Transfer Function Estimation for Electromagnetic Induction Studies: Evolution and State-of-the-Art

Gary D. Egbert

College of Oceanic and Atmospheric Sciences, Oregon State University

ABSTRACT

To use magnetotelluric (MT) data to image crustal structure one must first reduce long time series of electric and magnetic fields to transfer functions (TFs) or impedance tensors which vary smoothly with frequency. In many cases neither the natural source signal, nor the dominant noise sources are well modeled as stationary Gaussian processes, so the most obvious least squares approach to MT impedance estimation often fails catastrophically. Developments over the past two decades have improved the situation considerably, and with proper care modern MT surveys can routinely produce stable and repeatable impedance estimates, with meaningful estimates of statistical precision. Two develoments have been key to these improvements: adaptive schemes which automatically downweight or eliminate poor quality data, and the remote reference method, in which data from a second site is used to cancel local noise and improve data weighting. Initial developments in both of these areas largely involved *ad-hoc* schemes, such as coherency sorting for data screening, or simple analogues to least squares for remote reference processing. Subsequently more rigorously justifiable (but operationally quite similar) statistical schemes were introduced. Here I review some important aspects of these developments, including the regression M-estimate for single station impedance estimation, and its extension to the case of remote reference data. I will also consider the application of multivariate statistical methods to analysis of data from small (2-5 station) arrays of MT (or EM profiling) instruments. These methods allow a more full use of all data channels, and can ion some cases lead to significant improvements over estimates based on standard robust remote reference approaches. Finally, I consider jackknife methods for error bar estimation, presenting a comparison of error bars computed by several methods for a series of replicated MT impedance estimates from a pair of semi-permanent MT sites.

INTRODUCTION

The first step in the interpretation of magnetotelluric (MT) data involves estimating 10^2 frequency domain impedances, $\mathbf{Z}(\omega)$, from the raw electric and magnetic field time series $\mathbf{e}(t)$, $\mathbf{h}(t)$ (approx. 10^6 or more raw data/site) [e.g., *Swift*, 1967, *Sims et al.,* 1971]. Superficially, this initial data reduction step is almost trivial. Making the usual MT assumptions that the external sources are spatially uniform, and allowing for noise in the simplest way, **e** and **h** are related in the frequency domain via the linear statistical model

(1) $$\mathbf{e} = \mathbf{Z}\,\mathbf{h} + \varepsilon$$

where ε represents noise. The impedance **Z** can then be estimated quite simply by Fourier transforming the time series, and using linear least squares (LS) to minimize the misfit to equation (1) [*Sims et al*., 1971]. Unfortunately, this simple approach can fail catastrophically for noisy data, producing estimates which are heavily biased or wildly oscillatory [e.g., *Gamble et al*., 1979; *Jones et al*., 1989]. As a consequence a number of refinements to the simple LS approach have been proposed in an effort to guarantee impedance estimates which are useful for subsequent stages in the interpretation process.

The failure of the LS impedance estimate can be traced to two fundamental inadequacies of the simple model of equation (1). First, this linear statistical model is appropriate to the case where noise is restricted to the output, or "predicted" electric field channels, while the input magnetic fields are observed without error. It was recognized long ago [e.g., *Swift*, 1967; *Sims et al*., 1971] that the violation of this assumption would result in the downward bias of estimated impedance amplitudes. To avoid these bias errors the remote reference (RR) method, in which horizontal magnetic fields are recorded simultaneously at a second (remote) site and correlated with the EM fields at the local site, was proposed and developed by *Gamble et al*.[1979] and *Goubau et al*. [1978]. RR MT impedance estimates are often substantially improved (less biased and more stable) over those obtained using the single station LS approach.

Second, the simple LS estimator implicitly assumes a Gaussian distribution for the errors in equation (1). This assumption often fails for MT data due to the non-stationarity of both signal and noise. In essence, it is quite typical to find that some portions of the signal come much closer to satisfying the idealized relationship of equation (1) than others. This can result in a marginal error distribution in the frequency domain which is heavy tailed, or contaminated by outliers [*Egbert and Booker*, 1986; *Chave et al*., 1987; *Chave and Thomson*, 1989]. A number of MT processing methods have been proposed which attempt to exploit this fact by adaptive screening or weighting of the data. Early efforts in this direction were based on some sort of coherence weighted estimates [*Stodt*, 1983; *Jones and Jodicke*, 1984].

THE REGRESSION M-ESTIMATE

The regression M-estimate [RME; c.f., *Huber*, 1981] has been adapted to yield impedance estimates which are robust to violations of distributional assumptions and resistant to outliers [*Egbert and Booker*, 1986; *Chave and Thomson*, 1989; *Chave et al*., 1987; *Larsen*, 1989; *Sutarno and Vozoff*, 1991]. These procedures can be rigorously justified both mathematically, in terms of various technical definitions of optimality [*Huber*, 1981], and in practice, as demonstrated by numerous Monte Carlo studies [*Hampel et al.,* 1986].

The RME can be viewed as a data adaptive weighted least squares estimate, and is thus operationally similar to earlier, *ad hoc* data weighting methods. Data processing begins with a division of the time series into a sequence of short data segments, each of which is Fourier transformed, yielding a series of *I* complex data vectors identified with a fixed frequency ω. The impedance $\mathrm{Z}(\omega)$ is then estimated by minimizing the weighted residual sums of squares. For the *x* component of the electric fields the weighted squared misfit takes the form

(2) $$\sum w_i \left| e_{xi} - (Z_{xx} h_{xi} + Z_{xy} h_{yi}) \right|^2$$

with the weights w_i determined iteratively from the normalized residuals. For example, for the so-called Huber weights used by *Egbert and Booker* [1986] and *Chave et al*. [1987]

(3) $$w_i = \begin{cases} 1 & \text{if } \left| r_i \right| \le 1.5 \\ 1.5 / \left| r_i \right| & \text{if } \left| r_i \right| > 1.5 \end{cases}$$

(4) $$r_i = e_{xi} - (Z_{xx} h_{xi} + Z_{xy} h_{yi}) / \hat{\sigma} \,,$$

where $\hat{\sigma}$ is an estimate of the scale of the typical error. The determination of data weights is an iterative process for the RME. Initial estimates of the impedance can be obtained with unweighted LS. The resulting residuals are then used to determine a set of weights (which reduce the influence of data points which fit poorly). A new impedance estimate is then computed by minimizing the weighted squared misfit of equation (2), and the process is repeated to convergence (which is guaranteed for the Huber weights of equation (3); *Huber* [1981]).

A comparison of a variety of processing methods for long period MT data by *Jones et al*., [1989], demonstrated the overall superiority of methods based on the RME. A similar conclusion was reached in a comparison of processing methods applied to single site wide-band MT data by Egbert and Livelybrooks [19996]. *Chave and Thomson* [1989] proposed a straightforward extension of the RME to the case of remote reference. Standard formulae for calculation of remote reference estimates are used, with the addition of weights determined adaptively from the data exactly as in equation (3). The value of the robust remote reference (RRR) method is demonstrated by the comparisons given in *Jones et al*. [1989].

MULTIPLE STATION PROCESSING METHODS

Although modern MT data are highly multivariate (multiple components, recorded simultaneously at multiple stations) the most commonly used processing methods (including the RME and RRR estimates) are in fact based on univariate statistical procedures. *Egbert and Booker* [1989] described a general framework for processing data from arrays of simultaneously recording EM instruments based on classical methods of multivariate statistical analysis. Based on this multivariate approach, *Egbert* [1997] developed a practical robust processing scheme for small (2-5 station) arrays of MT and EM profiling sites. With this multivariate approach data from all channels can be used to improve signal-to-noise ratios, and to diagnose possible biases due to coherent noise. The possibility of noise (and outliers!) in all data channels is explicitly allowed for with this approach.

To allow for the possibility of coherent noise *Egbert* [1997] developed a two stage procedure. In the first stage the amplitude of the incoherent noise in each data channel is estimated and isolated (single channel) outliers are cleaned up. With all channels scaled into the non-dimensional units defined by incoherent noise amplitudes, the "coherence dimension" *M* of the array data (i.e., the number of distinct sources of coherent signal/noise which can be resolved by the array) can be estimated. This is accomplished by examination of the eigenvalues of the normalized spectral density matrix (i.e., the matrix of averaged cross-products of all data channels, scaled into non-dimensional signal-to-noise units). When there is no coherent noise, *M* will be two, corresponding to the two orthogonal plane wave MT sources. In this case the two dimensional coherent part of the array signal can be used to directly estimate MT impedances and inter-station transfer functions.

In contrast to the standard single station or RR approach, where a pair of data channels at one site is chosen (often rather arbitrarily) as a reference, all available channels are used to define the desired signal with this approach. This can improve signal-to-noise ratios and enhance detection of outliers restricted to one (or a few) channels. Using real and synthetic data, Egbert [1997] shows that RMEV estimates can be significantly better than those obtained with a more standard RRR estimator, particularly when outliers at the remote site severely contaminate remote reference estimates. In this case, the RMEV estimator, which treats all channels symmetrically can easily find and remove the contaminated sections.

If we find evidence for significant coherent noise (coherence dimension *M* > 2) we proceed to the second stage and attempt to separate coherent noise from the desired MT signal. This may often be difficult or impossible, but a careful array analysis can at least help to define the nature and extent of coherent noise problems. Furthermore in some cases this information allows us to choose sites, time segments and/or period ranges for which contamination by coherent noise is minimal, thereby allowing at least an approximate separation of coherent noise and signal. Egbert (1997) gives an example of two MT sites south of San Francisco California which are severely contaminated by coherent cultural noise (DC electric railways almost 100 km to the north). With the aid of diagnostics based on our multivariate analysis, time periods with minimal contamination could be selected, leading to dramatic improvements in MT apparent resistivity and phase estimates.

JACKKNIFE METHODS FOR COMPUTATION OF ESTIMATION ERRORS

Chave and Thomson [1989] first proposed the Jackknife method for computation of estimation errors regression M-estimates of MT impedances. The jackknife scheme [e.g., *Efron,* 1982] is conceptually quite simple, but the scheme can increase the overall computational burden an extreme amount. Each data point is omitted in turn, and the full estimation procedure is run on the remaining *I-1* data. The result is a set of *I* separate estimates $\hat{\mathbf{Z}}_{(i)}$ of the iimpedance tensor. It can be shown that the total deviation between these estimates $\sum(Z_{(i)} - \bar{Z})(Z_{(i)} - \bar{Z})^*$ is a good approximation to the statistical uncertainty in the impedance estimates. The primary advantage of the jackknife over standard asymptotic estimation procedures (as used, for example, by *Egbert and Booker*, [1986]), is that it accounts for the non-linearity of the estimator in a simple fashion. (Note that the RME is non-linear--the weights which multiply the data also depend on the data). However to literally implement the jackknife in this case one must repeat the full iterative robust calculation once for each data point. Since the total number of data segments collected at a wide-band MT site often exceeds 5000 or more, the increase in computational cost required by the jackknife can be extreme.

To make the jackknife more practical two possible modifications can be considered. *Chave and Thomson* [1989] suggested computing the weights for the RME once, then applying the jackknife to the weighted least squares problem (with the final weights fixed). This leads to a much more efficient estimation scheme, since the delete one estimates can be computed rapidly by applying row update methods to solve the required set of *I* LS problems with minimal additional computation. Indeed, it turns out that even for the remote reference case this "fixed weight" jackknife can be computed essentially for free; a closed form expression for the variance can be given [*Eisel and Egbert*, 1999]. Unfortunately, with the weights fixed the estimation problem becomes linear, and the primary justification for the jackknife disappears. However, as shown by *Hinkley* [1977], there is a second advantage to the jackknife, which applies even to the case of linear regression: the jacknifed error estimates are robust to violations of the assumption that the data error variances do not depend systematically on the signal power. Thus in some cases the simplified jackknife of Chave and Thomson [1989] may still be useful. A second way to make the Jackknife approach practical is to divide the total data set into a small number of subsets (e.g., 20 subsets each containing 5% of the data) and delete each subset in turn. With a small number of subsets it becomes morepractical to repeat the full non-linear RME procedure for each deleted set.

Eisel and Egbert (1999) used data from a continuously operating MT site from the Parkfield, California earthquake prediction experiment to compare error bars computed with several approaches, including the standard asymptotic approach, a full literal jackknife procedure, a subset deletion jackknife procedure, and the fixed weight jackknife procedure. For this comparison separate estimates were computed for each day in a two year period, and the deviations of the daily estimates, and their error bars, were compared to the long term mean impedance estimated from all data. This data was contaminated with low level coherent noise, which lead to small systematic biases on some days at periods around 20s. The standard asymptotic RME error estimates were small by a factor of more than two at this period. All of the jackknife approaches produced larger error bars, and were thus more realistic in this band. However, at other periods where the standard approach yielded reasonable estimates, the jackknife methods all overestimated error bars. The fixed weight scheme of Chave and Thomson [1997] was least effective at accounting for increases in error bars in the contaminated 20s band. Subset schemes with relatively large subsets (at least 5%) provide the most realistic error estimates.

REFERENCES

Chave, A.D., D.J. Thomson, and M.E. Ander, 1987, On the robust estimation of power spectra, coherences, and transfer functions, Journal of Geophysical Research, 92, 633-648.

Efron, B., 1982. The jackknife, the bootstrap and other resampling methods, Society for Industrial and Applied Mathematics, Philadelphia.

Egbert, G.D. and Booker, J.R. 1986, Robust estimation of geomagnetic transfer functions, Geophysical Journal of the Royal Astronomical Society, 87:173-194.

Egbert, G.D. and J.R. Booker, 1989, Multivariate analysis of geomagnetic array data I: The response space, Journal of Geophysical Research, 94, 14,227-14,248.

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

Egbert, G.D., 1997, Robust multiple station magnetotelluric data processing, Geophysical Journal International, 130, 475-496.

Eisel, M.E., and Egbert G.D., 1999, On the stability of magnetotelluric transfer functions and the reliability of their variances, to be submitted to Geophysical Journal International.

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

Goubau, W.M., T.D. Gamble, and J. Clarke, 1978, Magnetotelluric data analysis: removal of bias, Geophysics, 43, 1157-1166.

Hampel, F.R., E.M. Ronchetti, P.J. Rousseeuw, and W.A., Stahel, 1986, Robust Statistics; The Approach Based on Influence Functions, John Wiley, New York.

Hinkley, D.V., 1977, Jackknifing in unbalanced situations, Technometrics, 19, 285-292.

Huber, P., 1981, Robust Statistics, Wiley, New York.

Jones, A.G., and H. Jodicke, 1984, Magnetotelluric transfer function estimation improvement by a coherence-based rejection technique: Presented at 54th Annual International Meeting, Soc. of Expl. Geophys., Atlanta, Ga., Dec. 2-6.

Jones, A. G., A.D. Chave, G.D. Egbert, D. Auld, and K. Bahr, 1989, A comparison of techniques for magnetotelluric response function estimation, Journal of Geophysical Research, 94, 14,201-14,214.

Larsen, J.C., 1989. Transfer functions: smooth robust estimates by least squares and remote reference methods, Geophysical Journal International, 99, 655-663.

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.

Stodt, J.A., 1983, Noise analysis for conventional and remote reference magnetotellurics, Ph.D. dissertation, Univ. of Utah.

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

Swift, C.M., 1967, A magnetotelluric investigation of an electrical conductivity anomaly in the southwestern United States, Ph.D. dissertation, MIT.

Sixth International Congress of the Brazilian Geophysical Society