **4** and 5). In addition, the MT transfer function for the $85°E$ electric data (approximately normal to the coast line) shows a characteristic low phase and a linear rise toward long periods for log amplitude (Fig. **4)** similar to the observed TM mode (electric currents normal to coastline) for the MT transfer function at continental sites (Park et *al.*, 1991). These features in the TM mode are due to an excess of electric currents in the surface continental layer near the coast due to the currents induced in the conductive ocean. The train transfer functions for both components show a linear rise toward long periods in the log amplitude and a constant small phase. They approximate the transfer function for a high wavenumber source and indicate that the train signals induce electric currents mainly in the surface layer. Hence, the train transfer functions and the TM mode of the MT transfer functions have similar frequency characteristics. On the other hand, the log amplitude of the transfer function for the $5°W$ electric data (Fig. 5) is the TE mode (electric current parallel to coastline). It is much flatter in log amplitude with larger phases than the TM mode. This interpretation of the TM and TE modes of the MT transfer function and the train transfer functions indicates that the estimated smooth transfer functions are realistic.

More details on the analysis and interpretation of the MT data taken in the Larderello area can be found in our companion paper in this volume (Fiordelisi *et al.*, 1995)

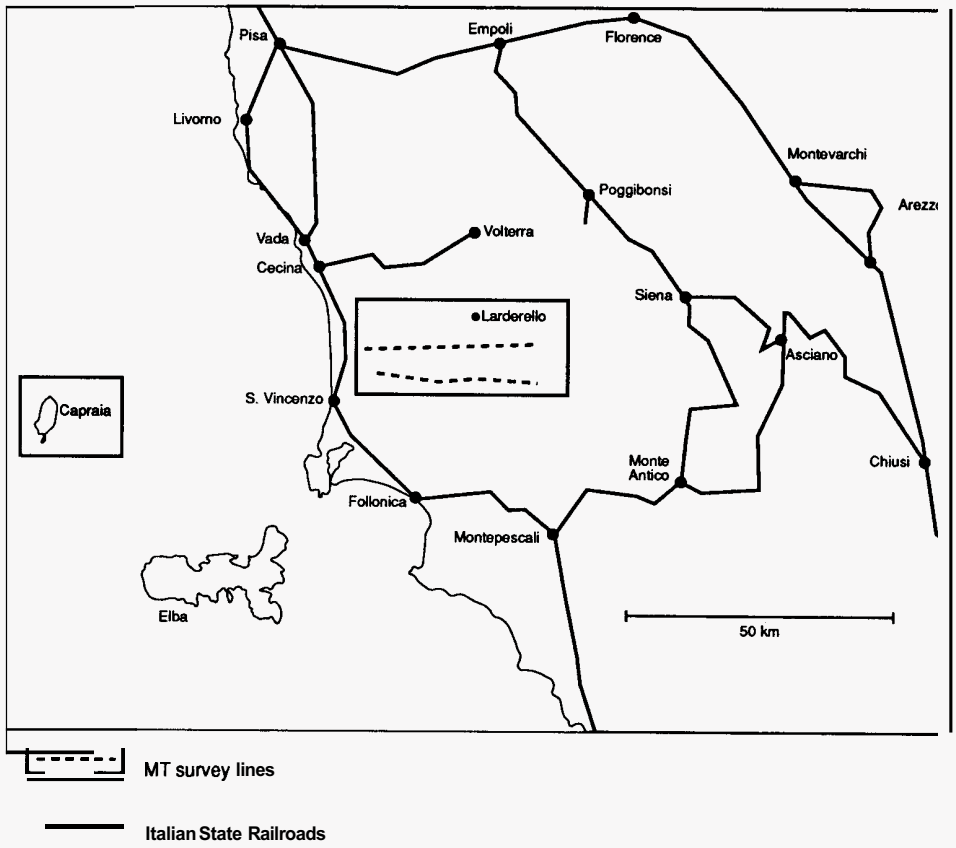

Figure 3: A map showing the Larderello MT survey lines in relation to the railways (heavy lines) and the remote island site of Capraia.

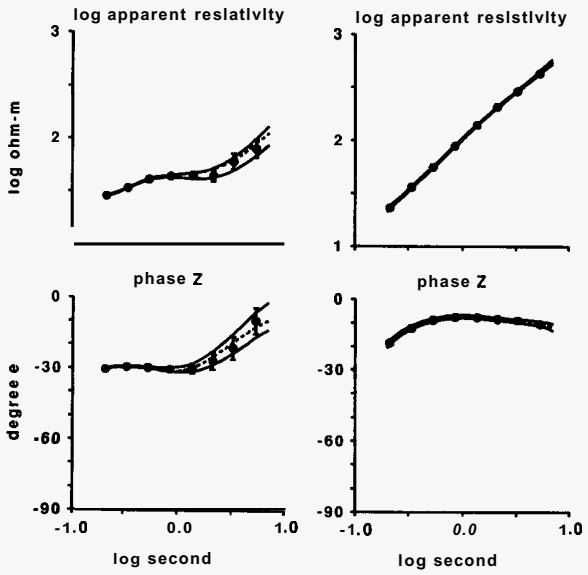

Figure 4: MT (left) and train (right) transfer functions for Larderello N01 site for the electric field direction of 85 deg E. Smooth estimates (pair of solid curves) are 95% confidence limits and band average estimates (dots with vertical lines) are 95% confidence limits and the dotted line is the best fitting 1-D transfer function.

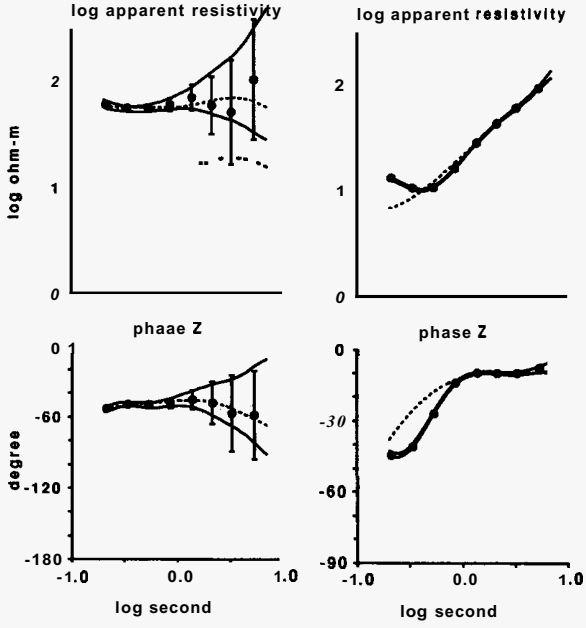

Figure 5: MT (left) and train (right) transfer functions for Larderello N01 site for the electric field direction of 5 deg W. Smooth estimates (pair of solid curves) are 95% confidence limits and band average estimates (dots with vertical lines) are 95% confidence limits and the dotted line is the best fitting 1-D transfer function.

9 Conclusions

The method provides three different estimates of the transfer function: (1) Smooth estimates using weighted and corrected data. They provide a compact representation of the transfer function for later use, can be generated for any particular frequency, and are used to identify individual outliers. (2) Band averages using weighted and corrected data. They provide estimates that are independent of the damping needed to stabilize the smooth estimates. (3) Band averages using unweighted but corrected data. Although they generally have much larger errors, they provide a check on whether there could be biasing effects by the frequency and times weights.

The present robust method is the first attempt to correct the magnetic data for outliers. Test with correlated noise confirms that the robust method is able to locate and remove intermittent outliers and can provide realistic estimates of the transfer functions. Robust methods based on section weighting did not remove the outliers and therefore yielded biased estimates of the transfer function.

The MT transfer functions can be extracted from data highly contaminated by electric train signals by using a remote magnetic site free of train signals. The method provides realistic robust estimates of the MT transfer function even when the MT signals are only 5% of the total variance.

Rotation of the horizontal magnetic components reduces collinearity and distortion in the phase for the major magnetic data.

Acknowledgments

Sincere appreciation goes to ENEL for providing the magnetotelluric data used in this study.

References

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

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

Egbert, G.D., Booker, J. R. and Schultz, A. (1992). Very long period magnetotellurics at Tucson Observatory: Estimation of impedances, *J. Geophys. Res.,* **97,** 15,113-15,128.

Fiordelisi, A., Mackie, R.L., Madden, T.R., Manzella, **A.** and Rieven, S.A. (1995). Application of the magnetotelluric method using a remote-remote reference system for characterizing deep geothermal systems, *World Geothermal Conference 1995.*

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

Larsen, J.C. (1980). Electromagnetic response functions from interrupted and noisy data, *J. Geomng. Geoelectr.,* **32 SI,** 89-103.

Larsen, J.C. (1989). Transfer functions: smooth robust estimates by least-squares and remote reference methods, *Geophys. J.,* **99,** 645-663.

Larsen, J.C., Mackie, R.L., Manzella, **A.,** Fiordelisi, A. and Rieven, S. (1995). Smooth MT transfer functions, *submztted to Geophys. J. Int.*

Linington, R.E. (1974). The magnetic disturbances caused by DC electric railways, *Prospezioni Archeologzche,* **9,** 9-20.

Park, S.K., Biasi, G.P., Mackie, R.L. and Madden, T.R. (1991). Magnetotelluric evidence for crustal suture zones bounding the southern Great, Valley, California, *J. Geophys. Res.,* **96,** 353 376.

Rousseeuw, P.J. and Leroy, **A.M.** (1987). *Robust Regression and Outlier Detectzon,* John Wiley & Sons, New York, 329 pp.

Schultz, A. and Larsen, J.C. (1987). On the electrical conductivity of the mid-mantle: I. Calculation of equivalent scalar magnetotelluric response functions, *Geophys. J. R. astr. Soc.*, **88,** 773-761.

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

Szarka, L. (1988). Geophysical aspects of man-made electromagnetic noise in the Earth **A** review, *Surveys in Geophyszcs,* **9**, 287-318.