

## Time-domain estimation of MT impedance tensor

Umberto Spagnolini\*

### ABSTRACT

The spectral analysis of magnetotelluric (MT) data for impedance tensor estimation requires the stationarity of measured magnetic (H) and electric (E) fields. However, it is well known that noise biases time-domain tensor estimates obtained via an iterative search by a descent algorithm to determine the least-mean-square residual between measured and estimated E data obtained from H data. To limit the noise that slows down, or even prevents convergence, the steepest descent step size is based upon the statistics of the residual (Bayes' estimation). With respect to uncorrelated noise, the time-domain technique is more robust than frequency-domain techniques. Furthermore, the technique requires only short-time stationarity.

The time-domain technique is applied to data sets (Lincoln Line sites) from the EMSLAB Juan de Fuca project (Electromagnetic Sounding of the Lithosphere and Asthenosphere Beneath the Juan de Fuca Plate), as well as to data from a southern Italian site. The results of EMSLAB data analysis are comparable to those obtained by robust remote reference processing where larger data sets were used.

### INTRODUCTION

Measurements of electric and magnetic telluric fields on the surface of the earth can be used to determine the electrical resistivity distribution in the subsurface. For an anisotropic or a laterally inhomogeneous earth model the relation between the electric (E) and magnetic (H) fields is given by the impedance tensor. The fundamental relationship of the magnetotelluric (MT) method (at a given angular frequency  $\omega$ )

$$\begin{bmatrix} E_x(\omega) \\ E_y(\omega) \end{bmatrix} = \begin{bmatrix} Z_{xx}(\omega) & Z_{xy}(\omega) \\ Z_{yx}(\omega) & Z_{yy}(\omega) \end{bmatrix} \begin{bmatrix} H_x(\omega) \\ H_y(\omega) \end{bmatrix}, \quad (1)$$

was first proposed in the frequency domain in Neves (1956) and Cantwell (1960). Estimation of the matrix transfer function (impedance tensor)

$$\mathbf{Z}(\omega) = \begin{bmatrix} Z_{xx}(\omega) & Z_{xy}(\omega) \\ Z_{yx}(\omega) & Z_{yy}(\omega) \end{bmatrix} \quad (2)$$

from the measured E field  $\mathbf{E}(\omega) = [E_x(\omega) \ E_y(\omega)]^T$  and H field  $\mathbf{H}(\omega) = [H_x(\omega) \ H_y(\omega)]^T$  ( $[\cdot]^T$  indicates the matrix transposition) can be carried out in the frequency domain as proposed in Swift (1967) by their orthogonal horizontal field components ( $E_x(\omega)$ ,  $E_y(\omega)$ ,  $H_x(\omega)$ ,  $H_y(\omega)$ ). The use of the frequency-domain technique means that long-time stationarity of the measured time series is necessary for a reliable estimate of the auto-power and cross-power. In fact, noise in the H and E data biases the estimated tensor elements. Sims et al. (1971) have proposed a way to quantify severe noise bias and thus select the least biased estimator.

The time-domain approach to impedance tensor estimation is less sensitive to noise produced by spikes, as shown in McMechan and Barrsdale (1985), but requires the inversion of large matrices for the resolution of low frequencies. Similarly, in robust processing, Larsen (1989) demonstrated the efficacy of time weights for the reduction of bias resulting from noise.

The reference-model approach is one of the simplest and most flexible methods for impedance tensor estimation. The H measurements are used as an input for a model whose parameters are then adjusted to reduce the error between the E data simulated from the model and the measured E data. If the initial model is reasonable and the algorithm for adjusting the model is properly designed, the residual vanishes and the model converges to an equivalent representation of the unknown impedance tensor. For instance, an autoregressive moving-average (ARMA) model was chosen by Yee et al. (1988) to estimate the impedance tensor. The computational costs of the adaptive approach to tensor estimation are lower than those for matrix inversion with the advantage that the estimate can be obtained while the data are being collected.

The present paper describes an adaptive time-domain estimation of the impedance tensor from measured H and E

Manuscript received by the Editor March 9, 1992; revised manuscript received May 14, 1993.

\*Dipartimento di Elettronica e Informazione, Politecnico di Milano, Piazza Leonardo da Vinci 32, I-20133 Milano, Italy.

© 1994 Society of Exploration Geophysicists. All rights reserved.

data. Owing to the flexibility of both the adaptive approach and the selected model (i.e., finite-length impulse response), this estimation technique seems quite appropriate for dealing with the short time stationarity of real data.

After a general overview of the time-domain technique, a discussion follows on the consequences of data correlation in impedance tensor estimation. Adaptive estimation uses an iterative search for the least-mean-square (LMS) solution in the time domain that minimizes the mean-square (MS) error between measured and simulated E data. Therefore, Bayes' estimation of the mismatch in the reference model permits an additional constraint on its adaptation. Several factors may slow down or even prevent convergence of the LMS approximate solution during the iterative search. This problem is examined and practical solutions are proposed.

As illustrations, both synthetic and real data examples are analyzed, and time- and frequency-domain techniques are compared. The results of the single-site adaptive estimation of EMSLAB Lincoln Line sites data are compared with the robust frequency-estimation technique presented in Jones et al. (1989) and obtained using the remote reference data. Finally, a time-domain remote reference processing is proposed and the results are compared to those of frequency-domain, remote-reference processing.

### TIME-DOMAIN IMPEDANCE TENSOR ESTIMATION

The impedance tensor relation (1) can be equivalently expressed in the time domain (t) where the convolutional relation becomes

$$\begin{bmatrix} e_x(t) \\ e_y(t) \end{bmatrix} = \begin{bmatrix} z_{xx}(t) & z_{xy}(t) \\ z_{yx}(t) & z_{yy}(t) \end{bmatrix} * \begin{bmatrix} h_x(t) \\ h_y(t) \end{bmatrix} + \begin{bmatrix} \epsilon_{e_x}(t) \\ \epsilon_{e_y}(t) \end{bmatrix} \quad (3)$$

(lower case indicates the corresponding time-domain orthogonal components discussed in the relation (1) and  $\epsilon_{e_x}(t)$  and  $\epsilon_{e_y}(t)$  are the noise components in the measured E data) or can be represented more concisely using matrix notation

$$\mathbf{E}(t) = \mathbf{Z}(t) * \mathbf{H}(t) + \mathbf{N}_E(t). \quad (4)$$

Here it is assumed that the additive noise  $N_E(t)$  is uncorrelated with the E data  $E(t)$  and that  $H(t)$  is noise free. The time-domain approach estimates the impulse or pulse response (PR) matrix  $Z(t)$  directly from the time data series. The impedance tensor in the frequency domain  $\mathbf{Z}(\omega)$  is obtained from the Fourier transform of the elements of PR matrix  $Z(t)$ .

The impedance tensor estimation corresponds to a decomposition of the E data into two components: one component is a linear combination of the H data and gives the impedance tensor estimate, and the other is uncorrelated to the H data and yields the noise  $N_E(t)$ . Given the symmetry of each term in equation (3), and providing both measured H field components and one E field component are known, the estimation of the impedance tensor elements can be split into the estimates of the two terms. In the flow chart in Figure 1, there are two input and one output linear subsystems. Because of the symmetry between  $e_x(t)$  and  $e_y(t)$ , the analysis will be conducted on the estimation of the sampled PRs  $z_{yy}(i)$  and  $z_{yx}(i)$  using all or part of the  $e_y(t)$  time

samples, here indicated as a vector  $\mathbf{E}_y$  (the estimation of PRs  $z_{xx}(i)$  and  $z_{xy}(i)$  requires the time series  $\mathbf{E}_x$ ).

The noisy  $\mathbf{E}_y$  time series can be modeled as a linear combination of the measured  $h_x(i)$  and  $h_y(i)$ , and the estimated  $\hat{e}_y(i)$  depends on the PRs

$$\hat{e}_y(i) = \sum_{j=-L/2}^{L/2} z_{yx}(j)h_x(i-j) + \sum_{j=-L/2}^{L/2} z_{yy}(j)h_y(i-j). \quad (5)$$

Hereafter, convolutional relationships like equation (5) will be expressed using the matrix notation

$$\hat{e}_y(i) = \mathbf{Zyx}^T \mathbf{Hx}(i) + \mathbf{Zyy}^T \mathbf{Hy}(i), \quad (6)$$

where the PR vectors are  $\mathbf{Zyy} = [z_{yy}(L/2), \dots, z_{yy}(0), \dots, z_{yy}(-L/2)]^T$ ,  $\mathbf{Zyx} = [z_{yx}(L/2), \dots, z_{yx}(0), \dots, z_{yx}(-L/2)]^T$ , and  $\mathbf{Hx}(i)$ ,  $\mathbf{Hy}(i)$  vectors represent H time series symmetrically windowed around the  $i$ th sample (i.e.,  $\mathbf{Hx}(i) = [h_x(i-L/2), h_x(i-L/2+1), \dots, h_x(i+L/2)]^T$ ). The estimation of the impedance tensor relies upon the identification of those PR vectors  $\mathbf{Zyx}$  and  $\mathbf{Zyy}$  that predict, in the LMS sense, the measured  $\mathbf{E}_y$  data from  $\mathbf{Hx}$  and  $\mathbf{Hy}$ . The MS error  $\overline{\epsilon^2}(\mathbf{Zyx}, \mathbf{Zyy}) = \overline{[e_y(i) - \hat{e}_y(i)]^2}$  in the time domain depends on the PR vectors (the overline indicates the time average). The PRs are obtained when the gradient of the MS error with respect to  $\mathbf{Zyx}$  and  $\mathbf{Zyy}$  is zero;

$$\begin{bmatrix} \overline{\mathbf{HxHx}^T} & \overline{\mathbf{HxHy}^T} \\ \overline{\mathbf{HyHx}^T} & \overline{\mathbf{HyHy}^T} \end{bmatrix} \begin{bmatrix} \mathbf{Zyx} \\ \mathbf{Zyy} \end{bmatrix} = \begin{bmatrix} \overline{\mathbf{EyHx}^T} \\ \overline{\mathbf{EyHy}^T} \end{bmatrix} \quad (7)$$

or, more concisely,

$$\mathbf{R}_{HH} \mathbf{Z} = \mathbf{P}_{EH}, \quad (8)$$

where  $\mathbf{R}_{HH}$  is a  $(2(L+1) \times 2(L+1))$  matrix of H data correlations and  $\mathbf{P}_{EH}$  is the  $(2(L+1) \times 1)$  crosscorrelation vector. The LMS estimate of the impedance tensor elements requires inversion of the symmetric correlation matrix  $\mathbf{R}_{HH}$ .

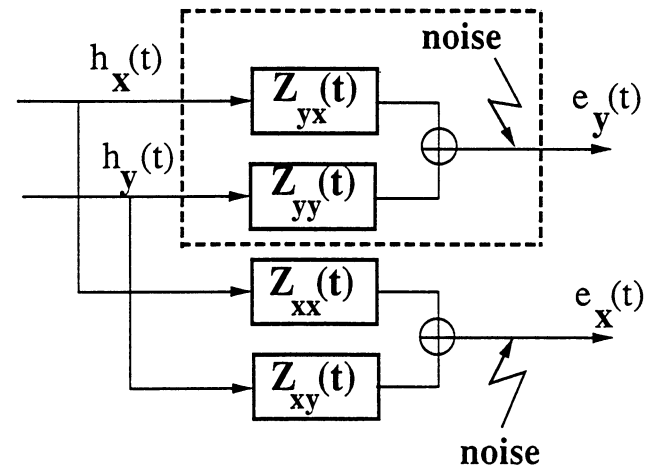


FIG. 1. Block diagram of the model adopted for impedance tensor estimation. The two subsystems denote two symmetrical tensor element estimation problems (e.g., dashed line) in a linear system with two inputs and one output.

Inversion of  $\mathbf{R}_{HH}$  is not always feasible and a full rank correlation matrix is required for a unique solution. To obtain a unique LMS estimate, the H time series can be decomposed into mutually correlated (or polarized) Hx, and Hy, and mutually uncorrelated (or unpolarized) Hx, and Hy, data (Fowler et al., 1967). The correlation matrix  $\mathbf{R}_{HH}$  is now separated into contributions for uncorrelated ( $\mathbf{R}_u$ ) and correlated ( $\mathbf{R}_c$ ) data:

$$\mathbf{R}_{HH} = \mathbf{R}_u + \mathbf{R}_c = \begin{bmatrix} \overline{\mathbf{Hx}_u \mathbf{Hx}_u^T} & \mathbf{0} \\ \mathbf{0} & \overline{\mathbf{Hy}_u \mathbf{Hy}_u^T} \end{bmatrix} + \begin{bmatrix} \overline{\mathbf{Hx}_c \mathbf{Hx}_c^T} & \overline{\mathbf{Hx}_c \mathbf{Hy}_c^T} \\ \overline{\mathbf{Hy}_c \mathbf{Hx}_c^T} & \overline{\mathbf{Hy}_c \mathbf{Hy}_c^T} \end{bmatrix}. \quad (9)$$

Assuming an uncorrelated H time series, the matrix becomes  $\mathbf{R}_{HH} = \mathbf{R}_u$ , and the LMS estimate is given using only one H time series for each PR estimation (1-D PR estimation)

$$\mathbf{Z}_{1D} = \mathbf{R}_u^{-1} \mathbf{P}_{EH}, \quad (10)$$

provided that the correlation matrix is full rank. Generally, for full-rank correlation matrices, the LMS estimate becomes

$$\mathbf{Z} = [\mathbf{I} - \mathbf{R}_{HH}^{-1} \mathbf{R}_c] \mathbf{Z}_{1D}. \quad (11)$$

The correlation matrix may become ill-conditioned because of correlated data. Under this condition the solution contains some nonobservable combinations of Z. In other words, the null space of the solution caused by the correlation of the H time series coincides with the nonobservable subspace of the LMS solution. For instance, the LMS estimate for the trivial case when Hx = Hy leads to the solution

$$\mathbf{Z}_{yx} + \mathbf{Z}_{yy} = \overline{\mathbf{HxHx}^T}^{-1} \overline{\mathbf{EyHx}^T}, \quad (12)$$

that is undetermined unless another constraint is available (i.e., the data are gathered from a well-known 1-D structure so that  $|\mathbf{Z}_{yx}| > |\mathbf{Z}_{yy}|$ ).

The time-domain approach generally requires estimates of  $\mathbf{R}_{HH}$  and  $\mathbf{P}_{EH}$ . The inversion of the correlation matrix  $\mathbf{R}_{HH}$  is computationally expensive and could be unstable even if it is computed only once for the four components of the MT tensor. The time-domain approach leads to the same solution uncertainty (e.g., because of H data correlation) as in frequency-domain impedance tensor estimation (Sims et al., 1971). However, unbiased correlation matrix estimates are required for short time stationary data. This means that several LMS estimates of the impedance tensor could be obtained for overlapped time windows and the estimated impedance tensor should, in some way, be averaged to reduce the bias effects caused by noise and data nonstationarity.

### ADAPTIVE TIME DOMAIN ESTIMATION

The adaptive approach to time-domain PR impedance tensor estimation is more robust and computationally less expensive than the time-domain PR estimation previously

shown. As adaptive PR estimation simultaneously performs the indirect  $\mathbf{R}_{HH}$  estimation and the search for the minimum of the MS error function  $\epsilon^2(\mathbf{Z})$ , the H data correlation still limits the reliability of the estimated PRs. In adaptive impedance tensor estimates, the MS error function is estimated from the data by optimizing the PR vectors toward the LMS solution.

Assuming that the MS error is known, an iterative search for the minimum can be performed by the steepest descent algorithm. This iterative technique is one way to compute the inverse of the correlation matrix  $\mathbf{R}_{HH}^{-1}$ . The adaptive approach is based on the iterative estimate of the LMS PR vectors from the gradient

$$\mathbf{Z}_{i+1} = \mathbf{Z}_i - \mu \frac{\partial \overline{\epsilon^2}(\mathbf{Z}_i)}{\partial \mathbf{Z}}, \quad (13)$$

where  $\mu$  scales the step size of the descent algorithm. The MS error can be approximated by the local error (Widrow et al., 1976)

$$\overline{\epsilon^2}(\mathbf{Z}_i) \approx [e_y(i) - \hat{e}_y(i)]^2, \quad (14)$$

so that the steepest descent iteration (13) and data from the model (6) allow the modification of the algorithm into an adaptive search [one iteration per time sample  $e_y(i)$ ], using (Figure 2)

$$\mathbf{Z}_{i+1} = \mathbf{Z}_i + 2\mu [e_y(i) - \hat{e}_y(i)] \begin{bmatrix} \mathbf{Hx}(i) \\ \mathbf{Hy}(i) \end{bmatrix}. \quad (15)$$

In the adaptive approximation (15), the scale  $\mu$  regulates the convergence and the stability of the adaptation by integrating the effects of the errors  $[e_y(i) - \hat{e}_y(i)]$  over the samples. When convergence has been reached, only noise in the E data randomly moves the estimated PR vectors. One of the most interesting properties of the adaptive impedance tensor

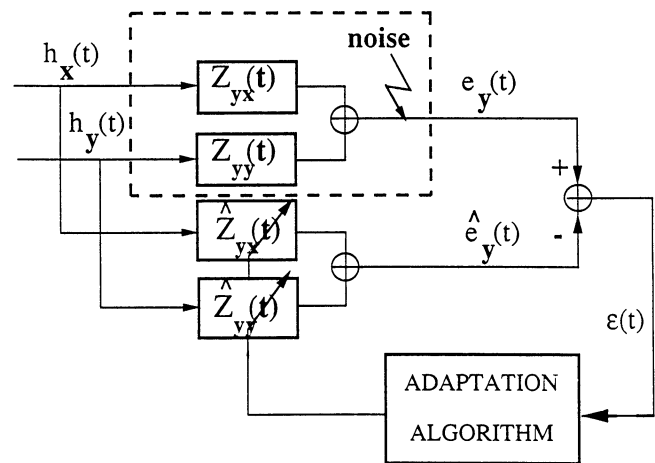


FIG. 2. Block diagram of the adaptive time-domain technique for impedance tensor estimation. Dashed line represents one of the two symmetric unknown subsystems derived from the model in Figure 1. The adaptation algorithm adjusts the estimated PR vectors  $\hat{\mathbf{Z}}_{yx}$ ,  $\hat{\mathbf{Z}}_{yy}$  so as to reduce the MS error  $\epsilon^2(t)$  between the model and the measured E data, given the same H measurements.

estimate is the ability to operate in a nonstationary environment. Because of the stationarity of true PRs (i.e., the earth is stationary), the update of the PR vectors  $\mathbf{Zy}_x$ , and  $\mathbf{Zy}_y$  is less sensitive to short time nonstationarity in the time series. However, stationarity is required in the short time window of the PR vector length.

Measured H and E data are noisy; therefore, the step-size scaling  $\mu$  should be small to reduce the influence of noise on the tensor estimation. The adaptive search method (15) can be very sensitive to spikes as well as to step-like noise, both common in real data. A large isolated spike in the measured data  $e_y(i)$  can affect the estimated tensor because it strongly moves the PR vectors  $\mathbf{Zy}_x$  and  $\mathbf{Zy}_y$  even if a small value of  $\mu$  has been chosen. Removal of spikes and step noise should be carried out in the adaptation loop to prevent noisy convergence (e.g., the PR vectors show large rapid variations). Since the adaptive estimation of PRs is basically a way to evaluate the LMS solution without the computationally expensive estimation and inversion of the correlation matrix  $\mathbf{R}_{HH}$ , the main advantage lies in the way the adaptive approach limits the influence of noise in PR estimation.

### Nonlinear adaptation

The adaptation algorithm (15) interprets noise in the E data as a mismatch between the unknown and the estimated tensor so that the updating of the PR vectors could lead to difficult convergence. In the presence of strong noise, the adaptation becomes unreliable and the true mismatch should be estimated, or the step size  $\mu$  reduced.

During the adaptation, the estimated E data of each channel

$$\hat{e}(t) = e_0(t) + \varepsilon_A(t) \quad (16)$$

differs from the “true” data  $e_0(t)$  because of an incomplete adaptation by  $\varepsilon_A(t)$ . However, it has been assumed in the convolutional relationship (3) that the measured E data are affected by noise  $\varepsilon_N(t)$ . Thus the estimation mismatch that updates the PR vector in the adaptation loop (15)

$$\varepsilon(t) = \varepsilon_A(t) + \varepsilon_N(t) \quad (17)$$

depends on both the estimate of the incomplete adaptation  $\varepsilon_A(t)$  and the E noise  $\varepsilon_N(t)$ , while the adaptation should be performed using the  $\hat{e}(t)$  only. The estimator of the adaptation error is obtained from a nonlinear function  $g[\cdot]$  of the overall error  $\varepsilon$  such that the mean-square error

$$E\{\varepsilon_A - g[\varepsilon]\}^2 \quad (18)$$

is minimized (for the sake of simplicity the time dependence is understood). Provided that the sequences  $\varepsilon_A$  and  $\varepsilon_N$  are independent random variables, the nonlinearity  $g[\cdot]$  is estimated using Bayes’ criterion (i.e., the mean-square error estimator of  $\varepsilon_A$  is the conditional mean  $E\{\varepsilon_A | \varepsilon\}$ )

$$g[\varepsilon] = E\{\varepsilon_A | \varepsilon\} = \frac{\int \varepsilon_A f_A(\varepsilon_A) f_N(\varepsilon - \varepsilon_A) d\varepsilon_A}{f_\varepsilon(\varepsilon)} \quad (19)$$

that depends on the probability density function (pdf) of the adaptation error  $f_A$  and the pdf of the E noise  $f_N$  (the error

pdf  $f_\varepsilon$  is derived). For the purpose of obtaining the nonlinear function  $g[\varepsilon]$ , realistic pdfs should be specified.

The noise pdf  $f_N$  is assumed to be a result of a superposition of Gaussian noise and spikes. The spikes occur with probability  $p$  and, for simplicity, have a Gaussian pdf with variance  $\sigma_p^2$ . The noise has variance  $\sigma_n^2$  and occurs with probability  $(1 - p) \gg p$ .

The pdf that describes the random variables  $\varepsilon_N$  is the “normal mixture”

$$f_N(\varepsilon_N) = (1 - p)G(\varepsilon_N; \sigma_n^2) + pG(\varepsilon_N; \sigma_p^2), \quad (20)$$

where  $G(\cdot)$  indicates the zero-mean Gaussian pdf. The probability  $p$  describes the occurrence of strong spikes with larger tails (i.e.,  $\sigma_p^2 \gg \sigma_n^2$ ) in a background noise. It seems reasonable to assume a Gaussian pdf of the adaptation error with variance  $\sigma_A^2$ :  $f_A(\varepsilon_A) = G(\varepsilon_A; \sigma_A^2)$ . The analytic solution of the nonlinearity  $g[\varepsilon]$  is obtained from the relationship (19), but for the Gaussian  $\varepsilon_A$  it depends on the first derivative of the error pdf  $[f'_\varepsilon(\varepsilon)]$

$$E\{\varepsilon_A | \varepsilon\} = -\sigma_A^2 \frac{f'_\varepsilon(\varepsilon)}{f_\varepsilon(\varepsilon)}. \quad (21)$$

A simplification of the noise model is to consider  $\sigma_n^2 \ll \sigma_A^2$ , so that  $G(\varepsilon_N; \sigma_n^2)$  is approximately impulsive.

Obviously, the specified pdfs are dissimilar from the modeled ones during the first iterations of the descent algorithm. When the estimation is approaching convergence, the model becomes more realistic. The random variables are independent and the overall error pdf shows larger tails as described by normal mixture  $f_\varepsilon(\varepsilon)$  (Figure 3). The nonlinear estimator of  $\varepsilon_A$  basically depends on the spike occurrence probability  $p$  and on the E noise compared to the adaptation mismatch\* Figure 4 shows the normalized  $g[\varepsilon]$  for a few values of the estimator parameters. The clipping for high-noise errors corresponds to the lack of adaptation when strong spikes occur. The threshold is proportional to the probability  $p$  and to the normalized spike variance  $\sigma_p^2/\sigma_A^2$ .

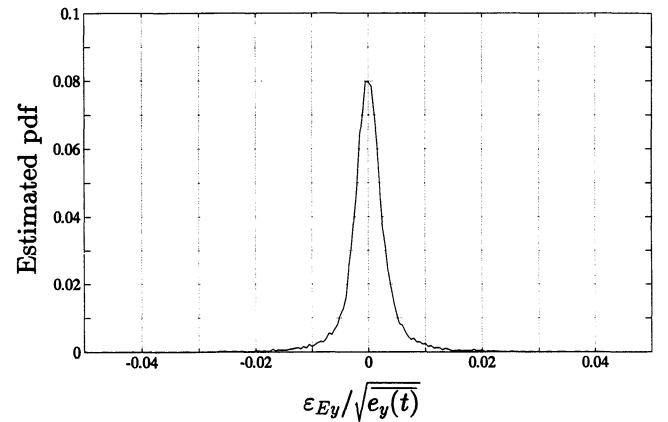


FIG. 3. Histogram of overall residual  $\varepsilon_{Ey} = e_y(t) - \hat{e}_y(t)$  normalized to the root mean square value of one component of E data after MT tensor estimation. The mathematically appealing Gaussian residual is not appropriate because the tails of the specified normal mixture pdf decay at rates lower than the rate of decay of a Gaussian pdf.

The slope of the linear part of  $g[\varepsilon]$  below the threshold basically depends on the ratio  $1/(1 + \sigma_n^2/\sigma_A^2)$ .

The Bayes' estimation of the residual depends on the noise level of the E channel that is used for the adaptation and corresponds to a modification of the step scale

$$\mu[\varepsilon(i)] = \mu \frac{g[\varepsilon(i)]}{\varepsilon(i)} \quad (22)$$

based on the noise pdf. In this way, when a strong spike occurs the adaptation is limited, and a low residual is interpreted as estimation mismatch. Since in the last stages of adaptive PR estimation the variance of adaptation error  $\sigma_A^2$  is reduced, as well as the ratio with the E noise, it is advisable to reduce the threshold of the nonlinear scale  $\mu[\varepsilon(i)]$ , as shown in Figure 4 for increasing  $\sigma_p^2/\sigma_A^2$ . The value of the step size is also reduced according to the ratio  $\sigma_A^2/\sigma_n^2$  for improving the final convergence (routine use suggests a reduction of approximately 1/10).

### Convergence of adaptive estimation

An adaptive search for the minimum MS error is equivalent to the iterative inversion of the correlation matrix  $\mathbf{R}_{HH}$ . Owing to this equivalence, the choice of the scale  $\mu$  in the adaptive PR estimation depends on the biggest eigenvalue  $\lambda_{max}$  of the correlation matrix (see Appendix);

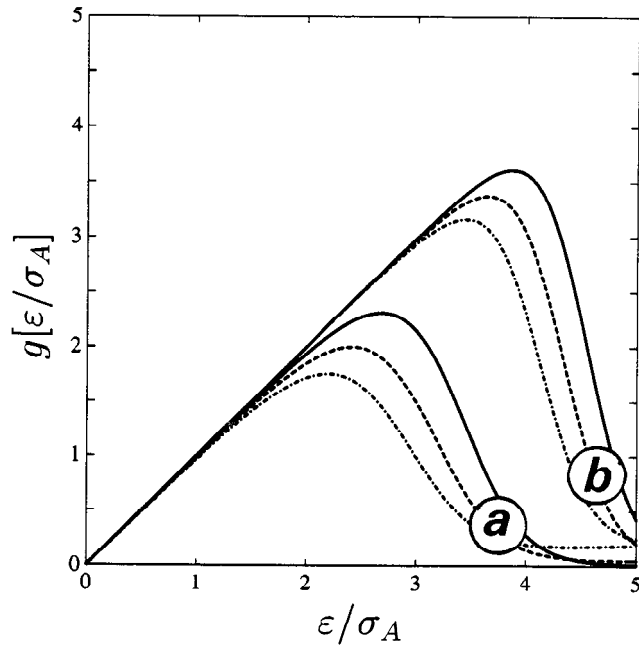


FIG. 4. Bayes' estimation  $g[\varepsilon]$  of the adaptation error  $\varepsilon_A$  from the residual  $\varepsilon = \varepsilon_A + \varepsilon_N$  given the E noise and adaptation error pdf. Only low values of the residual are interpreted as adaptation error. The threshold of the nonlinearity depends on the spike noise level estimated on the data and decreases as the spike occurrence probability  $p$  increases [(a):  $p = 10$  percent and (b):  $p = 0.1$  percent] and as the normalized spike noise variance  $\sigma_p^2/\sigma_A^2$  increases (solid line:  $\sigma_p^2/\sigma_A^2 = 5$ ; dashed line:  $\sigma_p^2/\sigma_A^2 = 10$ ; dash-dot line:  $\sigma_p^2/\sigma_A^2 = 25$ ). The slope below the threshold depends on the ratio  $1/(1 + \sigma_n^2/\sigma_A^2)$ .

i.e.,  $\mu < 1/\lambda_{max}$ . The suggested normalization for the scaling term of the step size

$$\mu = \frac{\mu_0}{\text{Tr}(\mathbf{R}_{HH})}, \quad (23)$$

depends on the power of the H data and gives a simple relationship for the convergence:  $0 < \mu_0 < 1$ . Under this condition the mean of the estimated PRs obtained with the adaptive technique converges to the optimum [equation (7)], as the number of iterations approaches infinity.

The adaptation algorithm (15) updates the estimated PR impedance tensor until convergence has been reached, then it moves the PRs randomly so that their mean value corresponds to the LMS estimate. The analysis of the residual supplies a criterion for convergence. However, a nonstationary time series makes the estimation of the statistical properties of the residuals far from an easy problem. Usually an assigned length data set, or a windowed one, is available, and for each time sample a steepest descent iteration is performed. For practical purposes, convergence is reached when a consecutive estimation on the same data set (using PR vector initialized from the previous estimation on the same data set) leads to a negligible reduction of the MS error over all the data. The MS error  $\varepsilon_{k+1}^2$  at the  $(k+1)$ th iteration of the same data set should be unchanged from the  $k$ th iteration. Choosing the convergence parameter  $\eta \leq 1$ , the convergence condition becomes:

$$\frac{\varepsilon_{k+1}^2}{\varepsilon_k^2} \leq \eta. \quad (24)$$

Because of the equivalence of the adaptive approach with iterative inversion of the correlation matrix, the convergence rate of the adaptive approach depends on the matrix eigenvalues. In Appendix A it is shown that the number of adaptive steps required is approximately proportional to the condition number of  $\mathbf{R}_{HH}$ ; i.e., to the ratio between the biggest and the smallest ( $\lambda_{min}$ ) eigenvalues. The condition number is limited by the ratio of the maximum to the minimum values of the H data power spectral density  $S_{HH}(\omega)$  (Haykin, 1986):

$$\frac{\lambda_{max}}{\lambda_{min}} \leq \frac{\max_{\omega}\{S_{HH}(\omega)\}}{\min_{\omega}\{S_{HH}(\omega)\}}. \quad (25)$$

Even an approximate prewhitening of the H data increases the convergence rate. A prewhitening of H data is advisable in MT data only over the frequency range where the power spectral ratio changes are more than 10 dB (e.g., for power spectra of telluric data the prewhitening should be performed for periods longer than 10 s). Since the prewhitened step-like noise becomes impulsive, the noise rejection capability of nonlinear adaptation is more effective for prewhitened E data.

If the noise in the H data is white and uncorrelated, the eigenvalues  $\lambda_j$  of the noisy correlation matrix  $\mathbf{R}_{HH}$  are biased by the noise power ( $\sigma_n^2$ ); i.e.,  $\lambda_j + \sigma_n^2$ . Because of the noise, the matrix  $\mathbf{R}_{HH}$  is full rank. Iterative PR estimation moves the PR vectors toward the noisy nonobservable subspace of the solution as shown by equation (12). The

adaptive solution contains not only a stable component because of the uncorrelated data but also an unstable component because of noise and data correlated (e.g., the correlation  $H_x = H_y$  leads to the trivial PRs estimate  $Z_{yx} = Z_{yy}$ ). To control the convergence for correlated data, an asymmetric choice of the nonlinear step size is proposed. In this way, a separate adaptation can be performed on both input channels

$$\begin{cases} \mathbf{Zyx}_{i+1} = \mathbf{Zyx}_i + 2\mu_{yx}[\varepsilon(i)]\varepsilon(i)\mathbf{Hx}(i) \\ \mathbf{Zyy}_{i+1} = \mathbf{Zyy}_i + 2\mu_{yy}[\varepsilon(i)]\varepsilon(i)\mathbf{Hy}(i) \end{cases} \quad (26)$$

so that adaptation can be reduced when the data show noise or short-time correlation. On real data, it seems reasonable to assume that the structure is 1-D, at least for the first iterations, thus leading to the choice of  $\mu_{yx} > \mu_{yy}$ . In practice, an empirical ratio of  $\mu_{yx}/\mu_{yy} \approx 10 - 50$  seems to provide good convergence properties.

Noise in the H data biases the LMS tensor estimate given by relationship (8). Including the effect of white-noise bias, the new tensor estimate  $\mathbf{Z}_n$  becomes approximately (for simplicity's sake,  $\sigma_n^2 < \lambda_j$  is assumed)

$$\mathbf{Z}_n = \mathbf{V}(\mathbf{\Lambda} + \sigma_n^2\mathbf{I})^{-1}\mathbf{V}^T\mathbf{P}_{EH} \approx (\mathbf{I} - \sigma_n^2\mathbf{R}_{HH}^{-1})\mathbf{Z}, \quad (27)$$

where  $\mathbf{\Lambda}$  is a diagonal matrix consisting of the eigenvalues of the correlation matrix  $\mathbf{R}_{HH}$ , and  $\mathbf{V}$  is the matrix of the eigenvectors associated with these eigenvalues. Noise in the H data facilitates convergence but may bias the estimated tensor. The correlation matrix for prewhitened H data is approximately diagonal and the downward bias of the PRs depends only on the signal-to-noise ratio (SNR) of the H data. In principle, relationship (27) can be written for several narrow-band problems. Since filtering is a linear processing, then for each PR estimation problem, the bias depends on the SNR evaluated within that band. To estimate the bias effect in the frequency domain, it is also convenient to express the PR estimates (27) in terms of their odd and even parts. The bias of odd and even parts of PR estimates is approximately the same. Therefore, the amplitude of the impedance tensor elements in frequency domain, rather than the phase, is biased by the noise in H data.

Noise in the E data increases the MS error when the convergence condition (24) has been met and randomly moves the estimated PR of the impedance tensor about the minimum point. Assuming that the mean value of the PR estimate at the latter adaptation stages corresponds to the tensor estimate at convergence (i.e.,  $\overline{\varepsilon_A} = 0$  since the adaptive algorithm is convergent in the mean), the estimated PRs, as well as the variance of the PR coefficients, are obtained from the average of the various estimated tensors obtained during the adaptation. Since there is no straightforward relationship that transforms the confidence limits of a time-domain estimated PR to the corresponding frequency-domain estimate, the uncertainty of the estimated impedance tensor in the frequency domain is evaluated by averaging Fourier transformed PR estimations for windowed data. Thus the way of estimating time-domain confidence limits closely resembles those frequently used in frequency-domain methods.

The convergence rate of the iterative descent method for prewhitened H data increases linearly with the length  $L$  of

the PR (Appendix A), so the use of other adaptive minimization techniques (e.g., the Newton method) can considerably improve the convergence rate of PR estimation. However, the number of multiplications per step required by other techniques is approximately  $6(L+1)^2$  compared to  $2(L+1)$  required for the steepest descent using equation (15). Consequently, the computational expense of the proposed technique is not excessive, and the procedure can be used directly in the field for quality control and for estimating the impedance tensor while data are still being collected.

## EXAMPLES AND COMMENTS

The following applications allow a comparison of the adaptive time-domain and frequency-domain approaches for MT impedance tensor estimation.

The first example is a synthetic 2-D structure where  $Z_{yx}(\omega)$  and  $Z_{yy}(\omega)$  are the impedance tensor elements. As independent variables, real H data have been chosen to verify the convergence of the adaptive tensor estimation for correlation of H time series. Figure 5 shows the results of the synthetic, and the correspondingly estimated, Z tensors when the E data were affected by noise (5 percent Gaussian noise or 100 percent of spike noise with occurrence probability of 1 percent). The results have been compared with the frequency-domain MT tensor estimation obtained using the algorithm proposed by Sims et al. (1971). Because of the power spectrum dispersal in real H data, the prewhitening is required to achieve an appreciable improvement of convergence rate. Even an approximate a priori pdf of noise (here it has been used as a nonlinear adaptation with  $p = 0.1$  percent and  $\sigma_p^2/\sigma_A^2 \approx 5$  shown in Figure 4) makes the time-domain approach more robust than the frequency-domain approach. Since the PRs are estimated using only E data that have a good SNR, the adaptive algorithm is more robust to spike noise than is the standard frequency-domain approach (Figure 5). Moreover, the importance of an intermediate time-domain step for spike removal in frequency-domain tensor estimations has been shown in Larsen (1989).

As the next example, the adaptive time-domain technique has been applied to a limited time series [five days of the so-called "active" data, cfr. Jones et al. (1989)] from EMSLAB Lincoln Line site data. The apparent resistivities have been derived from the impedance tensor elements through the relationship

$$\rho(f) = 0.2 \frac{|Z(f)|^2}{f}, \quad (28)$$

where  $Z(f)$  is the Fourier transform of the PR vectors and  $f$  is the frequency. The five-day time series has been divided into overlapped subsequences. Figure 6 shows the single-site adaptive time-domain estimates from all the subsequences (dots) compared with the robust remote reference impedance estimation (solid) presented by Jones et al. (1989).

Comparison of the adaptive time domain with the EMSLAB robust remote reference resistivities shows a bias of the impedance tensor estimation in the short period (below 200 s). The downward bias of the tensor is a result of noise in the H data, as described in relationship (27); however, the phase of the tensor elements shows less

dispersed values as predicted. In the field, a 12-bit A/D converter that allows 75 dB of overall dynamic range was used. Because of the dynamic range of the recording instruments, it is likely that the quantization noise in the H data contributes to a downward bias of the tensor estimation below 200 s. Compared to single-site processing, the bias of the adaptive time-domain solution benefits from the use of remote reference H data filtering. Figure 7 shows the result of apparent resistivities when H data have been thus filtered. The improved SNR of the H data has reduced the downward bias below 200 s, thus increasing the scattering of the solution as compared with Figure 6. Owing to the limited length of the PR vectors in the time domain ( $L = 250$  samples), the estimations for long periods ( $> 10^3$  s) are more dispersed, as frequency resolution depends on the length of the PR vectors. Nevertheless, since the number of operations grows linearly with the length of the PR vectors, it is advisable to limit the bandwidth to approximately one decade.

Another example from a southern Italian site shows that both time- and frequency-domain (Sims et al., 1971) approaches achieve comparable results when tensor estimates of good quality data are compared (Figure 8). The time-

domain technique shows a “hole” in the tensor element estimates at 150 Hz. This effect depends on the different E field coupling and on the gains of the analog filters used in the field to reject the third harmonic of power line noise (50 Hz). In any case, “holes” in the tensor estimate can be recovered easily during the processing.

Only simple averaging over overlapped windows have been shown here in the examples. Obviously, an additional statistical analysis on impedance tensors elements in the frequency domain can considerably improve the final quality of time-domain estimates.

CONCLUSIONS

MT time-domain processing is an iterative technique for the estimation of impedance tensor PRs that predict, in a LMS sense, the measured E data from the H data. The adaptive estimation of the impedance tensor performs an implicit iterative inversion of the H data correlation matrix. As with all MS iterative techniques, a larger eigenvalue spread of the H data correlation matrix reduces the convergence rate. However, convergence is speeded up even by an approximate prewhitening of the H data. Nonlinear adapta-

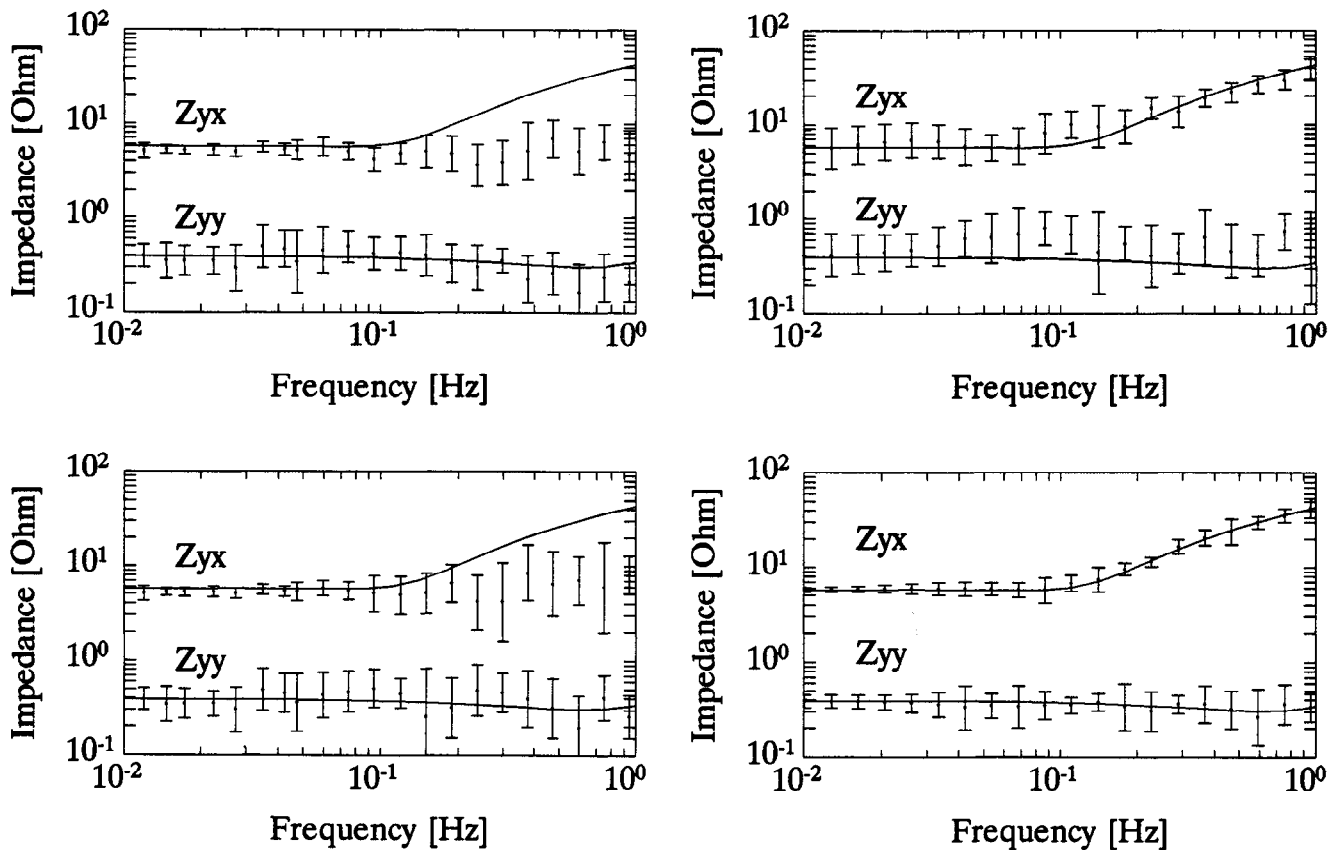


FIG. 5. Comparison of synthetic  $Z_{yx}(\omega)$  and  $Z_{yy}(\omega)$  impedance tensor estimates using the frequency-domain technique of Sims et al. (1971) in (a) and (c), and the adaptive time domain (data prewhitening and separate adaptation of  $\mu_{yx}/\mu_{yy} = 20$  applied) in (b) and (d). In all four panels  $Z_{yy}(\omega)$  is downward displaced by 1 decade for clarity; vertical bars represent 95 percent confidence region for Gaussian noise of 5 percent in (a) and (b), and 100 percent Gaussian spikes with occurrence probability 0.01 in (c) and (d).

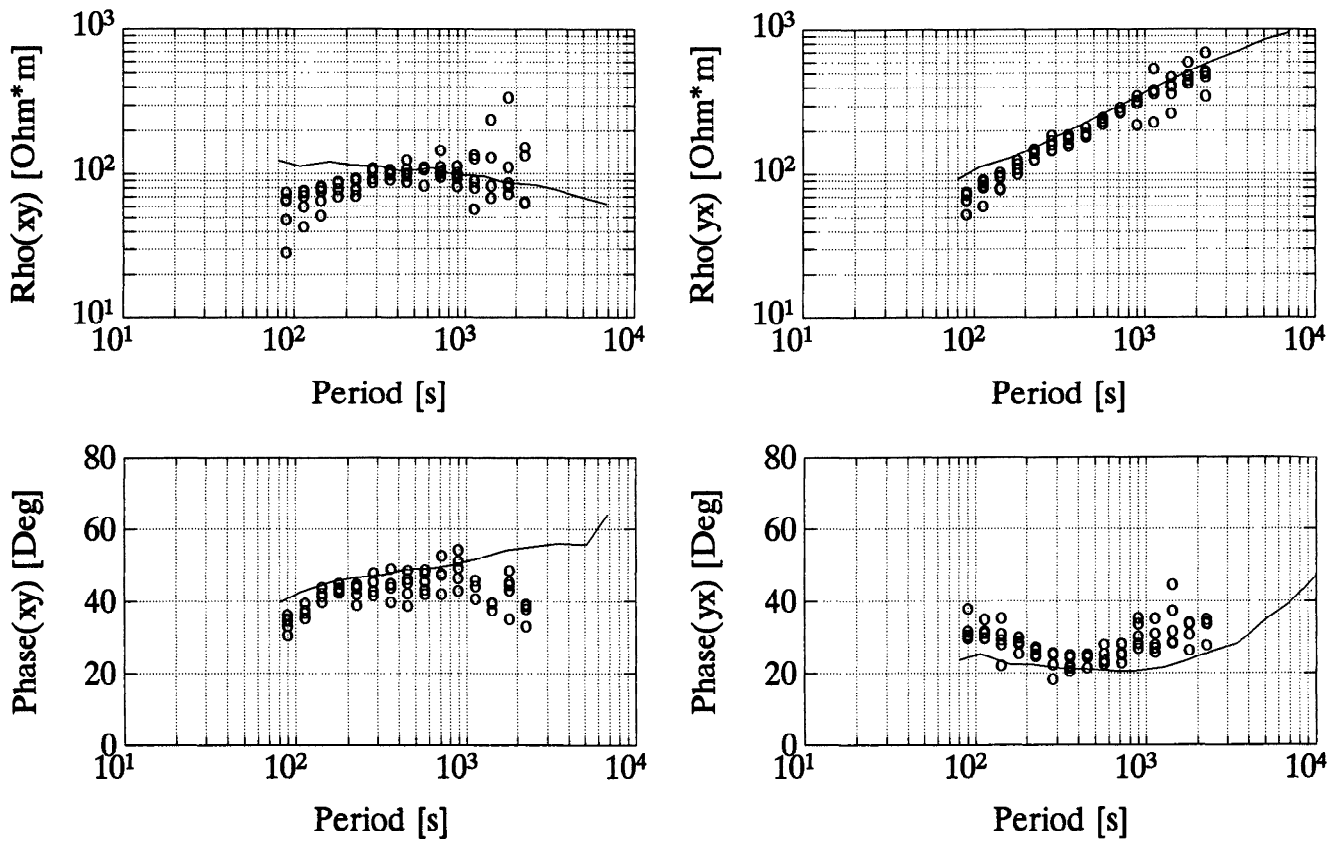


FIG. 6. Comparison between the apparent resistivity estimate that results from remote reference robust spectral analysis of 10 weeks of EMSLAB data (solid) and the one that is obtained from the single-site adaptive time-domain processing (dots) of overlapped subsequences of 5 days of data extracted from the entire EMSLAB data (Jones et al., 1989). The biases in the time-domain estimates below 200 s are caused by the low SNR of the H data.

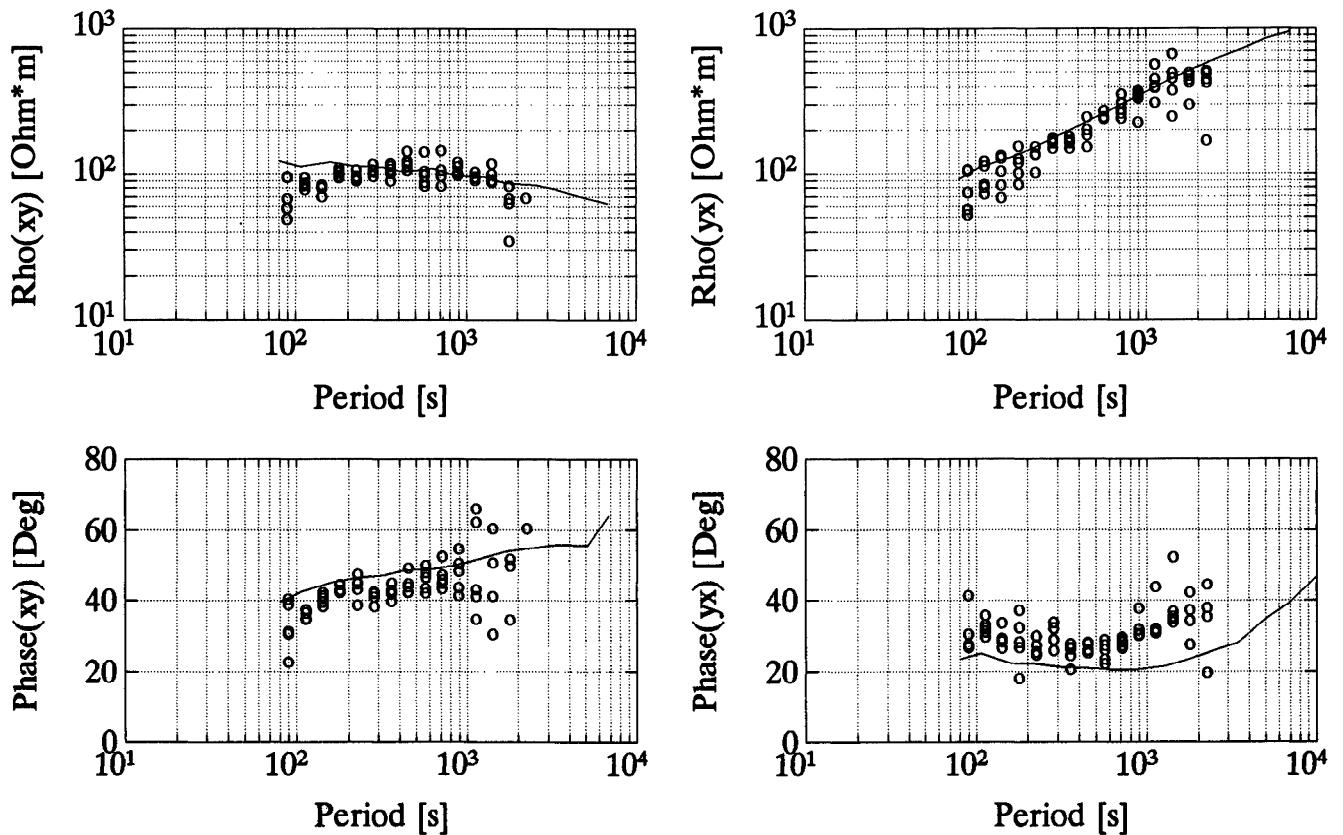


FIG. 7. Figure 6 after lowering the noise in the H data by using the remote reference data. The bias below 200 s has been limited, but the remote data noise has now increased the scattering of the estimated tensor elements over the full bandwidth.

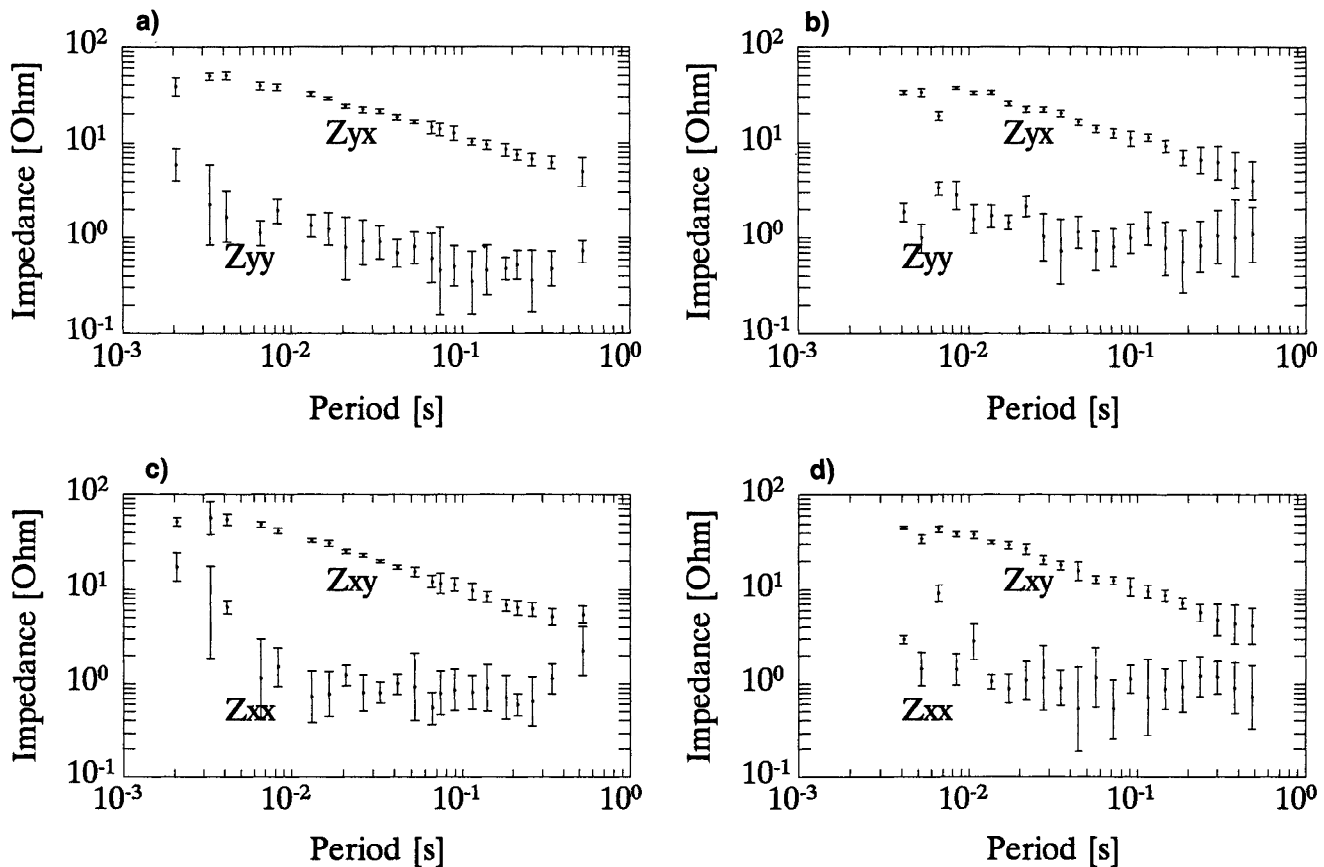


FIG. 8. Comparison between single-site impedance tensor estimates of low noise real data of a southern Italian site. In (a) and (c) using the frequency-domain technique of Sims et al. (1971), in (b) and (d) using adaptive time-domain processing.  $Z_{xx}(\omega)$  and  $Z_{yy}(\omega)$  are downward displaced by 2 octaves for clarity; vertical bars represent 95 percent confidence region. For good quality data (i.e., high SNR of E and H data, low H data correlation), both techniques achieve comparable results even for a 2-D structure. The “hole” in the time-domain tensor element estimate at 150 Hz is a result of the third harmonic power line coupling (50 Hz).

tion is a valuable tool for the control of the mismatch between the E data predicted by the PRs at the previous iteration and the measured E data. Using the Bayes' criterion, the data can be classified through analysis of the error as either too noisy (i.e., not reliable for PR estimations) or as useful for adaptation of PR vectors. Since in the adaptive time-domain technique the tensor is preferably estimated through the E data with the higher SNR, this classification appears as an advantage with respect to frequency-domain methods. Noise and high correlation between the two H time series can lead to a biased tensor estimation. Real data examples illustrated here, as well as routine use of the time-domain tensor estimation technique, show a lower sensitivity to H noise in the phase estimate than in the amplitude estimate. The separate adaptation of both H data may increase the quality of the results whenever it is possible to use any a priori information available from the geology of the investigated site. Remote reference H data have also been used to reduce the bias as a result of low SNR. However, to be effective over the complete bandwidth, the proposed remote reference time-domain processing could be improved by a selective noise cancellation of the H data using the remote reference.

The computational advantages of frequency-domain tensor estimation, because of the use of FFT algorithms, should be compared to the more flexible noise control of time-domain estimation. However, limiting the bandwidth to approximately one decade (e.g., by filtering and decimation of time-domain data) considerably improves the computational efficiency of the time-domain technique.

#### ACKNOWLEDGMENTS

I wish to express my gratitude to F. Rocca for many fruitful discussions and to G. Santarato for involving me in this stimulating research and for making critical evaluation during its course. I also want to thank Alan G. Jones for providing Lincoln Line site data (EMSLAB Juan de Fuca project).

#### REFERENCES

- Cantwell, L., 1960, Detection and analysis of low-frequency magnetotelluric signals: Ph.D. thesis, Massachusetts Institute of Technology.
- Fowler, R. A., Kotick, B. J., and Elliott, R. D., 1967, Polarization analysis of natural and artificially induced geomagnetic micropulsations: *J. Geophys. Res.*, **72**, 2871-2883.
- Haykin, S., 1986, Adaptive filter theory: Prentice Hall, Inc.

Jones, A. J., Chave, A. D., Egbert, G., Auld, D., and Bahr, K., 1989, A comparison of techniques for magnetotelluric response function estimation: *J. Geophys. Res.*, 94, 201-213.  
 Larsen, J. C., 1989, Transfer functions: Smooth robust estimates by least-squares and remote reference methods: *J. Geophys. Int.*, 99, 645-663.  
 McMechan, G. A., and Barrodale, I., 1985, Processing electromagnetic data in the time domain: *Geophys. J. Roy. Astr. Soc.*, 81, 277-293.  
 Neves, A. S., 1956, The magnetotelluric method in two-dimensional structures: Ph.D. thesis, Massachusetts Institute of Technology.  
 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.  
 Swift, Jr., C. M., 1967, A magnetotelluric investigation of electrical conductivity anomaly in the southwestern United States: Ph.D. thesis, Massachusetts Institute of Technology.  
 Widrow, B., McCool, J. M., Larimore, M. G., and Johnson, Jr., C. R., 1976, Stationary and nonstationary characteristics of the LMS adaptive filter: *Proc. Inst. Elect. Electron. Eng.*, 64, 1151-1162.  
 Yee, E., Kosteniuk, P. R., Paulson, K. V., 1988, The reconstruction of the magnetotelluric impedance tensor: An adaptive parametric time-domain approach: *Geophysics*, 53, 1080-1087.

## APPENDIX

### CONVERGENCE PROPERTIES OF TIME DOMAIN MT PROCESSING

The relationships derived here are adapted from the adaptive theory literature (Widrow et al., 1976; Haykin, 1986) for the estimation of the MT impedance tensor.

Given the MS error (the time dependence is omitted):

$$\bar{\varepsilon}^2(\mathbf{Z}) = \bar{\varepsilon}_y^2 + \mathbf{Z}^T \mathbf{R}_{HH} \mathbf{Z} - 2\mathbf{P}_{EH}^T \mathbf{Z}, \quad (\text{A-1})$$

the gradient search (13) becomes

$$\mathbf{Z}_{i+1} = \mathbf{Z}_i - 2\mu(\mathbf{R}_{HH}\mathbf{Z}_i - \mathbf{P}_{EH}). \quad (\text{A-2})$$

It is more convenient to define the deviation from the LMS solution of the PRs to show the equivalence between the adaptive technique and iterative  $\mathbf{R}_{HH}^{-1}$  estimation

$$\Delta\mathbf{Z}_i = \mathbf{Z}_i - \mathbf{R}_{HH}^{-1}\mathbf{P}_{EH}. \quad (\text{A-3})$$

Equation (A-2) now depends on the PRs deviation and on the iteration

$$\Delta\mathbf{Z}_{i+1} = (\mathbf{I} - 2\mu\mathbf{R}_{HH})\Delta\mathbf{Z}_i = (\mathbf{I} - 2\mu\mathbf{R}_{HH})^{i+1}\Delta\mathbf{Z}_0. \quad (\text{A-4})$$

The condition for the adaptive technique to be convergent in the mean, from every starting condition  $\mathbf{Z}_1$ , is that  $\Delta\mathbf{Z}_i \rightarrow 0$ . From the eigenvector decomposition of  $\mathbf{R}_{HH}$  (i.e.,  $\mathbf{V}^T \mathbf{R}_{HH} \mathbf{V} = \mathbf{\Lambda}$ ), the relationship (A-4) becomes

$$\Delta\mathbf{Z}_{i+1} = \mathbf{V}(\mathbf{I} - 2\mu\mathbf{\Lambda})^{i+1}\mathbf{V}^T\Delta\mathbf{Z}_0. \quad (\text{A-5})$$

For convergence of the adaptive algorithm,

$$\lim_{i \rightarrow \infty} (\mathbf{I} - 2\mu\mathbf{\Lambda})^i = \mathbf{0}; \quad (\text{A-6})$$

the scaling term in the step size is limited to the maximum eigenvalue  $0 < \mu < 1/\lambda_{max}$  (the eigenvalues are real and positive). The trace of the correlation matrix is an upper bound of  $\lambda_{max}$  and since

$$\text{Tr}(\mathbf{R}_{HH}) = \sum_j \lambda_j > \lambda_{max} \quad (\text{A-7})$$

holds, then, the relationship (23) always guarantees convergence of the adaptive technique.

For evaluation of the convergence rate, the quadratic MS error with respect to PR equation (A-1) is referred to the LMS value of the error  $\bar{\varepsilon}_{min}^2$ . The MS error at the  $i$ th iteration depends on the starting point of the PRs, which from equation (A-5) is

$$\begin{aligned} \bar{\varepsilon}^2(\mathbf{Z}_i) &= \bar{\varepsilon}_{min}^2 + \Delta\mathbf{Z}_i^T \mathbf{R}_{HH} \Delta\mathbf{Z}_i \\ &= \bar{\varepsilon}_{min}^2 + \Delta\mathbf{Z}_i^T \mathbf{V} \mathbf{\Lambda} (\mathbf{I} - 2\mu\mathbf{\Lambda})^{2i} \mathbf{V}^T \Delta\mathbf{Z}_0. \end{aligned} \quad (\text{A-8})$$

The term  $(\mathbf{I} - 2\mu\mathbf{\Lambda})^{2i}$  now regulates the convergence rate as the eigenvectors in  $\mathbf{V}$  represent only a change of coordinate for PRs. The rate of convergence is limited by the smallest eigenvalue  $\lambda_{min}$ . The number of iterations  $n$  for 1 percent of reduction of the MS error from the starting point  $\Delta\mathbf{Z}_0$  and for its slower rate [i.e.,  $(1 - 2\mu\lambda_{min})^{2n} = 0.011$  can be approximated as  $1/n \approx 2\mu\lambda_{min}$ . For the maximum step size allowed by equation (A-6), the convergence rate depends on the eigenvalue ratio  $n \approx \lambda_{max}/2\lambda_{min}$  as shown. Since choice (23) guarantees the unconditional convergence, the convergence rate becomes

$$n \approx \frac{\sum_j \lambda_j}{2\mu_0 \lambda_{min}} > \frac{\lambda_{max}}{2\mu_0 \lambda_{min}}. \quad (\text{A-9})$$

In practice, since  $\mu_0 \approx 0.1 \div 0.001$  is a reasonable choice for PR estimation of MT data, this value considerably decreases the convergence rate. For prewhitened data,  $\lambda_j \approx \lambda$  and the convergence rate is linearly dependent on the length of the PRs [ $n \approx (L + 1)/\mu_0$ ].