Geophys. J. Int. (1997) **130**, 475–496

Robust multiple-station magnetotelluric data processing

Gary D. Egbert

College of Oceanic and Atmospheric Sciences, Oregon State University, Corvallis, OR 97331-5503, *USA. E-mail: egbert@oce.orst.edu*

Accepted 1997 April 4. Received 1996 October 17; in original form 1996 January 7

SUMMARY

Although modern magnetotelluric (MT) data are highly multivariate (multiple components, recorded at multiple stations), commonly used processing methods are based on univariate statistical procedures. Here we develop a practical robust processing scheme which is based on multivariate statistical methods. With this approach we use data from all channels to improve signal-to-noise ratios, and to diagnose possible biases due to coherent noise. To illustrate our approach we use data from two- and three-station wide-band MT arrays from an area south of San Francisco, California, where contamination of the MT signal by spatially coherent cultural electromagnetic noise is severe at some periods. To deal with such coherent noise we adopt a two-stage procedure. In the first stage we focus on reducing the effects of incoherent noise, and testing for the presence of coherent noise. To this end we have developed a robust multivariate errors-in-variables (RMEV) estimator, which estimates background noise levels, cleans up outliers in all channels, and determines the 'coherence dimension' of the array data. In the absence of coherent noise, the coherence dimension of the data will be two (corresponding to two polarizations of the plane-wave MT source fields). In this case the RMEV estimator provides direct estimates of MT impedances and inter-station transfer functions. We show, with synthetic and real data examples, that in some cases these estimates can be significantly better than those obtained with more standard robust remote reference estimators. When MT data is severely contaminated by coherent noise (as for our example arrays for periods of 4–50 s) the coherence dimension of the data will exceed two. The RMEV estimate thus provides a clear warning of coherent noise contamination. Although there appears to be no completely general automatic way to deal with this circumstance, useful results can be obtained from severely contaminated data in some cases. We show in particular how the RMEV estimator can be adapted to separate the MT signal from coherent noise for two special cases: when at least one site is unaffected by coherent noise, and when coherent noise sources are intermittent. We give examples of significant improvements in MT impedance estimates obtained with the RMEV estimate for each of these cases.

Key words: coherent noise, magnetotellurics, multivariate statistics, robust estimation, transfer functions.

1 INTRODUCTION

The magnetotelluric (MT) method for imaging subsurface electrical conductivity has proved to be a useful geophysical tool both for fundamental studies of continental structure and tectonics (e.g. Wanamaker *et al.* 1989; Stanley *et al.* 1990), and for solving applied problems in exploration for hydrocarbon or geothermal resources (e.g. Orange 1989; Goldstein 1988). MT uses naturally occurring electromagnetic (EM) field variations, and thus can offer significant cost advantages over active-source surface EM methods. However, the use of passive sources also presents a problem—signal and noise levels can be highly variable, and are largely beyond the control of the experimenter. Success with the MT method thus requires great care in the acquisition and initial reduction of EM time series. The simple single-station least-squares approach used by early MT practitioners (e.g. Swift 1967; Sims, Bostick & Smith 1971) can yield estimates of apparent resistivities and phase that are heavily biased or wildly oscillatory.

Two developments have improved this situation greatly—the remote reference (RR) method, in which horizontal magnetic fields recorded simultaneously at a second remote site are correlated with the EM fields at the local site (Gamble, Goubau & Clarke 1979), and various sorts of robust data-adaptive weighting schemes (e.g. Jones & Jodicke 1984; Egbert & Booker 1986; Stodt 1986; Chave, Thomson & Ander 1987;

Chave & Thomson 1989; Larsen 1989; Sutarno & Vozoff 1991; Spangolini 1994; see Jones *et al.* 1989 for a general discussion and comparison). However, all too often there are still critical sites and/or frequency ranges where useful results are not obtained, even with some sort of robust processing of RR data. Typically, these difficulties are most severe for periods from 1 to 10 s (the 'dead band'), and in areas where urban cultural EM noise is significant.

For example, in the EMSLAB experiment (Booker & Chave 1989), MT data in the Willamette Valley were contaminated by cultural noise, leading to poor estimates for periods of 1–10 s. As a result, the conductivity structure of the valley basement was poorly constrained. Two very different models (Jiracek *et al.* 1989; Wannamaker *et al.* 1989), which suggested very different interpretations of the physical significance of mid-lower crustal conductors across much of the EMSLAB land profile, were both allowed by the large error bars on the dead-band impedances.

As a more dramatic example of the difficulties that can be encountered, we plot in Fig. 1 robust RR apparent resistivity and phase estimates from a pair of MT sites from near the epicentre of the 1989 Loma Prieta earthquake (Mackie & Madden 1992). These two simultaneously occupied sites were approximately 20 km apart (A and B in Fig. 2), a distance that would seem to be great enough for successful RR processing. However, apparent resistivity and phase curves [computed with a robust leverage-controlled remote reference procedure (Egbert & Booker 1986; Chave & Thomson 1989)] vary rapidly at periods of 1–10 s, where apparent resistivity (ρ_a) increases steeply and phases (ϕ) become negative.

Mackie & Madden (1992) inferred that the behaviour of ρ_a and ϕ evident in Fig. 1 resulted from contamination of the MT signal by non-uniform EM cultural-noise sources in the San Francisco Bay area to the north. In particular, the Bay Area Rapid Transit System (BART, a DC electric rail system) had previously been shown to overwhelm natural sources at distances up to several tens of kilometres from the railway tracks (Fraser-Smith & Coates 1978). Due to return currents leaking from the tracks into the Earth, BART should look like a series of grounded electric dipoles, with temporally varying geometries and currents (Fraser-Smith & Coates 1978). Measurements made in the near-field of this EM source would not be consistent with the usual MT assumption of spatially uniform sources, and could well lead to distortions of ρ_a and ϕ of the sort seen in Fig. 1. Robust RR methods (at least of the sort used for Fig. 1) are apparently ineffective when confronted with these sorts of difficulties.

In this paper we report on our efforts to develop improved methods for processing multiple-station wide-band MT data. We explore two interrelated themes. First, modern MT data are typically highly multivariate with multiple-channel time series recorded at multiple stations, sometimes in the presence of cultural noise which may have a complex inter-component correlation structure. The statistical methods that have traditionally been applied to MT processing are basically univariate in nature, with each noisy output channel treated separately. Here we take a multivariate statistical approach, extending the geomagnetic array processing scheme of Egbert & Booker (1989; hereinafter referred to as EB). A second theme is robustness. There is no question that the basic idea of adaptively computed data weights has had a major impact on the quality of MT impedance estimates. However, the best way to apply this general idea to a problem as complex as RR MT processing is far from obvious. Again, methods in common use (i.e. robust regression) were in fact developed for a comparatively simple univariate statistical problem—a single dependent real variable predicted by several real independent variables. Given the multivariate character of multi-station MT data, with noise (and possibly outliers!) in all channels, the whole question of

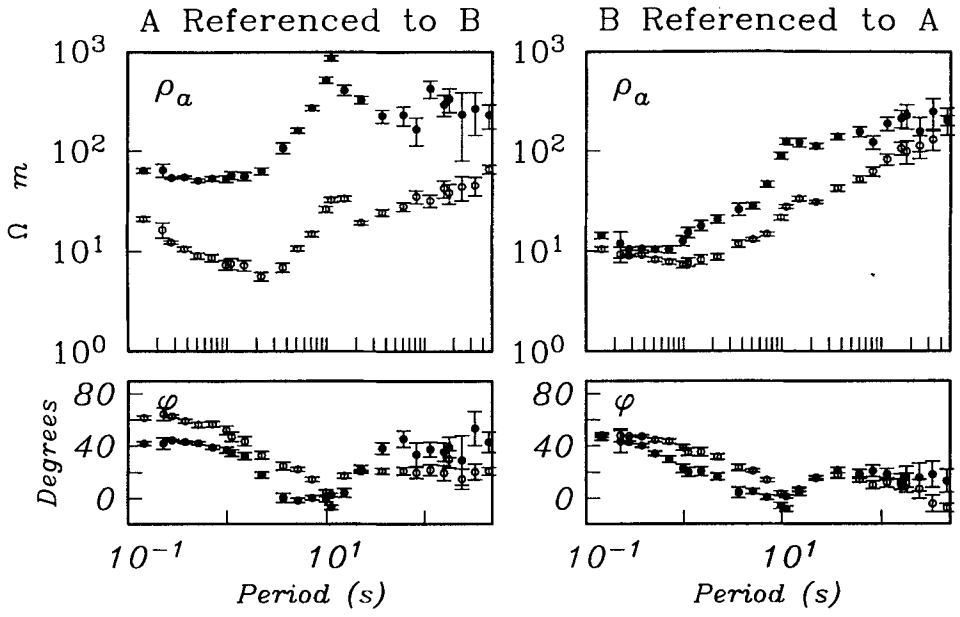

Figure 1. Apparent resistivity and phase curves for a pair of MT sites ('A' and 'B') near the epicentre of the 1989 Loma Prieta, California earthquake. The two sites were occupied simultaneously, and estimates were computed using a leverage-controlled robust remote reference (RRR) procedure. At both sites there is evidence of cultural-noise contamination in the period range 1–100 s. See Fig. 2 for station locations.

robustness in MT data processing deserves a much closer look. Here we explore the notion of robustness in a multivariate context with the goal of developing a practical robust multiple-station MT processing scheme. The methods we discuss are in particular tailored to processing data from small (two or three station) MT arrays in circumstances where signal-to-noise ratios are low and/or cultural EM noise is significant.

In Section 2 we briefly review the multivariate model and outline the two-stage procedure we have developed to allow for the possibility of coherent noise. In the first stage, discussed in Section 3, 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 or noise ratio which can be resolved by the array—can be estimated. When there is no coherent noise, M will be two, corresponding to the two orthogonal plane-wave MT sources. In this case, the 2-D coherent part of the array signal can be used to estimate MT impedances and inter-station transfer functions directly.

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. In Section 4 we illustrate the basic ideas, and demonstrate the effectiveness of the proposed scheme, using data from a three-station wide-band MT array from the same general area as the Loma Prieta survey (stations 1, 4 and 5 in Fig. 2).

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 will perhaps often be impossible, but a careful array analysis can help to define the nature and extent of coherent-noise problems. At least in some cases this information will allow 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. We consider two special cases. In Section 5 we consider the case of a quiet reference site, relatively unaffected by coherent noise. In Section 6, we consider the case of intermittent coherent noise. We apply these specialized methods to the two- and three-station example arrays of Fig. 2 (A, B, and 1, 4, 5).

The proposed scheme is unquestionably more complicated than methods currently in use. However, the scheme provides a clear diagnostic for the presence of coherent noise, and should thus be of some value when such noise might be encountered. In the absence of coherent noise, the scheme works in a more-or-less automatic fashion, and, at least in some cases, yields estimates that are significantly better than those obtained with simpler approaches (e.g. robust RR). We must stress, however, that our scheme does not provide a foolproof automatic way to eliminate the effects of coherent noise. Successful application of the second stage of the estimation procedure (required when coherent noise is found) will only be possible in some cases. Furthermore, the appropriate way to proceed in such cases may be highly variable, and will require some guidance from a user who understands both the multivariate approach and the nature of EM noise.

2 THE MULTIVARIATE STATISTICAL MODEL

We work in the frequency domain with Fourier coefficients derived from a series of short overlapping time segments (Egbert & Booker 1986). Estimates for each frequency band are computed independently, so we generally omit specific reference to frequency in the following. To simplify the discussion we assume that there are five channels of data observed at each of J stations, so that the total number of channels observed in the array is $K = 5J$. Generalization to other array configurations is straightforward. Assuming spatially uniform external sources, and allowing for noise in all channels, the frequency-domain MT array data vectors satisfy

$$\mathbf{X}_i = \begin{pmatrix} \mathbf{h}_{1i} \\ \mathbf{e}_{1i} \\ \hline \vdots \\ \hline \mathbf{h}_{Ji} \\ \mathbf{e}_{Ji} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\eta}_{11} \\ \boldsymbol{\zeta}_{11} \\ \hline \vdots \\ \hline \boldsymbol{\eta}_{J1} \\ \boldsymbol{\zeta}_{J1} \end{pmatrix} \beta_{1i} + \begin{pmatrix} \boldsymbol{\eta}_{12} \\ \boldsymbol{\zeta}_{12} \\ \hline \vdots \\ \hline \boldsymbol{\eta}_{J2} \\ \boldsymbol{\zeta}_{J2} \end{pmatrix} \beta_{2i} + \boldsymbol{\varepsilon}_i = \mathbf{U}\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i . \quad (1)$$

Here $\mathbf{h}_{ji}$ and $\mathbf{e}_{ji}$ are the magnetic and electric field Fourier coefficients computed for the ith time segment at the jth site, the parameters β_{li}, $l = 1, 2$, define the polarization of the source magnetic fields for this data segment, and the vectors $\boldsymbol{\varepsilon}_i$ represent all sources of noise, which we take to be statistically independent of the natural-source MT signals.

The columns of $\mathbf{U}$ are K-dimensional complex vectors, which ideally represent the magnetic and electric fields that would be observed at all sites for idealized quasi-uniform magnetic sources, linearly polarized N–S ($l=1$) and E–W ($l=2$). Note, however, that $\mathbf{U}$ is only determined up to multiplication on the right by an arbitrary invertible 2×2 matrix $\mathbf{A}$, and is not uniquely identifiable for the array data (EB). Physically, this indeterminacy results from the impossibility of uniquely determining the external source polarization from observations of total (internal plus external) fields. It is easy to check that impedance tensors for each site, as well as vertical field and inter-station transfer functions, are simply determined from $\mathbf{U}$ (independently of the arbitrary matrix $\mathbf{A}$; EB). For example, the impedance tensor for site j is given in terms of the elements of $\mathbf{U}$ corresponding to the x and y components of the electric and magnetic fields observed at station j:

$$\mathbf{Z}_j = \begin{bmatrix} \zeta_{xj1} & \zeta_{xj2} \\ \zeta_{yj1} & \zeta_{yj2} \end{bmatrix} \begin{bmatrix} \eta_{xj1} & \eta_{xj2} \\ \eta_{yj1} & \eta_{yj2} \end{bmatrix}^{-1} . \quad (2)$$

Note that because of the indeterminacy in $\mathbf{U}$ we may assume that $\mathbf{U}^*\mathbf{U} = \mathbf{I}$, where the asterisk denotes the complex conjugate transpose.

Eq. (1) defines a multivariate linear statistical model, which in various forms has been referred to as the structural relationship, factor analysis, linear functional equation, and (the name we adopt) multivariate errors-in-variables (MEV) model. These models for formal parameter estimation are closely related to more exploratory methods for data analysis known as principal components or empirical orthogonal functions. A general review from a statistical perspective is given by Anderson (1984), and

an extensive discussion of applications to geomagnetic array data is given by EB.

Unbiased [and maximum-likelihood for Gaussian errors; see Gleser (1981)] estimates of $\mathbf{U}$ can be obtained by solving the generalized eigenvalue problem

$$\mathbf{Su} = \lambda \boldsymbol{\Sigma}_N \mathbf{u}. \tag{3}$$

Here $\mathbf{S} = 1/I \, \Sigma \, \mathbf{X}_i \mathbf{X}_i^*$ is the spectral density matrix (SDM, the $K \times K$ matrix of all possible component cross-product averages) and $\boldsymbol{\Sigma}_N$ is the covariance matrix of the noise. The two columns of $\mathbf{U}$ can be estimated using the eigenvectors associated with the two largest eigenvalues of (3). More explicitly, $\mathbf{U}$ can be estimated as

$$\hat{\mathbf{U}} = \boldsymbol{\Sigma}_N^{-1/2} \mathbf{U}' (\mathbf{U}'^* \boldsymbol{\Sigma}_N^{-1} \mathbf{U}')^{-1}, \tag{4}$$

Where $\mathbf{U}'$ is a $K \times 2$ matrix consisting of the two dominant eigenvectors of $\mathbf{S}'$:

$$\mathbf{S}' = \boldsymbol{\Sigma}_N^{-1} \mathbf{S} \boldsymbol{\Sigma}_N^{-1/2}. \tag{5}$$

Note that in (4) we have enforced the condition $\mathbf{U}^*\mathbf{U} = \mathbf{I}$. Application of this 'eigenvector estimate' to MT data processing has been proposed previously by Jupp (1978) and Park & Chave (1984), who suggested that this method could be used to overcome bias problems in single-station impedance estimates. This idea has not proved particularly useful in practice, primarily because it is essential to estimate the noise covariance $\boldsymbol{\Sigma}_N$ accurately to eliminate bias, and this turns out to be difficult to accomplish with a small number of data channels (EB; Park & Chave 1984).

If we were to allow for a completely general form for the noise covariance, the sample SDM $\mathbf{S}$ could obviously be matched exactly with a variety of possible combinations of noise ($\boldsymbol{\Sigma}_N$) and signal ($\mathbf{U}$) parameters. In particular, we could estimate $\hat{\boldsymbol{\Sigma}}_N = \mathbf{S}$ (i.e., the data is pure noise), leaving $\mathbf{U}$ (the MT parameters of interest) completely unconstrained. To make any progress at all, we thus must assume a restricted form for $\boldsymbol{\Sigma}_N$. The simplest model for the noise covariance is the diagonal form

$$\boldsymbol{\Sigma}_N = \mathrm{diag}(\sigma_1^2 \ldots \sigma_K^2), \tag{6}$$

which implies that all noise is incoherent between stations, and also between channels at a single station. Arguments given in Anderson & Rubin (1956) show that if (6) and (1) hold, signal and noise parameters will be uniquely identifiable provided $K \geqslant 5$. Thus, in principle, incoherent-noise variances for each channel should be estimable from single-station 5-component MT data. In practice, with finite data, this conclusion appears to be optimistic (Egbert 1987). Fortunately, the feasibility of reliably estimating the statistical properties of the noise increases with the total number of data channels K. With at least two MT stations, estimation of the incoherent-noise variances of (6) is reasonably simple, as we shall show below. Furthermore, more complicated models for $\boldsymbol{\Sigma}_N$ may be assumed. For example, Egbert (1987) shows that, with two or more stations, models for local noise can be arbitrary, assuming only that noise is uncorrelated between sites. Allowing for noise which is coherent *between* sites presents much more serious difficulties, even when the number of stations is very large. Without precise *a priori* knowledge of coherent noise geometry it is not possible to include these noise sources in a parametric model for $\boldsymbol{\Sigma}_N$.

To allow for coherent noise in a general way we have thus found it useful to begin our analysis with a modified version of (1):

$$\mathbf{X}_i = \mathbf{U}\boldsymbol{\beta}_i + \mathbf{V}\boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_i = [\mathbf{U} \quad \mathbf{V}] \begin{bmatrix} \boldsymbol{\beta}_i \\ \boldsymbol{\gamma}_i \end{bmatrix} + \boldsymbol{\varepsilon}_i = \mathbf{W}\boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}_i, \tag{7}$$

where $\mathbf{V}$ is a $K \times N$ matrix whose columns represent coherent-noise sources. Now $\boldsymbol{\varepsilon}_i$ represents only *incoherent* noise, guaranteeing that (6) is a reasonable model for $\boldsymbol{\Sigma}_N$. With this approach we initially treat any possible coherent noise essentially as signal, and combine $\mathbf{U}$ and $\mathbf{V}$ into a single $K \times M$ matrix $\mathbf{W}$ of coherent signal/noise vectors ($M = N + 2$). As for the case of $\mathbf{U}$, $\mathbf{W}$ is determined only up to multiplication on the right by an arbitrary $M \times M$ invertible matrix $\mathbf{A}$. We may thus assume that $\mathbf{W}$ is a unitary matrix (satisfying $\mathbf{W}^*\mathbf{W} = \mathbf{I}$). Here the indeterminacy presents a serious difficulty, in that each of the columns of the estimated matrix $\mathbf{W}$ can be an arbitrary mixture of the columns of the signal and coherent-noise matrices ($\mathbf{U}$ and $\mathbf{V}$, respectively). By adopting (7) we are thus only delaying the inevitable difficulty of separating coherent noise from signal to a later stage of the analysis. However, we now start with a model that is general enough to include the true situation, and from which we can learn at least some useful things about our data. With this model we can estimate incoherent-noise levels in each channel, and clean up local isolated outliers. Most importantly, we can estimate the 'coherence dimension' M of the data (i.e. the number of columns in $\mathbf{W}$), and thus the dimensionality $N = M - 2$ of coherent noise.

3 ROBUST ESTIMATION FOR THE MEV MODEL

For the MEV model of (7) with the noise covariance given by (6), the unknown parameters are the incoherent-noise variances $\sigma_k^2, k = 1, K$, the coherence dimension $M = N + 2$, and the elements of the $K \times M$ matrix $\mathbf{W}$ (modulo the fundamental indeterminacy noted above). In this section we present a practical robust scheme for the estimation of all of these unknowns. If we find that $M = 2$, $\mathbf{W}$ can be identified with $\mathbf{U}$, the response of the Earth to quasi-uniform external sources. MT impedances (or other transfer functions) can then be computed using (2) (or analogues). If $M > 2$, we have clear evidence for coherent noise. We are then faced with the difficult task of separating this noise (i.e., $\mathbf{V}$) from the desired MT signal ($\mathbf{U}$). We consider some possible strategies for dealing with this difficult problem in Sections 5 and 6.

Although the details of the full multivariate robust scheme are somewhat involved, the basic ideas are quite simple. First note that if we knew (or had good estimates of) $\sigma_k^2, k = 1, K$ estimates of M and $\mathbf{W}$ would be relatively straightforward to compute. With the data transformed as

$$C'_{ik} = X_{ik}/\sigma_k, \tag{8}$$

incoherent noise results in an isotropic scatter of unit variance in the transformed data space. Hence, in this space, only those directions corresponding to coherent parts of the data should have variances significantly greater than one. Thus, if we use the rescaled data to construct $\mathbf{S}' = \boldsymbol{\Sigma}_N^{-1/2} \mathbf{S} \boldsymbol{\Sigma}_N^{-1/2}$, the number of eigenvalues of this Hermitian matrix significantly greater than one provides an estimate of M. The corresponding

eigenvectors (scaled back into the original data coordinates) provide an estimate of **W**; see eqs (4)–(5).

To proceed with this strategy we must first estimate the incoherent-noise variances, without knowing M or **W**. The incoherent part of the data in each channel is by definition not predictable from data in the remaining channels. Thus, if we use multiple linear regression to fit data for channel k to the remaining $K-1$ channels, the magnitude of the residuals provides at least a rough estimate of the incoherent-noise variance for channel k. That is, we can estimate σ_k^2 as the residual variance obtained from fitting the linear model

$$X_{ik} = \sum_{k' \neq k} b_{k'} X_{ik'} + \varepsilon_{ik}. \tag{9}$$

This simple estimate can be refined. In particular, bias in the estimates of σ_k^2 resulting from the presence of noise in all channels can be at least approximately removed (Appendix A). Furthermore, once we have a rough estimate of M we can improve the stability and accuracy of the variance estimates. These ideas are discussed further below, and in Appendix A. The key point to note is that the simple initial estimate of σ_k^2, $k = 1, K$, proposed here does not require any prior information about the other unknown parameters, and is thus suitable as a starting point for the robust iterative scheme defined more precisely below.

The final key component of our estimation scheme is to use adaptively determined weights to clean up isolated outliers in individual channels, and to downweight unusual data vectors. In broad outline, the approach used is very similar to the regression M-estimate previously applied to geomagnetic transfer function estimation problems by Egbert & Booker (1986) and Chave *et al.* (1987). We begin with further details on these robust methods.

3.1 Robust estimation of the SDM

Since our estimates of M and **W** ultimately depend on **S**, it is useful to consider the problem of robust estimation of the SDM, without making specific model assumptions concerning signal or noise. With this approach, we focus on limiting the effect of a few unusual data vectors on the estimated SDM.

Huber (1981, Chapter 8) discusses a variety of approaches to robust covariance matrix estimation. Any of these methods could be applied to robust estimation of the SDM, since **S** is essentially just the complex sample covariance matrix for the frequency-domain data vectors $\mathbf{X}_i$. Here we adapt the affinely invariant approach of Huber to allow for complex data vectors. This approach is similar to the M-estimate—estimates of the SDM are obtained as maximum-likelihood estimates for a heavy-tailed 'ellipsoidal' family of multivariate densities. Note that we may assume that the Fourier coefficients have zero mean, significantly simplifying the more general scheme discussed by Huber. To derive the estimator, one assumes a family of multivariate probability densities of the form

$$f(\mathbf{x}; \boldsymbol{\Psi}) = |\det \boldsymbol{\Psi}| f(|\boldsymbol{\Psi}\mathbf{x}|), \tag{10}$$

where f is a spherically symmetric probability density, and the multi-dimensional shape of the distribution is given by the 'pseudo-covariance matrix' $(\boldsymbol{\Psi}\boldsymbol{\Psi}^*)^{-1}$. For the K-dimensional complex Gaussian distribution, $f(r) = (4\pi)^{-K} \exp[-r^2/4]$. To make the procedure robust a heavy-tailed f is used. For example, we have taken

$$f(r) = \begin{cases} C_0 \exp[-r^2/4] & \text{if} \quad r \leq r_0, \\ C_1 \exp[-r] & \text{if} \quad r \geq r_0, \end{cases} \tag{11}$$

where C_0 and C_1 are chosen to make $f(r)$ a continuous density function. Other choices of f are discussed in Huber (1981).

Formally, the robust estimate of the SDM is given by

$$\mathbf{S} = (\boldsymbol{\Psi}\boldsymbol{\Psi}^*)^{-1}, \tag{12}$$

where $\boldsymbol{\Psi}$ maximizes (10). In practice, the estimate can be computed as a weighted cross-product matrix,

$$\mathbf{S} = \sum_{i=1}^{I} w_i \mathbf{X}_i \mathbf{X}_i^*, \tag{13}$$

with the weights determined by the data:

$$w_i = \frac{-f'(|\mathbf{Y}_i|)}{2|\mathbf{Y}_i| f(|\mathbf{Y}_i|)}, \quad \text{where } \mathbf{Y}_i = \mathbf{S}^{-1/2}\mathbf{X}_i. \tag{14}$$

The computation thus proceeds iteratively, much as for the regression M-estimate used by Egbert & Booker (1986), with the Cholesky decomposition of the weighted SDM **S** from the previous iteration used to calculate the weights of (14) for the next update.

3.2 A robust MEV estimator

The robust SDM estimate described above is invariant under all rotations of the data vectors (in the full K-dimensional data space). This means that outliers are assumed to be equally likely in all directions. In fact, outliers are much more likely to occur in specific directions in the data space. For example, there may be outliers due to coherent noise which occur only in a relatively low-dimensional subspace (dictated by the geometries of noise sources and the MT array). By identifying this 'coherent noise subspace', the performance of a robust estimator might be enhanced. On the other hand, instrumental problems or localized sources of noise might be expected to cause outliers in individual channels. Under these circumstances it would make more sense to focus on outliers in the coordinate directions of the measured data. We consider this simpler case first, putting off the problem of explicitly allowing for coherent noise outliers until Sections 5 and 6.

To allow for outliers in individual channels we develop a robust estimate for the MEV model of (7). This means that initially we treat coherent noise as part of the signal. To start we assume that M and σ_k^2, $k = 1, K$, are known. In practice, these parameters will of course have to be estimated, an issue to which we return below. By transforming the data as in (8) we may assume that $\boldsymbol{\Sigma}_N = \mathbf{I}$. For the case of isotropic noise, the MEV estimates for **W** can be obtained by minimizing (Gleser 1981)

$$\sum_{ik} \left| X_{ik} - \sum_{m=1}^{M} \alpha_{im} W_{km} \right|^2, \tag{15}$$

over **W** and the generalized polarization parameters $\boldsymbol{\alpha}_i$, $i = 1, I$. This minimization can be accomplished by solving an eigenvalue problem, or by an alternating conditional least-squares algorithm (Egbert 1991). Emulating the regression M-estimate (Egbert & Booker 1986; Chave *et al.* 1987), we can replace the quadratic misfit of (15) by the more general

loss function

$$\sum_{ik} \rho\left(X_{ik} - \sum_{m=1}^{M} \alpha_{im} W_{km} \right). \tag{16}$$

Minimizing (16) over the unknowns W_{km} and α_{im} leads to a series of coupled weighted least-squares normal equations:

$$\sum_{k} w_{ik} r_{ik}^{*} W_{km} = 0 \qquad i = 1, I \quad m = 1, M, \tag{17}$$

$$\sum_{i} w_{ik} r_{ik}^{*} \alpha_{im} = 0 \qquad k = 1, K \quad m = 1, M, \tag{18}$$

where r_{ik} and w_{ik} are the residuals and weights:

$$r_{ik} = X_{ik} - \sum_{m=1}^{M} \alpha_{im} W_{km}, \qquad w_{ik} = \rho'(r_{ik})/r_{ik}. \tag{19}$$

Defining

$$\tilde{Z}_{ik} = w_{ik} X_{ik} + (1 - w_{ik}) \sum_{m=1}^{M} \alpha_{im} W_{km} = w_{ik} X_{ik} + (1 - w_{ik}) \hat{X}_{ik}, \tag{20}$$

a simple calculation shows that the normal equations (18)–(19) can also be written

$$\sum_{k} \left[\tilde{X}_{ik} - \sum_{m'=1}^{M} \alpha_{im'} W_{km'} \right]^{*} W_{km} = 0 \qquad i = 1, I \quad m = 1, M, \tag{21}$$

$$\sum_{i} \left[\tilde{X}_{ik} - \sum_{m'=1}^{M} \alpha_{im'} W_{km'} \right]^{*} \alpha_{im} \qquad k = 1, K \quad m = 1, M. \tag{22}$$

In (21)–(22), $\tilde{X}_{ik}$ can be interpreted as the kth component of a 'cleaned data vector' $\tilde{\mathbf{X}}_i$. Each component is a weighted average of predicted and observed data. Note that with the assumption that $\mathbf{W}^{*}\mathbf{W} = \mathbf{I}$, (21) implies that the estimates of $\boldsymbol{\alpha}_i$ and $\mathbf{W}$ satisfy

$$\hat{\alpha}_{im} = \sum_{k=1}^{K} \hat{W}_{km}^{*} \tilde{X}_{ik}. \tag{23}$$

That is, the estimated generalized polarization parameters are just the inner product of the cleaned data vectors $\tilde{\mathbf{X}}_i$ with the columns of $\hat{\mathbf{W}}$.

The following iterative scheme for minimizing (16) is thus suggested.

(1) Estimate the SDM using the rotationally invariant robust approach of the previous subsection, and compute a preliminary estimate of $\mathbf{W}$ using the dominant M eigenvectors.

(2) Use the estimate of $\mathbf{W}$ and the raw data vectors in (23) to estimate $\boldsymbol{\alpha}_i$, $i = 1, I$.

(3) Calculate residuals, weights, and cleaned data vectors using (19) and (20).

Using the cleaned data vectors from step (3), we can then repeat steps (1)–(3), iterating to convergence.

Note that after step (2) is completed, the estimates of $\mathbf{W}$ and $\boldsymbol{\alpha}_i$, $i = 1, I$, for the current iteration satisfy (21) and (22), but with the cleaned data vectors calculated from the previous iteration. To see this, note that (21) and (22) are the normal equations for the least-squares problem (15), but with the actual data vectors replaced by $\tilde{\mathbf{X}}_i$. $\mathbf{W}$ is computed from the eigenvectors of the corresponding 'cleaned SDM', and thus, with $\boldsymbol{\alpha}_i$ defined through (23), satisfies (21) and (22) (Gleser 1981). Thus, if the iterations converge, the resulting $\mathbf{W}$ and $\boldsymbol{\alpha}_i$ must satisfy (20)–(22) and hence (17)–(18). We do not know what conditions will guarantee either that this scheme will converge to a solution to the normal equations, or that any resulting solution will be a global minimum of (16). However, using a Huber-type penalty functional, modified for complex residuals, i.e.

$$\rho(r) = \begin{cases} |r|^2 & \text{if} \quad |r| \le r_0, \\ r_0 |r| & \text{if} \quad |r| > r_0, \end{cases} \tag{24}$$

the procedure generally seems to be well behaved. Convergence to an (apparently reasonable) solution is generally rapid. We have tried a range of values of r_0 in (24); $r_0 = 1.4$ seems to work reasonably well.

3.3 Estimation of incoherent noise variances

An initial estimate of σ_k^2, $k = 1, K$, can be based on residual variances obtained from fitting each channel to the remaining $K - 1$ channels. The iterative robust MEV estimation scheme suggests an alternative approach. Note first that, after the first iteration of steps (1)–(3) above, we have estimates of the polarization parameters $\boldsymbol{\alpha}_i$, and for subsequent iterations step (1) can be replaced by the following.

(4) Using the cleaned, but unnormalized, data vectors $\tilde{\mathbf{X}}_i$, and the current estimates of the polarization vectors $\boldsymbol{\alpha}_i$, $i = 1, I$, solve the normal equations (22) for $\mathbf{W}$. Use Gram–Schmidt orthogonalization to make the columns of the new estimate orthonormal. With the new estimate of $\mathbf{W}$ return to step (2).

Step (4) corresponds exactly to doing a multiple linear regression of each (cleaned) data channel on the current estimates of the polarization parameters. This suggests that residual variances from these regressions could be used to provide approximate estimates of σ_k^2, $k = 1, K$. We use a variant on this idea, which also incorporates elements of our simple initial estimate of incoherent-noise variances.

Again, for channel k we restrict attention to the remaining $K - 1$ channels. Using these reduced $(K - 1)$-dimensional data vectors we execute steps (1)–(3), obtaining estimates of $\hat{\boldsymbol{\alpha}}_{ik}$, $i = 1, I$. The estimated polarization vectors (which are slightly different for each channel) are then used as the predictor variable in step (4) to compute the incoherent residual variance for channel k. The resulting set of K residual variances will in general be biased estimates of σ_k^2, because noisy estimates, rather than the true polarization vectors $\boldsymbol{\alpha}_i$, are used as the predictor variables. However, once the residual variances have been computed, approximately unbiased estimates of the incoherent-noise variances can be computed by solving a $K \times K$ system of linear equations. Details are given in Appendix A.

This variance estimation scheme can be made more robust in two ways. First, the regression step (4) can be accomplished with a robust scheme in which outliers in channel k are pulled towards predicted values. Second, after data vectors have been 'cleaned' for all channels, polarization parameters can be re-estimated by repeating (2), and the calculation of incoherent-noise variances can be repeated.

Three aspects of our scheme for estimating σ_k^2 deserve further comment. First, using the estimated polarization parameters $\hat{\boldsymbol{\alpha}}_{ik}$, $i = 1, I$, to predict each channel is significantly more stable than using the original data vectors $X_{ik'}$, $k \neq k'$, $i = 1, I$. Our approach is essentially identical to regression on principal

components (e.g. Hawkins 1977; Seber 1984), a commonly used way to stabilize regression when the predictor variables are highly correlated (as we should expect of the components of the original data vectors). Second, until we have good estimates of σ_k^2, accurate estimation of M is difficult, while to estimate the polarization vectors we apparently need to know M. Fortunately, as we show in Appendix A, the estimates of incoherent-noise variances based on the estimated polarization vectors $\hat{\mathbf{a}}_{ik}$ (computed by excluding channel k), should be roughly correct provided our estimate of M is at least as large as the true M. With the proposed scheme it is thus not necessary to have a good estimate of M to proceed with the initial estimation of the incoherent-noise variances.

Finally, it is relatively straightforward to generalize our incoherent noise model to allow for noise which is coherent within a single site, but incoherent between sites. In this model, $\mathbf{\Sigma}_N$ would be block-diagonal, with all inter-station blocks identically zero. To estimate this more complicated covariance matrix we compute incoherent residual vectors for site j by predicting all channels at this site using the remaining $5(J-1)$ channels for the other $J-1$ sites for the predictor variables. With this alternative model we can explicitly distinguish between noise which is coherent between sites, and noise which is coherent within a single site.

3.4 Summary of the RMEV estimation scheme

We summarize the full procedure for estimating the unknowns σ_k^2, M, and $\mathbf{W}$.

(a) Compute the robust scaled SDM $\mathbf{S}' = \mathbf{\Sigma}_N^{-1/2}\mathbf{S}\mathbf{\Sigma}_N^{-1/2}$, using the current version of the cleaned data vectors $\tilde{\mathbf{X}}_i$, and the current estimate of σ_k^2, $k = 1, K$. To initialize, just use the raw data and unit noise variances.

(b) For each component k extract the L dominant eigenvectors of the submatrix of $\mathbf{S}'$ formed by deleting the kth row and column. Initially, choose $L = K-1$; subsequently, choose L conservatively to be at least as large as the estimated M. Using these eigenvectors of the reduced SDM, and the scaled data components $X'_{ik'} = X_{ik'}/\sigma_{k'}$, $k' \neq k$, do step (2) once, then iterate steps (3) and (4) to convergence. Save the final cleaned data components $\tilde{X}_{ik}$, along with estimates of residual noise variance for each channel. Residual noise variances are corrected to allow for the downweighting of large residuals as described in Egbert & Booker (1986, Appendix A), and for the bias effects discussed in Appendix A.

(c) Return to step (a) using the updated cleaned data vectors and incoherent-noise variances. Steps (a) and (b) can be iterated to convergence. A few (three or so) steps generally seem to be sufficient.

(d) Now do step (a) once more, using the cleaned data and estimates of σ_k^2, $k = 1, K$, from the final iteration of (b). Extract the eigenvectors from the full matrix $\mathbf{S}'$, providing initial estimates of M and $\mathbf{W}$.

(e) Iterate steps (2)–(4) above (with σ_k^2 fixed) to refine the estimates of M and $\mathbf{W}$.

We will refer to the full scheme outlined in steps (a)–(e) as the Robust Multivariate Errors-in-Variables (RMEV) estimator.

3.5 Calculation of error bars

When $M = 2$, the usual MT impedances and inter-station transfer functions can be calculated from the estimates of $\mathbf{U}$ ($= \mathbf{W}$), by using (2). To calculate error bars for these estimates, we apply the asymptotic (large sample) results of Gleser (1981), as modified for complex data by EB. To be explicit, we consider error bars for transfer functions using the horizontal magnetic fields at the first site as a normal field reference. By suitable reordering of components, these formulae can be applied to compute error bars for any inter-station or inter-component transfer functions. Throughout this subsection, we use primes to denote quantities expressed in the non-dimensional units defined by the incoherent-noise standard deviations σ_k^2.

Results are most simply expressed in terms of the $K \times 2$ matrices $\mathbf{U}$ and $\mathbf{U}'$ defined in (4) and (5). The estimated transfer functions for all channels relative to the two reference channels are given by the $(K-2) \times 2$ complex matrix $\hat{\mathbf{T}} = \mathbf{U}_2\mathbf{U}_1^{-1}$, where $\mathbf{U}_1$ is the 2×2 sub-matrix consisting of the first two rows of $\mathbf{U}$, and $\mathbf{U}_2$ contains the remaining $K-2$ rows. In non-dimensional incoherent noise units, the transfer function estimates are of course $\hat{\mathbf{T}}' = \mathbf{U}_2'\mathbf{U}_1'^{-1}$. Define the $(K-2) \times (K-2)$ matrices $\mathbf{\Sigma}_R'$ and $\mathbf{\Sigma}_R$ as

$$\mathbf{\Sigma}_R' = \mathbf{I}_{K-2} + \hat{\mathbf{T}}'\hat{\mathbf{T}}'^*,$$

$$(\mathbf{\Sigma}_R)_{ll'} = \sigma_{l+2}\sigma_{l'+2}(\mathbf{\Sigma}_R')_{ll'}, \quad l, l' = 1, K-2. \tag{25}$$

Next, define the 2×2 matrices

$$\mathbf{\Sigma}_H' = \mathbf{U}_1'\left[I^{-1}\sum_{i=1}^{I}\hat{\mathbf{a}}_i\hat{\mathbf{a}}_i^* - \mathbf{I}_2\right]\mathbf{U}_1'^*,$$

$$(\mathbf{\Sigma}_H)_{mm'} = \sigma_m\sigma_{m'}(\mathbf{\Sigma}_H')_{mm'}, \quad m, m' = 1, 2. \tag{26}$$

After some algebraic manipulation, the results given in EB for the large sample approximation to the covariance of the elements of $\hat{\mathbf{T}}$ can be expressed as

$$\mathrm{Cov}(\hat{T}_{km}\hat{T}_{l'm'}) = I^{-1}[(\mathbf{\Sigma}_H^{-1})_{mm'} + (\sigma_m\sigma_{m'})^{-1} \times (\mathbf{\Sigma}_H'^{-1}\mathbf{U}_1'\mathbf{U}_1'^*\mathbf{\Sigma}_H'^{-1})_{mm'}] \times (\mathbf{\Sigma}_R)_{ll'}, \tag{27}$$

$m, m' = 1, 2; \quad l, l' = 1, K-2.$

The terms in this expression have a relatively simple interpretation, which suggests a simple modification to allow at least approximately for downweighting of data by the robust procedure. First, it can be easily shown that $\mathbf{\Sigma}_R$ is an approximation to the covariance matrix of the predicted field component residuals $X_{ik} - \Sigma_{m=1}^2 T_{km}X_{im}$, $k = 3, K$ [the approximation, which comes from using the estimated transfer functions in (25), improves as the sample size I increases]. The second term in the product (subscripted by ll') thus characterizes the combined noise in the predicted and predicting channels. On the other hand, $I^{-1}\mathbf{\Sigma}_H$ provides an estimate of the total signal power covariance in the reference magnetic fields, while the non-dimensional 2×2 matrix $\mathbf{\Sigma}_H'$ provides an estimate of the average array signal-to-noise ratio. It can thus be shown that as signal-to-noise ratios increase, the first term in the product can be approximated as $(\mathbf{\Sigma}_H^{-1})_{mm'}$.

For transfer functions between a single output channel and two noise-free input magnetic channels, the usual least-squares expression for error covariance is

$$\mathrm{Cov}(T_m T_{m'}) = \sigma^2\left[\left(\sum_{i=1}^{I}\mathbf{h}_i\mathbf{h}_i^*\right)^{-1}\right]_{mm'}, \tag{28}$$

where σ^2 is the noise variance for the predicted output channel, and the cross-product matrix gives the total horizontal magnetic field power. Clearly, (27) bears a close resemblance to (28); for example, $(\mathbf{\Sigma}_R)_{kk}$ (variance of residuals for channel k)

in (27) should be identified with σ^2 in (28). For the robust regression M-estimate, (28) should be changed to allow for the downweighting of data. Following Huber (1981), Egbert & Booker (1986) show this can be accomplished by (1) using the cleaned data in the calculation of the residual variance [i.e. the estimate of σ^2 in (28)]; and (2) multiplying the right-hand side of (28) by a correction factor $\chi = [\mathbf{E}\rho'']^{-2}$, where ρ'' is the second derivative of the loss function used, and $\mathbf{E}$ is the expectation operator. For the Huber loss function of (24), $\mathbf{E}\rho''$ can be estimated by the fraction of 'good' data points (with residuals magnitudes less than r_0). See Egbert & Booker (1986) for details. We use a similar approach to allow for downweighted data in the RMEV estimate. Specifically, we use cleaned data to calculate $\mathbf{\Sigma}_R$, and multiply the right-hand side of (27) by the appropriate estimated correction factor χ.

These error bars are only approximately correct, for several reasons. First, in contrast to the mathematically rigorous development followed for the standard M-estimate, our treatment here is purely heuristic. Second, we make no allowance for uncertainties in the estimated incoherent noise variances. Third, the asymptotic error bars proposed by Gleser (1981) strictly require fourth moments of the error distribution. In (27), these are approximated by using estimated second moments to calculate fourth moments under the assumption of a Gaussian error distribution. Note, however, that the fourth moments enter the error expression through $(\mathbf{\Sigma}_H'^{-1}\mathbf{U}_1'\mathbf{U}_1'^{*}\mathbf{\Sigma}_H'^{-1})_{mm'}$, a term which we have argued will generally be small. Failure of the Gaussian error assumption may thus not be so serious.

Further development of improved error estimates [for example using jackknife methods as proposed by Chave & Thomson (1989)] is clearly warranted, but is beyond the scope of this paper.

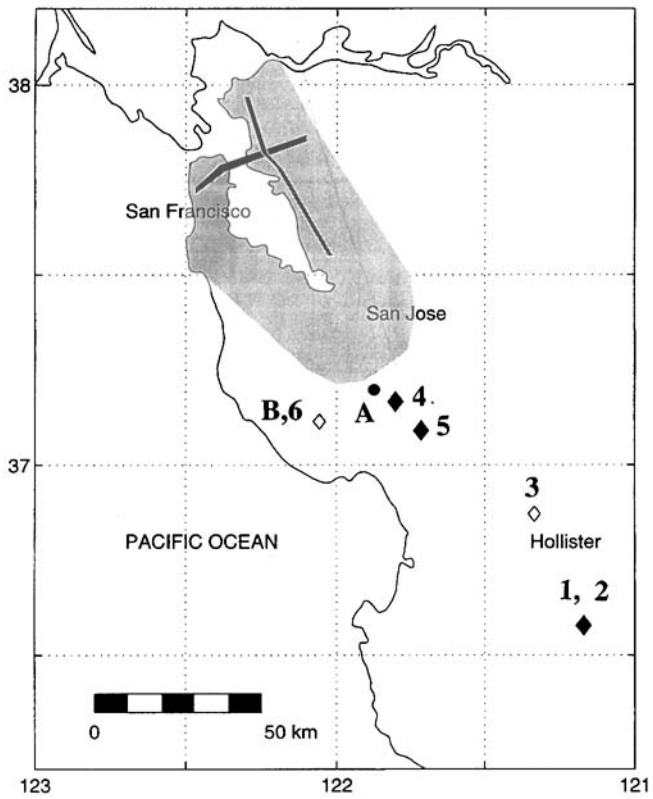

Figure 2. Map showing locations of sites A and B referred to in Fig. 1, as well as the three stations (1, 4, and 5) used for the example three-station array. Shading denotes urban areas; tracks for the Bay Area Rapid Transit (BART) DC electric railway are marked by a heavy line.

4 EXAMPLES

To illustrate the methods of the previous section we consider data from the three-station array 1–4–5 plotted in Fig. 2. At all sites, standard 5-component MT data were collected, so $K = 15$. For our examples we consider data from two sampling bands: 120 Hz data, collected in a series of short runs (total duration: ≈ 1000 s) at a range of times during daylight hours; and 10 Hz data sampled continuously for approximately 12 hours. In Fig. 3 we plot incoherent noise power spectra (i.e. estimates of σ_k^2, $k = 1, K$, as a function of period) for selected channels (k): the three H_y channels for the 120 Hz band, and the three E_y channels for the 10 Hz band. In both cases we present estimates of incoherent-noise power obtained both with and without the iterative cleaning used for the RMEV estimate. The raw data estimates (Figs 3a and b) thus give power spectra for the total incoherent-noise process (including isolated outliers), while the RMEV estimates (Figs 3c and d) give power spectra for the noise processes with outliers cleaned up.

Differences between the raw data and RMEV spectra provide clear evidence for the effectiveness of the multivariate robust procedures. Many of the localized peaks seen in the non-robust noise spectra of Figs 3(b) and 4(d) are reduced or even eliminated by the robust scheme. In particular, the prominent peak near 0.05 s in the station 1 H_y noise spectrum is reduced by two orders of magnitude in the RMEV spectrum. In this frequency band, noise at site 1 was thus dominated by an intermittent source, which could be effectively filtered out using the RMEV scheme. As we shall show below, estimates of MT impedances near 0.05 s are significantly improved using the RMEV estimate. Other peaks in the H_y noise spectra (at 0.7 s for station 5, and at 0.2 s for station 4) are also significantly reduced (by a factor of 2–3) with the RMEV estimator. Note also that there is a general tendency toward a reduction in overall noise levels in the RMEV curves for H_y.

For the 10 Hz sampling band, there are peaks near a period of 100 s in the raw data E_y noise spectra for all sites. These are eliminated by the robust scheme. This rather curious source of noise (incoherent between stations, but apparently present at all sites) is also only present sporadically. The small peak in noise power at a period of 1 s for E_y at site 1 is also removed by the RMEV estimator. Of course in many cases robust and raw data noise spectra will be similar. Indeed, this is the case for some of the other (unplotted) data channels for our example arrays. However, at least in some cases, the robust iterative scheme enhances signal-to-noise ratios, and as a result can improve estimates of MT impedances.

In Fig. 4 we plot, as a function of period, the ordered eigenvalues of the non-dimensional SDM $\mathbf{S}'$. The number of eigenvalues significantly greater than one provides an estimate of the true coherence dimension M. If there is no coherent noise we should find $M = 2$. To emphasize this in Fig. 4, curves for the two largest eigenvalues are plotted with solid lines, with the remaining $K - 2$ smaller eigenvalue curves dashed.

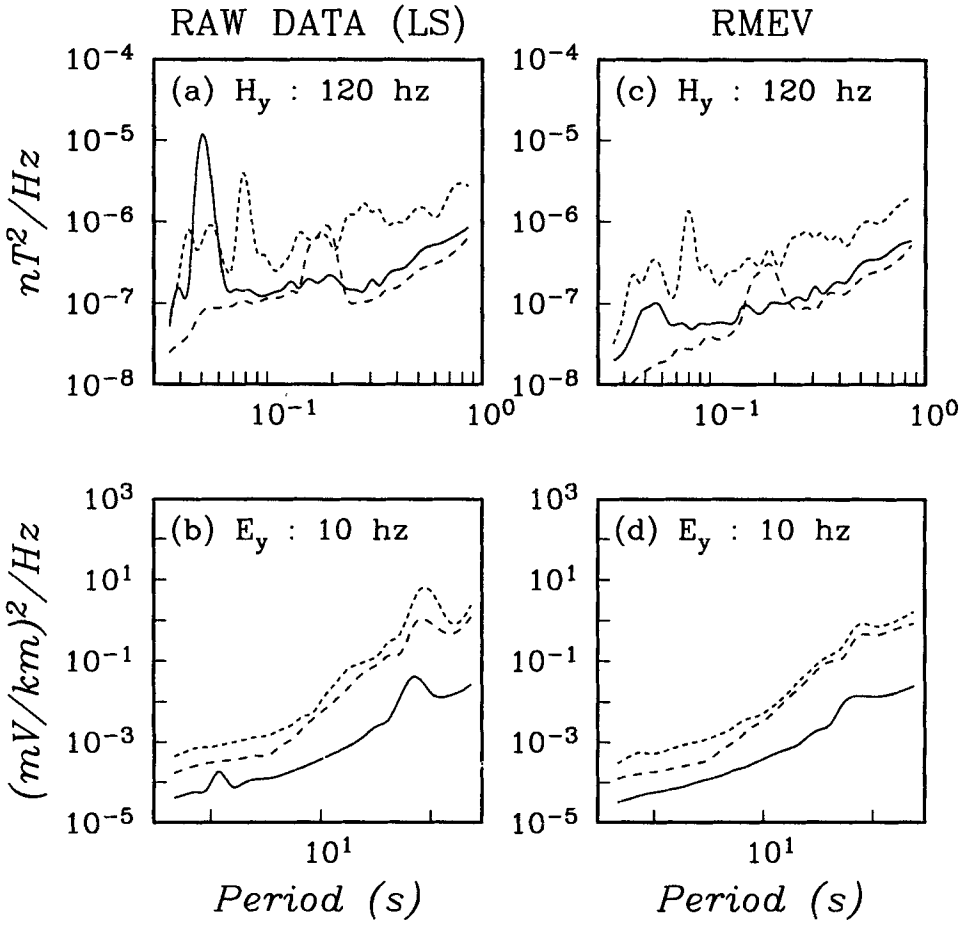

Figure 3. Examples of estimated noise spectra (i.e., σ_k^2 as a function of period) computed for selected channels from the three-station example array. In all panels solid lines give the spectrum for station 1, long dashes that for station 4, and short dashes that for station 5. In (a) raw-noise spectra are plotted for the three H_y channels from the 120 Hz sampling band. (c) gives noise spectra for the three E_y channels from the 10 Hz sampling band. For (b) and (d) spectra for the same channels were computed using the data with outliers cleaned up by the RMEV estimate. Some localized peaks in the raw spectra are eliminated by the robust estimator, implying that these peaks result from intermittent noise sources, which are effectively filtered out by the robust scheme.

When the number of large eigenvalues exceeds two, we have evidence for coherent noise of some sort. There are large sample hypothesis tests for the statistical significance of large eigenvalues in the case where signal and noise jointly have a multivariate Gaussian distribution (e.g. Morrison 1967; Giri 1977; EB). Given the strongly non-Gaussian nature of both signal and noise, and given that we must estimate $\sigma_k^2, k = 1, K$, these tests can only provide rough guidance at best. Better ways to test the coherence dimension of the data rigorously would be a useful development. For now we adopt a more qualitative exploratory approach, noting that with the data scaling of (8) the eigenvalues are non-dimensional, and can be interpreted as a signal-to-noise power ratio. Thus, quite aside from the question of statistical significance, we can at least qualitatively assess the relative magnitude of possible coherent noise in the MT array data.

For the 120 Hz sampling band $M \approx 6$ at the shortest periods. The magnitudes of these secondary eigenvalues decrease steadily with period until they merge with the background incoherent noise levels at $T \approx 0.2$ s. For longer periods in this sampling band $M = 2$, indicating that the uniform-source MT assumption is satisfied quite well. The short-period coherent noise turns out to have a rather mundane explanation. A bug in the data acquisition software caused sporadic small random errors in the actual sampling time for data at each site. The result is an apparent source of noise which is coherent between all components at a site, but completely incoherent between sites. This property of the noise can be verified by generalizing our model for $\mathbf{\Sigma}_N$ to allow for locally coherent noise, as sketched above (and in Appendix A). When the data are rescaled with the resulting estimate of $\mathbf{\Sigma}_N$, we find that the coherence dimension of the array data is two for the full 120 Hz sampling band (Fig. 4b).

Note also that in Figs 4(a) and (b) there are three broad peaks in the dominant eigenvalue spectra (at periods of 0.125, 0.07 and 0.05 s), where signal-to-noise ratios are enhanced. These peaks correspond to the Schuman resonances of the Earth-ionosphere cavity (e.g. Matsushita & Campbell 1967), where an enhancement of signal power should be expected.

For the 10 Hz sampling band (Fig. 4c) the coherence dimension M exceeds two for all periods longer than about 2 s. Indeed, for periods of 5 to 30 s, the second and third largest eigenvalues are nearly equal, so that the distinction between the largest two eigenvalues and the remaining 13 is no longer

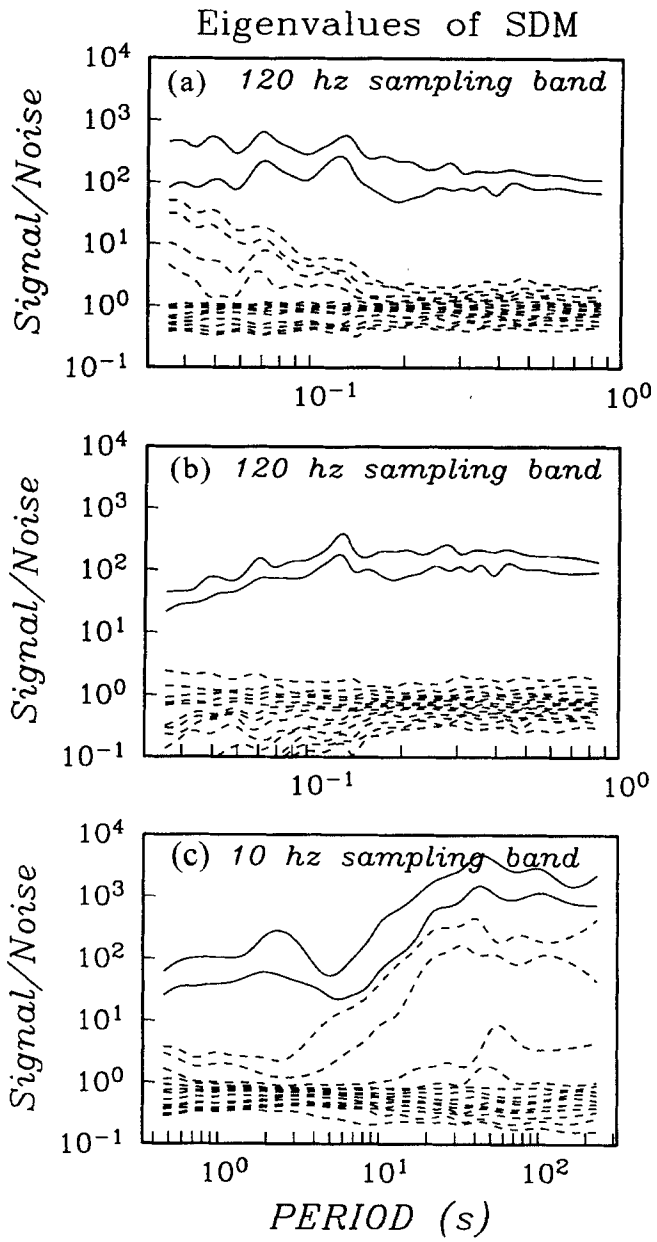

Figure 4. Eigenvalues of the scaled SDM $\mathbf{S}'$ for the three-station example array for (a)–(b) the 120 Hz sampling band, and (c) the 10 Hz sampling band. All channels are scaled into non-dimensional units by dividing by the estimated incoherent-noise standard deviations, so the eigenvalues are non-dimensionalized and can be interpreted as signal-to-noise power ratios. The number of eigenvalues significantly greater than unity thus gives an estimate of the effective coherence dimension M of the array signal. For the case of quasi-uniform sources contaminated only by incoherent noise $M = 2$. To emphasize this the two dominant eigenvalues are plotted with solid lines. In (a), where we have assumed a diagonal noise covariance, we have evidence for coherent noise at the shortest periods. However, generalizing the model for $\mathbf{\Sigma}_N$ to allow for noise that is locally coherent (b) demonstrates that this noise is not coherent between stations. In fact, the coherent noise at short periods can be traced to a problem with timing synchronization between stations. For this sampling band $M = 2$ for all periods, and there is no real problem with coherent noise. For the 10 Hz sampling band, (c), coherent noise is significant for periods longer than a few seconds. From about 5 to 30 s period the second and third largest eigenvalues are nearly equal, making the distinction between MT signal and coherent noise problematic.

truly justified. Allowing for locally coherent noise has little effect on the appearance of Fig. 4(c), demonstrating that the additional significant eigenvalues indeed result from sources of noise which are coherent between stations.

The $M = N + 2$ dominant eigenvectors of the scaled SDM $\mathbf{S}'$ provide us with an estimate of the span of the columns of the $K \times M$ matrix $\mathbf{W}$ [via the analogue of (4)]. When $M = 2$ ($N = 0$), $\mathbf{W}$ is identical to $\mathbf{U}$ and we can proceed to calculate the usual MT transfer functions. To assess the performance of the RMEV estimate in this simpler case where there is no coherent noise, we first consider a synthetic data example.

In Fig. 5 we compare the performance of the RMEV with a robust RR (RRR) estimate. This estimate is based on the approach of Egbert & Booker (1986), with the addition of procedures for leverage control, and generalized to the RR case as suggested by Chave & Thomson (1989). For this example, 1 Hz data were generated for two 5-channel sites over a 100 Ω m half-space. Random noise from a heavy-tailed distribution was then added (in the frequency domain) to all data channels. When compared to non-robust single-station and RR estimates (not shown), the RRR estimate (Fig. 5a) performs reasonably well. However, the RMEV estimates (Fig. 5b) are significantly better. There are two significant differences between the RRR and RMEV estimation schemes, which probably explain most of the difference in performance. First, there are outliers in all channels, including the RR channels. In contrast to the RMEV estimate, the RRR estimate does not specifically allow for outliers in the remote. Second, the RMEV estimator uses all data channels to define an optimal set of reference fields.

In Fig. 6 we compare RRR and RMEV estimates of apparent resistivities and phases for station 5, using actual data from the 120 Hz sampling band. For the RRR estimate we used horizontal magnetic fields from station 1 as a reference, while the RMEV estimator was applied to the two-station array consisting of stations 1 and 5. Improvements are significant with the array approach. Note in particular the improvement in the estimates at a period of $T \approx 0.05$ s. This is, in fact, the period where the RMEV scheme eliminated the peak in the H_y noise spectrum at site 1 (see Fig. 3, and the associated discussion). Clearly, outliers in the magnetic fields at site 1, which were used as the sole reference in the RRR estimate, are seriously degrading the RRR impedance estimates. Note that in this case similar results are obtained when all three sites are included in the multiple-station analysis.

For the 120 Hz sampling band there is no coherent noise, and we find that the RMEV impedance estimates are quite good. However, the eigenvalues plotted in Fig. 4(c) suggest that coherent noise will be a significant problem for periods beyond a few seconds in the 10 Hz sampling band. This is confirmed in Fig. 7, where we plot robust single-station and RMEV results for site 5. The single-station curves are fairly smooth, but ρ_a is locally very steep (for one mode the increase is more than an order of magnitude between 5 and 10 s), and the phase decreases suddenly (to negative values for one mode) at a period of 7–8 s. This apparently non-physical behaviour of the impedance from 4 to 40 s is consistent with the coherent noise implied by Fig. 4(c).

To apply the RMEV estimates here we have a problem: how do we decide which 2-D slice of the M-dimensional coherent data space (i.e., the span of the columns of $\mathbf{W}$) defines the

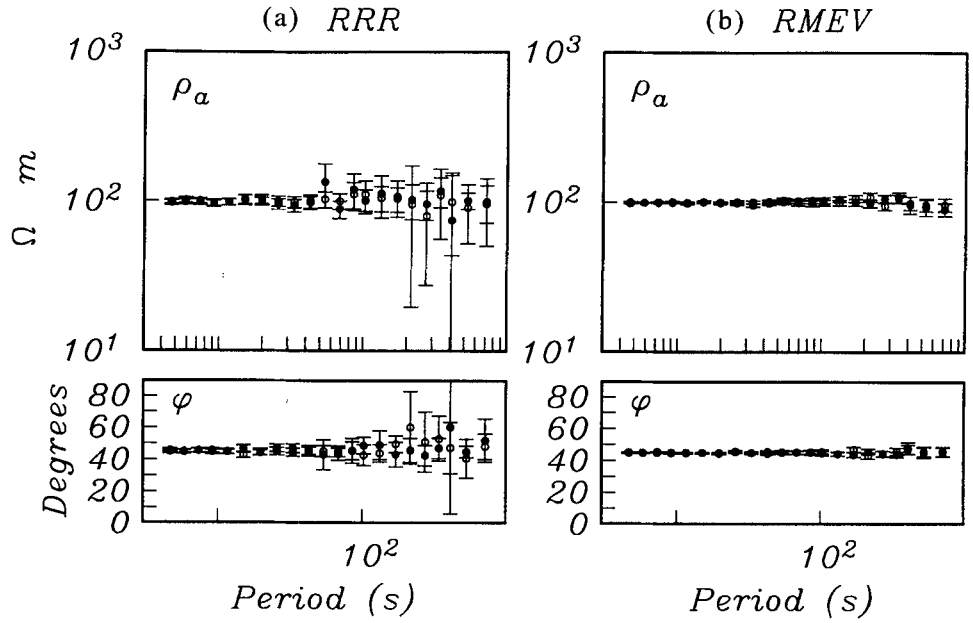

Figure 5. Apparent resistivity and phase estimates computed from very noisy synthetic MT data using (a) RRR, and (b) RMEV estimates. For this comparison synthetic data were generated for two stations over a 100 Ω m half-space, with noise from a heavy-tailed distribution added to all data channels (in the frequency domain). The RMEV estimate provides dramatically improved estimates by cleaning up outliers in all data channels, including those for the remote.

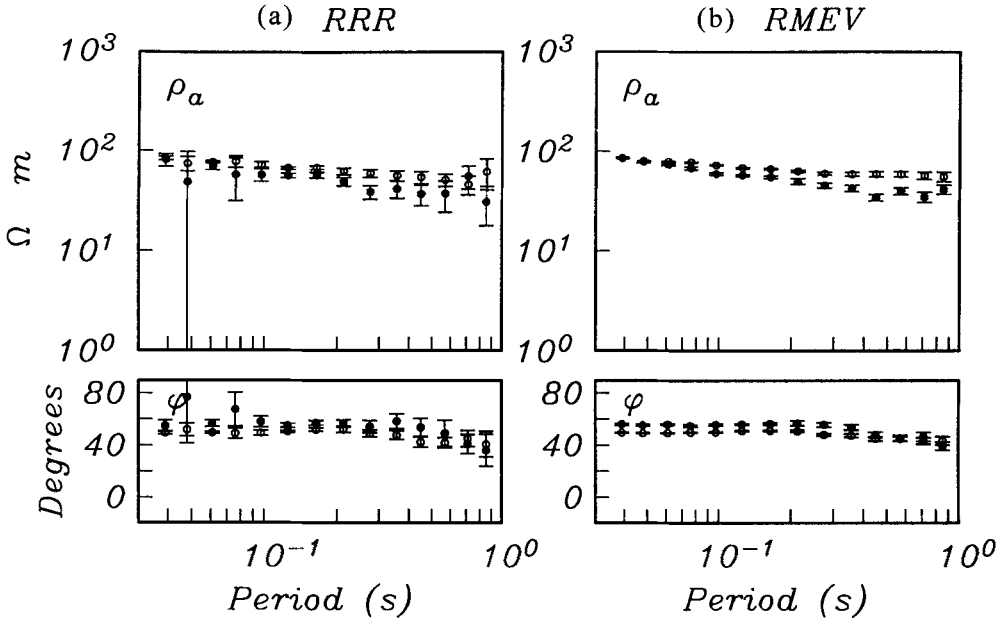

Figure 6. Apparent resistivity and phase estimates for site 5 from the three-station example array (120 Hz sampling band). (a) RRR estimates for site 5 using site 1 as a reference. (b) RMEV estimates for the two station array consisting of sites 1 and 5. Improvements are significant with the new approach. Note especially the improvement near $T = 0.05$ s, where the RMEV scheme was able to eliminate the peak in the H_y noise spectrum at site 1 (see Fig. 3). Clearly, outliers responsible for the peak in the noise spectra are seriously degrading the RRR estimates of (a).

response of the Earth to uniform sources? For much of the frequency range here, $M = 4$ (Fig. 4c). We plunge ahead blindly and assume that the signal dominates, so that the two dominant eigenvectors of the scaled SDM at least approximately define the desired uniform-source space. The results, given for site 5 in Fig. 7(b), are disappointing but hardly surprising. The non-physical behaviour is similar to that obtained with single-station processing, but the RMEV estimates are no longer smooth. This erratic behaviour is worst where the second and third eigenvalues are nearly degenerate (4–50 s). In this period range, the second eigenvector used for the RMEV estimate of **U** is obviously ill-defined, allowing the mix of coherent noise and signal in our naive estimate to vary rapidly with period.

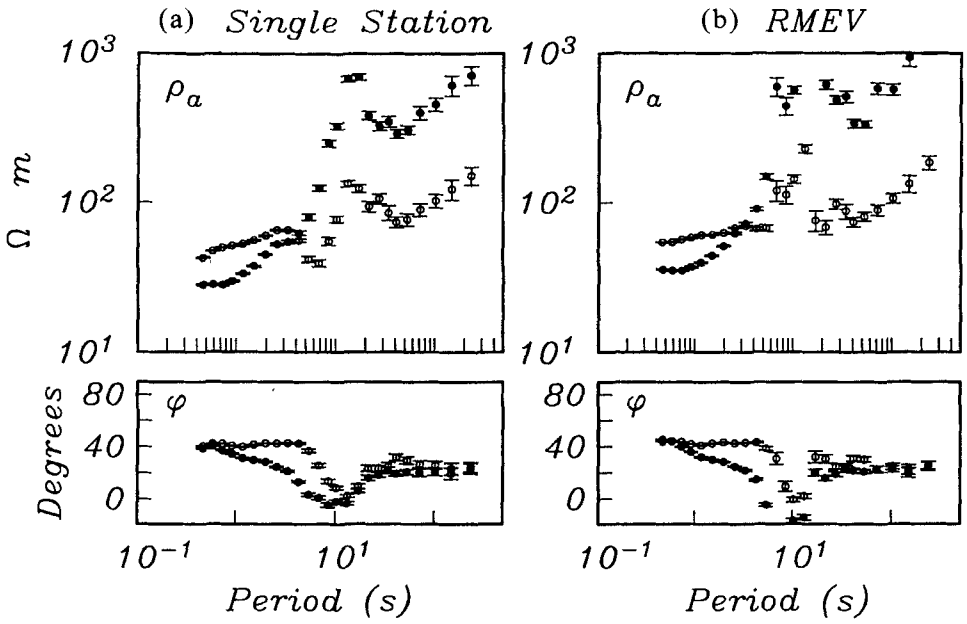

Figure 7. (a) Robust single station and (b) RMEV apparent resistivity and phase estimates for site 5, for the 10 Hz sampling band. In both cases the curves exhibit non-physical behaviour in the period range 4–50 s. To compute the RMEV estimates here we have assumed that the eigenvectors associated with the largest two eigenvalues of the scaled SDM define the MT source. Clearly this is not so, as suggested already by Fig. 4(c).

5 SEPARATION OF MT SIGNAL AND COHERENT NOISE

The RMEV estimator developed above provides an estimate of M, the coherence dimension of the array data. If $M = 2$, it also provides an estimate of $\mathbf{U} = \mathbf{W}$, equivalent to all possible inter-component and inter-station transfer functions. As we have shown, these estimates improve on the more standard robust RR transfer function estimates in some cases. Furthermore, if M exceeds 2, we are clearly warned of the presence of coherent noise. With $M > 2$, however, $\mathbf{W}$ is a mixture of coherent noise ($\mathbf{V}\boldsymbol{\gamma}$) and the desired MT signal ($\mathbf{U}\boldsymbol{\beta}$), and it is unclear how (or if) we can obtain reasonable estimates of the MT parameters.

The situation is illustrated in Fig. 8, where we plot the field components corresponding to each of the columns of $\mathbf{W}$, as estimated for our example three-station array at a period of 10 s, where $M = 4$. The horizontal magnetic fields in this plot exhibit large variations between the southernmost site 1 and sites 4 and 5. For MT sources, the horizontal magnetic field components should be approximately constant across the array, since internal anomalous fields are typically significantly smaller than the uniform primary (or normal) total fields. The large deviations from uniformity seen here suggest that coherent noise dominates the MT signal.

Although there is no general completely fool-proof method for separating MT signal from coherent noise, in some circumstances such a separation may be possible. We consider two particular cases. In the first, we make use of a quiet reference site to effect the separation. In the second case, considered in Section 6, we take advantage of intermittency in the coherent noise.

5.1 Use of a quiet remote

Site 1 in our three-station example array is 100 km south of the other two sites. Several lines of evidence suggest that this site is relatively unaffected by the coherent noise, which clearly contaminates the MT data at stations closer to the San Francisco Bay area (Fig. 7). First, single-station estimates for this site (not shown) are generally well behaved. Second, canonical coherence analysis shows that the dimensionality of coherent signal between site 1 and the other two sites is two.

The simplest way to use a quiet site to separate the MT signal from the coherent noise is to use the usual RR method. Here we consider some generalizations which allow us to make better use of all data channels. In essence, all sites can first be used to reduce the effects of incoherent noise, and then the quiet site can be used to separate the MT signal from the coherent noise. Consider again (7), with the MT signal represented by $\mathbf{U}\boldsymbol{\beta}_i$, and the coherent noise represented by $\mathbf{V}\boldsymbol{\gamma}_i$. Recall that the MT ($\boldsymbol{\beta}_i$) and coherent noise ($\boldsymbol{\gamma}_i$) polarization vectors are taken to be statistically independent. We also assume that at the first site the horizontal magnetic field components are unaffected by coherent noise. Finally, we assume that the columns of the unknown MT signal-space matrix $\mathbf{U}$ are normalized relative to the site 1 horizontal magnetic components, so that

$$\mathbf{U} = \begin{pmatrix} \mathbf{I}_2 \\ \mathbf{U}_1 \end{pmatrix}, \qquad \mathbf{V} = \begin{pmatrix} \mathbf{0} \\ \mathbf{V}_1 \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} \mathbf{I}_2 \\ \mathbf{0} \end{pmatrix}, \tag{29}$$

where all matrices are $5J \times 2$, partitioned to highlight the 2×2 upper blocks, corresponding to the quiet-site magnetic components. Note that the normalization assumption adopted here for $\mathbf{U}$ does not represent any real restriction, since it is only the span of the columns of $\mathbf{U}$ that is well defined in practice. With this modified definition of $\mathbf{U}$, the MT polarization parameters $\boldsymbol{\beta}_i$ are also modified, so that they now correspond to the horizontal magnetic components of the MT signal at site 1. The matrix $\mathbf{B}$ defined in (29) will be useful in the following.

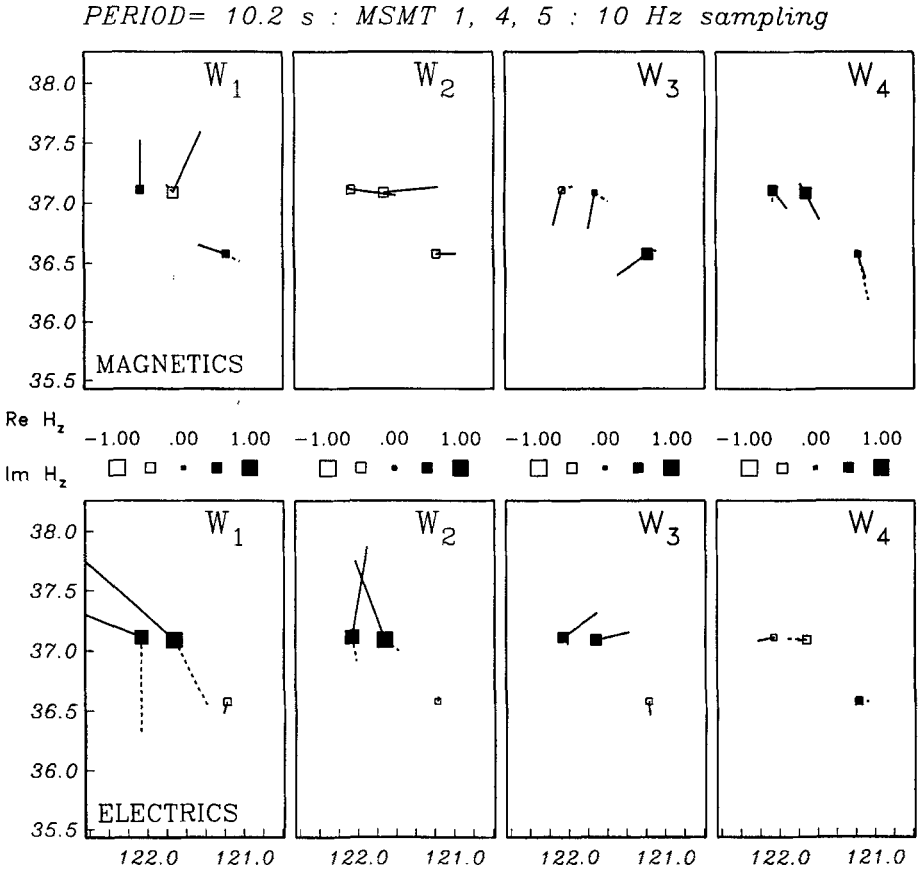

Figure 8. Coherent array mode vectors **W** for the example MSMT array at a period of 10 s. Here the coherence dimension of the array data is $M = 4$. Horizontal magnetic and electric components for the four columns of **W** are plotted in the upper and lower panels respectively, as vectors on a map of station locations. Real parts of vectors are solid, imaginary parts are dashed. Vertical magnetic components (H_z) are represented by the symbols (real parts in upper panels, imaginary parts lower). Symbol size is proportional to magnitude, with open symbols denoting negative H_z, and solid symbols positive. The phase of each 15-component vector (column of **W**) is chosen so that the RMS imaginary part of the horizontal magnetic field components is minimized. Electric field components are scaled to have the same units as the magnetic fields (nT) by dividing by the absolute value of the impedance for a 100 Ω m half-space. These plots display the spatial variation of the EM fields associated with the coherent signal/noise 'array modes'. For the idealized case of quasi-uniform sources, $M = 2$ and the horizontal magnetic field components for the two dominant modes should be approximately uniform. Here, the horizontal magnetic fields for the two dominant modes exhibit large variations, particularly between the base site and sites 4 and 5 to the north.

The RMEV estimate can be viewed as an iterated two-stage weighted least-squares procedure. In each iteration, we first use linear regression to estimate the generalized M-dimensional polarization vectors $\boldsymbol{\alpha}_i$ (i.e., step 2 of the RMEV estimate). Then, in the second stage, we regress on the estimated $\boldsymbol{\alpha}_i$ to determine **W** (step 4). When there is no coherent noise and $M = 2$, the vectors $\boldsymbol{\alpha}_i$ estimated in the first stage can be identified with the MT source polarization vectors $\boldsymbol{\beta}_i$, and the results of the second stage can be identified with **U**, i.e. the MT transfer functions. When $M > 2$, the MT source ($\boldsymbol{\beta}_i$) and coherent-noise ($\boldsymbol{\gamma}_i$) polarization vectors are mixed together in an unknown way in the estimated vectors $\boldsymbol{\alpha}_i$. If instead we could estimate $\boldsymbol{\beta}_i$, we could proceed to a modified step 4 in which we regressed $\mathbf{X}_i$ on these proxies for the MT source fields to obtain estimates of **U**, and hence the desired MT impedances. This is is just what a quiet site allows us to do. In particular, with the assumptions and notation of (29), the magnetic field components at the quiet site ($\mathbf{h}_{1i}$) can be viewed as unbiased estimates of $\boldsymbol{\beta}_i$, since

$$\mathbf{h}_{1i} = \mathbf{B}^*\mathbf{X}_i = \boldsymbol{\beta}_i + \mathbf{B}\boldsymbol{\varepsilon}_i, \qquad E\mathbf{B}\boldsymbol{\varepsilon}_i = \mathbf{B}E\boldsymbol{\varepsilon}_i = 0. \tag{30}$$

In fact, use of these simple estimates of $\boldsymbol{\beta}_i$ as the independent variable in a classical linear regression exactly reproduces the usual RR estimate.

The simplest way to make use of a quiet reference within our robust multivariate framework is thus just to use the cleaned local magnetic field channels from the quiet site to define the MT polarization parameters $\boldsymbol{\beta}_i$, and then proceed with step 4. Although very similar to the usual RRR estimate, there are some advantages to even this simple variant of the RMEV estimate. In particular, we still allow for outliers in all channels, including those at the reference site. All of the multivariate robust features described above are applied to clean up and/or downweight data. Only in the final step do we focus on data from a single site to define the reference signal.

Other ways to estimate the parameters $\boldsymbol{\beta}_i$ allow us to use data from additional channels. Most obviously, all five channels from a quiet site could be used. To do this we can compute the two dominant eigenvectors from the 5-D quiet-site data vectors. These can then be used in the analogue of (23) (i.e., with cleaned data vectors restricted to the quiet-site

components) to estimate $\boldsymbol{\beta}_i$. With this approach the reference is effectively an optimally weighted linear combinations of all five data channels. Note that the weights depend on the estimated σ_k^2, and will thus vary with frequency. The reference may thus be primarily defined by electric fields over part of the frequency range, and by magnetic fields over the rest. Compared to the standard RR approach, this approach can reduce noise in the reference. Of course this idea could be generalized further in some cases, for example data from several quiet sites could be used to define the reference.

There is a slightly more complicated way to use a quiet site to define the MT signal for the RMEV estimator, which makes even better use of all data channels. Assuming (as we did initially) that the columns of $\mathbf{W}$ have been orthonormalized, $\mathbf{WW}^*$ is the orthogonal projection onto the combined span of $\mathbf{U}$ and $\mathbf{V}$. A simple calculation thus yields

$$\mathbf{B}^*\mathbf{WW}^*\mathbf{X}_i = \boldsymbol{\beta}_i + \mathbf{B}^*\mathbf{WW}^*\boldsymbol{\varepsilon}_i = \hat{\boldsymbol{\beta}}_i. \tag{31}$$

The 2-D complex vectors $\hat{\boldsymbol{\beta}}_i$ are thus unbiased estimates of the MT source polarization parameters. These have a simple interpretation: they are the quiet-site horizontal magnetic components of the predicted data vectors

$$\hat{\mathbf{X}}_i = \mathbf{WW}^*\mathbf{X}_i. \tag{32}$$

With this approach, all data channels are thus used to form the best prediction of the coherent part of the data (MT signal + coherent noise). The predictions at the quiet site (which should still be free of coherent noise) are then used to define the reference signal. The advantage of this approach is that all data channels are averaged to reduce noise in the predicted reference channels. A disadvantage of this approach is that it can lead to a small bias in the resulting impedance estimates, even with an infinite number of data. The nature of this bias, and a simple correction are discussed in Appendix B. In spite of this extra complication, our experience so far suggests that this last scheme provides the most satisfactory way to make use of a quiet reference site in a MT array. Note that when $M = 2$ (no coherent noise) this estimate is exactly equivalent to the RMEV estimate proposed above, no matter which site is chosen as the reference.

5.2 Estimation of coherent-noise vectors

With a quiet reference site we can estimate $\mathbf{U}$ as well as $\mathbf{W}$. With a bit more effort we can also estimate the coherent-noise vectors $\mathbf{V}$. These can then be used to derive weights which allow us to emphasize the cleanest data sections. We sketch the basic idea. Details are provided in Appendix B. Proceeding as above, we use the quiet site to estimate the MT polarization vectors $\boldsymbol{\beta}_i$. These estimates are then used to predict and remove the MT part of the signal, leaving residuals that are a combination of coherent and incoherent noise. Using our estimates for incoherent-noise scales, an eigenvector analysis is used to estimate the incoherent-noise vectors $\mathbf{V}$.

This procedure allows us to separate the four vectors of the coherent data space ($\mathbf{W}$) plotted in Fig. 8 into the MT ($\mathbf{U}$) and coherent-noise ($\mathbf{V}$) parts (Fig. 9). Horizontal magnetic fields for the two MT signal vectors are indeed approximately uniform across the array (Figs 9a and b). For the coherent-noise vectors (Figs 9c and d), the electric and magnetic fields at the two northern sites (4 and 5) are of similar amplitude, direction, and phase. Corresponding components at the quiet site 1 are very small. Note that the phases of all vectors in Fig. 9 have been chosen to minimize the imaginary parts of the horizontal magnetic components. Thus the relative magnitudes and directions of the imaginary and real parts of the electric field vectors roughly define the impedance phase for each mode. For the quasi-uniform MT vectors, imaginary and real parts are collinear, and point in the same direction, indicating a physically reasonable first quadrant phase for the impedance. For the coherent-noise vectors, the imaginary parts are small, but tend to be reversed. This implies a physically unreasonable negative phase for this mode. The effect of this negative phase (and of the relatively large ratio of electric to magnetic components) on the single-station estimates for site 5 is clearly evident in Fig. 7.

With the columns of $\mathbf{W}$ decomposed into $\mathbf{U}$ and $\mathbf{V}$ as in Fig. 9, we can estimate the generalized polarization vectors of the quasi-uniform ($\boldsymbol{\beta}_i$) and coherent noise ($\boldsymbol{\gamma}_i$) components for each data segment i. Taking account of the variable noise in each channel, an appropriate weighted estimate of the polarization parameters is

$$\begin{bmatrix} \hat{\boldsymbol{\beta}}_i \\ \boldsymbol{\gamma}_i \end{bmatrix} = [\mathbf{W}^*\boldsymbol{\Sigma}_N^{-1}\mathbf{W}]^{-1}\mathbf{W}^*\boldsymbol{\Sigma}_N^{-1}\mathbf{X}_i, \tag{33}$$

and the predicted quasi-uniform ($\hat{\mathbf{X}}_i^U$), coherent-noise ($\hat{\mathbf{X}}_i^V$), and residual ($\mathbf{R}_i$) parts of the data vectors are given by

$$\hat{\mathbf{X}}_i^U = \mathbf{U}\hat{\boldsymbol{\beta}}_i, \qquad \hat{\mathbf{X}}_i^V = \mathbf{V}\hat{\boldsymbol{\gamma}}_i, \qquad \mathbf{R}_i = \mathbf{X}_i - \hat{\mathbf{X}}_i^U - \hat{\mathbf{X}}_i^V. \tag{34}$$

A useful statistic for summarizing the relative importance of coherent noise for period T, time segment i is

$$C(i, T) = \|\boldsymbol{\Sigma}_N^{-1/2}\hat{\mathbf{X}}_i^V\|^2 / \|\boldsymbol{\Sigma}_N^{-1/2}\hat{\mathbf{X}}_i^U\|^2. \tag{35}$$

$C(i, T)$ gives, as a function of time and period, the ratio of the total coherent-noise power to the signal power averaged over all components observed in the array. Note that in (35) we use the incoherent-noise scales ($\boldsymbol{\Sigma}_N^{1/2}$) to scale the individual components into reasonable non-dimensional units before averaging the power across components in the array.

$C(i, T)$ is plotted for the three-station example array (10 Hz sampling band) in Fig. 10. This figure shows that coherent noise is only significant for periods longer than ≈ 2 s, and is relatively most serious in the period range $4\text{ s} < T < 50\text{ s}$. Coherent noise virtually disappears from 1:30 am to 4:00 am. These are, in fact, exactly the hours that the BART DC electric railway shuts down on weekday nights. The rapid increase of coherent noise at precisely 4:00 am when BART trains start moving is especially striking. The noise then increases to a peak during the morning rush hour. This clearly demonstrates that the BART system is by far the dominant source of coherent noise in this array. The effects of BART on electromagnetic measurements in the San Francisco Bay area have been noted previously (e.g. Fraser-Smith & Coates 1978). This 'BART noise' almost certainly results primarily from leakage of DC current from the rails into the ground on the return path from the trains to the power substations (Fraser-Smith & Coates 1978). The BART noise sources thus should look like a series of grounded electric dipoles which are varying in time. This is completely consistent with the low phase and relatively large magnitude of the electric field components of the coherent-noise modes of Fig. 9.

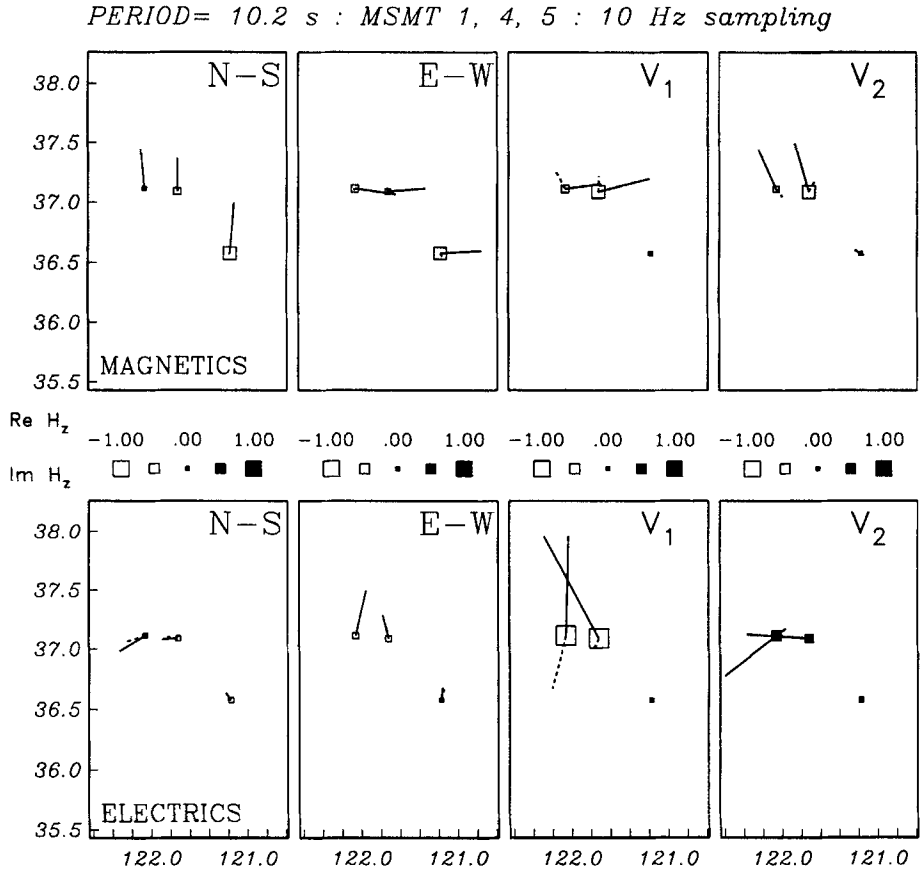

Figure 9. Results of separating the four coherent signal/noise vectors of Fig. 8 (**W**), into quasi-uniform source (**U**) and coherent-noise (**V**) parts. The separation is effected by using a preliminary estimate of **U** (derived by using site 1 as a reference) and assuming that coherent-noise sources are statistically independent of MT sources. Plot conventions are as in Fig. 8. The horizontal magnetic fields for the two quasi-uniform signal vectors (denoted N–S and E–W) are approximately uniform across the array, while the coherent-noise vectors ($\mathbf{V}_1$, $\mathbf{V}_2$) are dominated by coherent (and locally uniform) components at the two noisy northern sites. Note also that electric and vertical magnetic components are much larger in the coherent-noise vectors. Also note that the phase relationship between electric and magnetic fields is significantly different for the signal and coherent-noise vectors. Mixing of these sources (as in Fig. 8) causes the unphysical behaviour of the MT parameters seen in Fig. 7.

5.3 Downweighting coherent noise

As Fig. 10 makes clear, there is a 'window of opportunity' in the middle of the night when coherent noise is greatly reduced. The simplest way to take advantage of this circumstance would be to restrict processing to the relatively quiet interval 1:30–4:00 am. There is of course no reason why we cannot choose the time window used for processing as a function of period. Indeed, we could use $C(i, T)$ to determine directly which time segments should be included in estimates for each period. Since $C(i, T)$ can be computed as a by-product of the RMEV estimate, this data selection criterion can be included quite naturally as a (somewhat automatic) refinement to this estimator. This can be accomplished as follows:

(1) Use the quiet reference to provide an estimate of the MT vectors **U**.

(2) Estimate the coherent-noise vectors **V**, as described in Appendix B.

(3) Calculate $C(i, T)$, and then use only data segments for which $C(i, t) < p$ in a refined estimate of **U**.

(4) Using the refined estimate of **U**, redo steps (2) and (3). Iterate to convergence.

The results of applying this scheme (with $p = 0.2$) to array 1–4–5 are given for site 5 in Fig. 11(a). For comparison, RRR estimates obtained using the quiet site as a reference are given in Fig. 12(b). Relative to the RRR approach, this weighted RMEV scheme results in a clear improvement of MT impedance estimates. We stress, however, that the key to reducing coherent-noise bias is to have a quiet remote site. Both estimates in Fig. 12 essentially eliminate the bias seen in Fig. 7. Refinements due to the multiple-station approach, while non-negligible, are of second order.

Note that our proposed iterative refinement scheme can be usefully applied to reduce the effects of other sorts of intermittent noise, provided enough *a priori* information can be provided to get a reasonable starting estimate of the quasi-uniform response vectors **U**, or the coherent noise vectors **V**.

6 DETECTION OF INTERMITTENT COHERENT NOISE

If possible, one should obviously occupy a quiet (and distant) remote site for any MT survey conducted in an area likely to be contaminated by cultural noise. If coherent noise is present

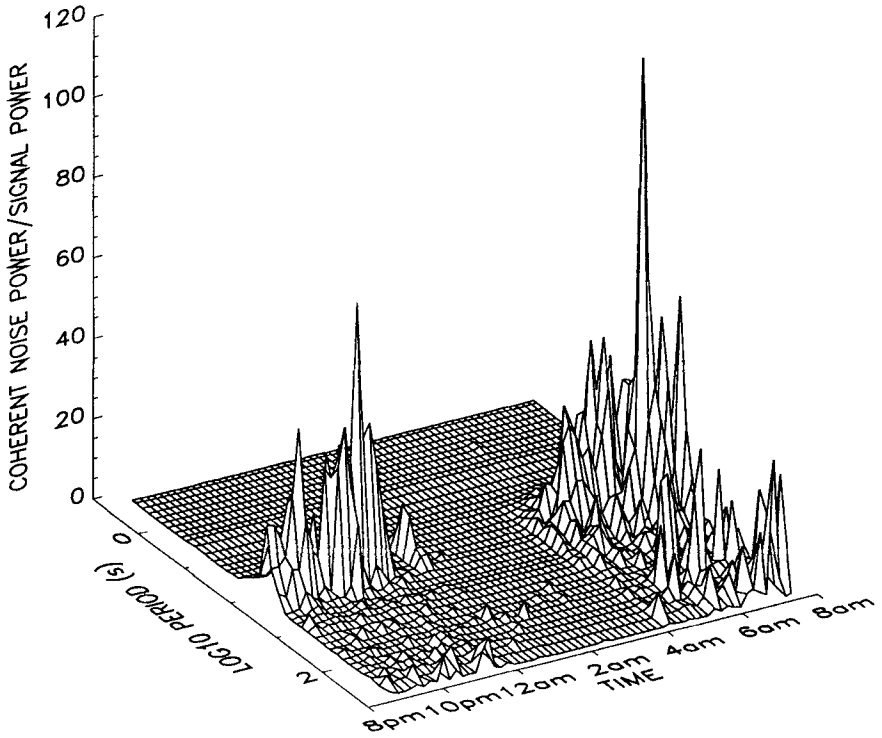

Figure 10. Ratio of total coherent-noise power $C(1, T)$ to total uniform source power as a function of period and time. The period range (4–50 s) and time periods (8:00 pm–12:00 am) and (4:00–8:00 am) where coherent noise overwhelms the MT signal are clearly evident. Coherent noise virtually disappears from 1:30 to 4:00 am, exactly when the BART DC electric railway is shut down for the evening. It is thus clear that BART is by far the dominant source of coherent noise in this array.

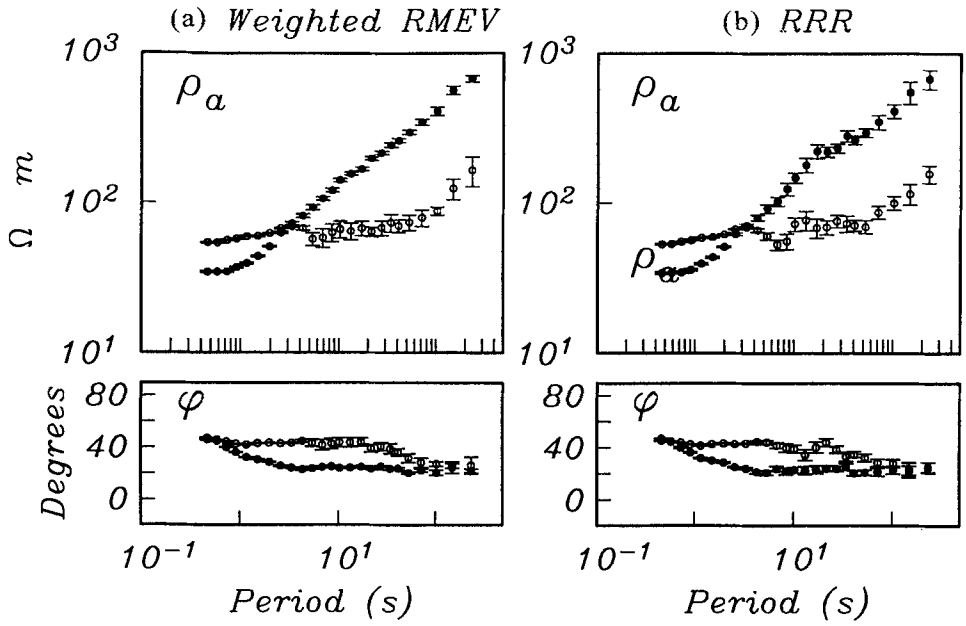

Figure 11. (a) Apparent resistivities and phases for site 5 (10 Hz sampling band) computed with the modified RMEV estimate, using predicted magnetic fields from site 1 as a reference, with downweighting of data segments significantly contaminated by coherent noise. For these estimates weights were set to zero for all segments for which the statistic $C(i, T)$ exceeded a cut-off value of $p = 0.2$. In (b) RRR estimates are given for the same site, computed using all of the data with the quiet site 1 as the reference. For both cases the erratic and unphysical behaviour seen in Fig. 7 is greatly reduced. However, the weighted RMEV approach results in significantly smoother estimates, with smaller error bars.

at all sites, it may be impossible to separate MT signal from coherent noise. However, the multivariate methods provide clear diagnostics for the presence of coherent noise. By applying these methods to short sections of data we might hope to find some time intervals uncontaminated by coherent noise. Here we consider some refinements to this simple data-screening approach which build upon the multivariate techniques developed in the previous sections.

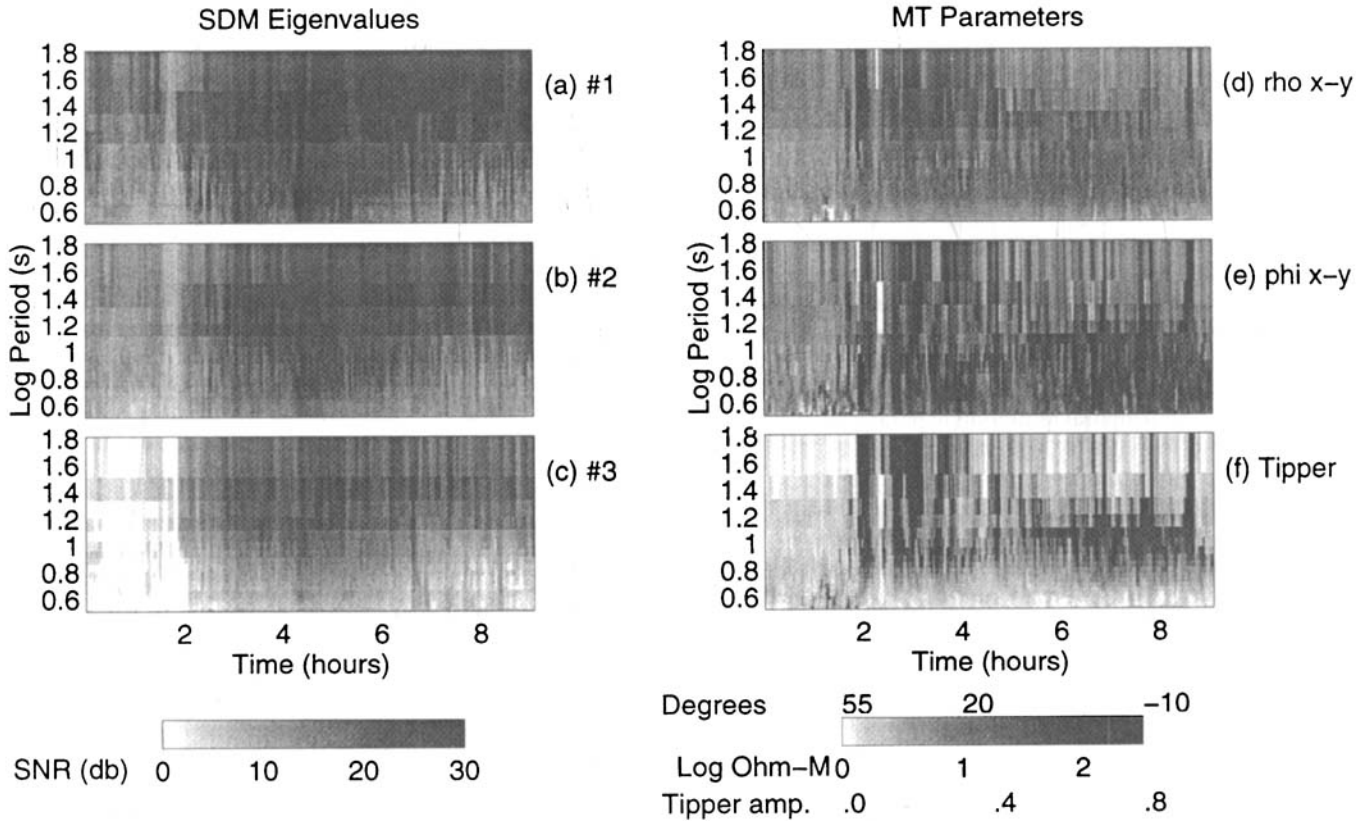

Figure 12. Plots of time-varying coherent-noise diagnostics computed for the data from the pair of Loma Prieta sites (A and B of Figs 1 and 2). To make this figure, the 10-D data vectors are first projected onto the coherent signal/noise space (here of dimension 3–4 depending on the period), and then diagnostics are computed in small local bands. The first column (a)–(c) gives the magnitude (signal-to-noise ratio in dB) of the three largest eigenvalues of the normalized SDMs as a function of (relative) time and period. The second column (d)–(f) gives time-varying estimates of apparent resistivity and phase for one MT mode, along with the tipper amplitude, all for station A. A sharp change occurs in all plots approximately two hours into the run. Before this change only the first two eigenvalues are significantly above background noise levels. After this time there are at least three significant eigenvalues at all periods longer than a few seconds. Note also the drop in phase, and increase in apparent resistivity and tipper after the second hour.

We begin by estimating $\mathbf{W}$ and $\mathbf{\Sigma}_N$ using the full data set, and we assume that we have found $M > 2$ (at least for some periods). Analogous to (34)–(36), we can compute the weighted projection onto the span of the coherent data subspace defined by the columns of $\mathbf{W}$:

$$\hat{\mathbf{X}}_i = \mathbf{W}[\mathbf{W}^*\mathbf{\Sigma}_N^{-1}\mathbf{W}]^{-1}\mathbf{W}^*\mathbf{\Sigma}_N^{-1}\mathbf{X}_i. \tag{36}$$

Note that

$$\hat{\mathbf{X}}_i = \hat{\mathbf{X}}_i^U + \hat{\mathbf{X}}_i^V, \tag{37}$$

but now we do not have enough information about signal and noise to directly separate the projected data vectors $\hat{\mathbf{X}}_i$ into component parts. However, by doing this projection we significantly reduce incoherent noise, especially if we use the cleaned data vectors $\tilde{\mathbf{X}}_i$ on the right-hand side of (36).

We can now group the projected, cleaned frequency-domain data vectors $\hat{\mathbf{X}}_i$ into small bands, localized in both time and frequency, form the SDM, and do an eigenvector analysis. Note that the number of non-zero eigenvalues for each time/frequency band will never exceed M, since the data vectors used to form the SDMs are first projected into the M-dimensional subspace defined by the span of the columns of $\mathbf{W}$. Moreover, if coherent noise ceases for some time sections, the number of eigenvalues significantly greater than one should drop to two for these sections.

As an example application of these ideas, we return to the Loma Prieta data used for the RRR results plotted in Fig. 1. Based on the close proximity of these stations to sites 4 and 5 in the three-station array, it is highly likely that coherent BART noise would contaminate impedance estimates at these sites also. As Fig. 10 indicates, this source of coherent noise should cease for at least a few hours every night. Results of the proposed diagnostic are plotted for this two-station array in Fig. 12. The three panels on the right give time–frequency sections of the three largest eigenvalues of the SDMs, normalized into non-dimensional units as for Fig. 4. There is a very clear change in the character of the signal just before the end of the second hour of data. After this time, the amplitudes of all eigenvalues increase sharply. More importantly, the bottom panel in this series shows that before this time there were only two significant eigenvalues at all periods, suggesting the absence of coherent noise.

We can also compute time-varying MT parameters such as apparent resistivities, phases, and tippers using the two dominant eigenvectors of the SDMs estimated from the projected data vectors. These are plotted for our example array in the left-hand panels of Fig. 12. Again, just before hour 2, the data character changes dramatically. Over most of the period range, apparent resistivities suddenly increase dramatically in amplitude, while phases drop to near (or below) zero. At the same

time, long-period tipper amplitudes jump from near zero to 0.75. In general, the MT parameters computed for later times exhibit the behaviour we have been able to associate with coherent BART noise in the three-station array 1–4–5. Clearly, the data before hour 2 are the 'good' data, with minimal coherent-noise contamination, while the later data are seriously contaminated by coherent noise.

Note that by projecting the data onto the coherent data space, incoherent noise is reduced, and the time-varying impedance estimates (each of which has only a small number of degrees of freedom) are significantly stabilized. As the number of components in the array increases, the degree of noise reduction and stabilization will be enhanced. We make use of redundancy in the array to reduce variances, and hence allow for improved temporal resolution of parameter estimates. Note that if we found $M = 2$ (no coherent noise), the time-varying plots of MT parameters would in fact be independent of time (and equal to the RMEV estimate).

Once we have identified sections of uncontaminated data, we can construct an initial estimate of **U**, and use this as a starting estimate for the coherent-noise downweighting scheme described in the previous section. We applied this method to the two Loma Prieta sites of Fig. 1. Resulting apparent resistivity and phase curves, combined with estimates from (uncontaminated) shorter-period sampling bands, are presented Fig. 13. The results of this processing represent a dramatic improvement relative to the initial RRR estimates presented in Fig. 1. To be fair, going back to the original time series (after doing this analysis!) one can see quite clearly where the character of the data changes, and coherent noise begins. Furthermore, using RRR on the 'good' data, one can achieve results nearly as good as those plotted in Fig. 13. If we can identify the 'good' data (assuming there is any), a variety of processing schemes should work adequately. The multivariate approach developed here can be a significant assist in the search for good data sections.

7 CONCLUSIONS

We have presented a general framework for understanding signal and noise characteristics in multiple-station MT array data. This framework includes a multivariate impedance estimator (RMEV) which automatically determines incoherent-noise levels, and then makes full and optimal use of all data channels to estimate the coherent part of the signal. The estimator is robust to isolated outliers in all channels, and provides a clear diagnostic for the presence of coherent noise. If there is no coherent noise, the RMEV estimator often performs better than previously proposed robust RR estimators, which apparently can be very sensitive to outliers in the remote channels.

When MT data are contaminated by coherent noise, there is no general way to guarantee that reasonable results will be obtained. However, using the approach developed here we can obtain a greatly enhanced understanding of signal and noise characteristics. Diagnostic statistics can be used to clarify the degree to which MT parameter estimates are likely to be contaminated by coherent noise, and to identify which parts of the data are least contaminated by incoherent noise. In the examples presented here it was possible to take advantage of this information to find a reasonable first-guess estimate of the quasi-uniform source signal vectors **U**. In this case, separation and optimal downweighting of coherent noise may be possible. Using several variants on this idea we have achieved very significant improvements in MT impedance estimates. We stress that the specific techniques which have proved useful here will not work in all circumstances. Nonetheless, we believe that the general framework and diagnostics developed here will prove useful in a broad range of circumstances.

ACKNOWLEDGMENTS

I thank Electromagnetic Instruments Incorporated (EMI) for the generous loan of MT instruments which made possible

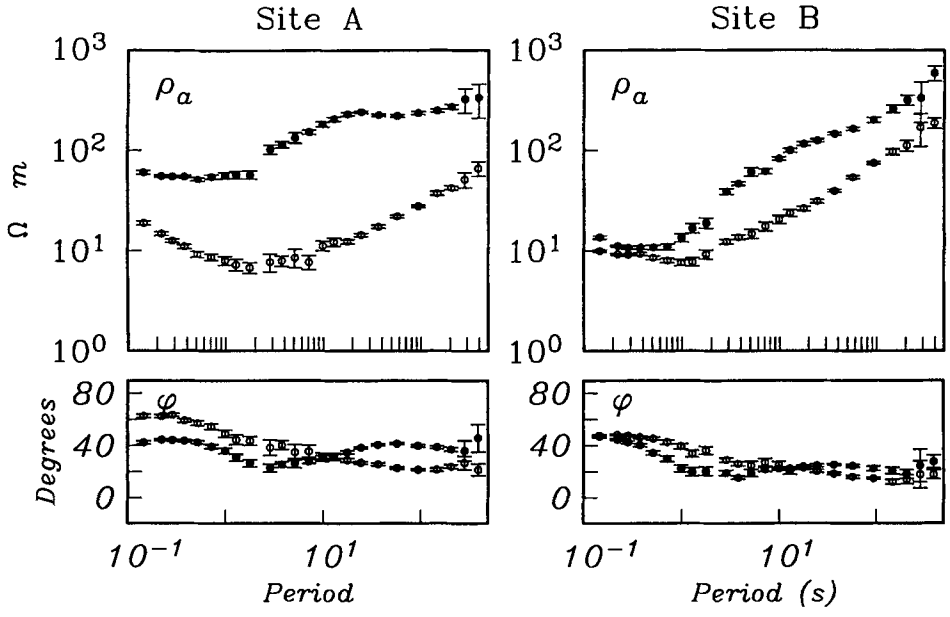

Figure 13. Apparent resistivities and phases computed for the Loma Prieta stations A and B, using data identified by the diagnostics of Fig. 12 as being free of coherent noise to define an initial estimate, followed by computation of a weighted RMEV estimate. Compared to the standard robust remote reference estimates given for the same two sites in Fig. 1, results are dramatically improved.

collection of the data used in this study. Professor H. Frank Morrison played a major role in this work, helping with data acquisition, and participating in many stimulating discussions of ideas and results. Weerachai Siripunvaraporn helped with collection and processing of the array data. This work was supported by the US Department of Energy through grant DE-FG06-92ER14277.

REFERENCES

Anderson, T.W., 1984. Estimating linear statistical relationships, *Ann. Statist.*, **12,** 1–45.

Anderson, T.W. & Rubin, H., 1956. Statistical inference in factor analysis, in *Proc. Third Berkeley Symp. on Mathematical Statistics and Probability,* **5,** 111–150, ed. Jerzy Neyman, University of California Press, Berkeley, CA.

Booker, J.R. & Chave, A.D., 1989. Introduction to the Special Issue on the EMSLAB-Juan de Fuca experiment, *J. geophys. Res.,* **94,** 14 093–14 098.

Chave, A.D. & Thomson, D.J., 1989. Some comments on magnetotelluric response function estimation, *J. geophys. Res.,* **94,** 14 215–14 226.

Chave, A.D., Thomson, D.J. & Ander, M.E., 1987. On the robust estimation of power spectra, coherences, and transfer functions, *J. geophys. Res.,* **92,** 633–648.

Egbert, G.D., 1987. A multivariate approach to the analysis of geomagnetic array data, *PhD thesis,* University of Washington.

Egbert, G.D., 1991. On the synthesis of a large geomagnetic array from small overlapping arrays, *Geophys. J. Int.,* **106,** 37–51.

Egbert, G. & Booker, J.R., 1986. Robust estimation of geomagnetic transfer functions, *Geophys. J. R. astr. Soc.,* **87,** 173–194.

Egbert, G.D. & Booker, J.R., 1989. Multivariate analysis of geomagnetic array data I: The response space, *J. geophys. Res.,* **94,** 14 227–14 248 (EB).

Fraser-Smith, A.C. & Coates, D.B., 1978. Large amplitude ULF electromagnetic fields form BART, *Radio Science,* **13,** 661–668.

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

Goldstein, N.E., 1988. Subregional and detailed exploration for geothermal-hydrothermal resources, *Geotherm. Sci. Tech.,* **1,** 393–431.

Giri, N.C., 1977. *Multivariate Statistical Inference,* Academic Press, New York, NY.

Gleser, L.J., 1981. Estimation in a multivariate 'errors-in-variables' regression model: large sample results, *Ann. Statist.,* **9,** 24–44.

Huber, P.J., 1981. *Robust Statistics,* Wiley, New York, NY.

Hawkins, D.M., 1977. On the investigations of the alternative regressions by principal components analysis, *Appl. Statist.,* **22,** 175–186.

Jiracek, G.R., Curtis, J.H., Ramirez, J., Martinez, M. & Romo, J., 1989. Two dimensional magnetotelluric inversion of the EMSLAB Lincoln Line, *J. geophys. Res.,* **94,** 14 145–14 152.

Jones, A.G. & Jodicke, H., 1984. Magnetotelluric transfer function estimation improvement by a coherence-based rejection technique, *paper presented at 54th Ann. Int. Mtg., Soc. Expl. Geophys.*

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

Jupp, D.L.B., 1978. Estimation of the magnetotelluric impedance functions, *Phys. Earth planet. Inter.,* **17,** 75–82.

Larsen, J.C., 1989. Transfer functions: smooth robust estimates by least squares and remote reference methods, *Geophys. J. Int.,* **99,** 655–663.

Mackie, R.L. & Madden, T.R., 1992. A magnetotelluric survey around the Loma Prieta fault zone, *EOS, Trans. Am. geophys. Un.,* **73,** 99.

Matsushita, S. & Campbell, W.H., 1967. *Physics of Geomagnetic Phenomena,* International geophysics series, vol. 11, Academic Press, New York, NY.

Morrison, D.F., 1967. *Multivariate Statistical Methods,* McGraw-Hill, San Francisco, CA.

Orange, A.S., 1989. Magnetotelluric exploration for hydrocarbons, *Proc. IEEE.,* **77,** 287–317.

Park, J. & Chave, A.D., 1984. On the estimation of magnetotelluric response functions using the singular value decomposition, *Geophys. J. R. astr. Soc.,* **77,** 683–709.

Rao, C.R., 1973. *Linear Statistical Inference and its Applications,* Wiley, New York, NY.

Seber, G.A.F., 1984. *Multivariate Observations,* Wiley, New York, NY.

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

Spangolini, U., 1994. Time domain estimation of MT impedance tensor, *Geophysics,* **58,** 712–721.

Stanley, W.D., Mooney, W.D. & Fuis, G.S., 1990. Deep crustal structure of the Cascade Range and surrounding regions from seismic refraction and magnetotelluric data, *J. geophys. Res.,* **95,** 19 419–19 438.

Stodt, J.A., 1986. Weighted averaging and coherence sorting for least-squares magnetotelluric estimates, *8th Workshop on Geomagnetic Induction in the Earth and Moon, Assoc. Geomag. Aeron.*

Sutarno, D. & Vozoff, K., 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, *PhD thesis,* MIT, Boston, MA.

Wannamaker, P.E., Booker, J.R., Jones, A.G., Chave, A.D., Filloux, H.H., Waff, H.S. & Law, L.K., 1989. Resistivity cross-section through the Juan de Fuca subduction system and its tectonic implications, *J. geophys. Res.,* **94,** 14 127–14 145.

APPENDIX A: ESTIMATION OF INCOHERENT NOISE VARIANCES

We consider the general case of data channels divided into two groups (for example the kth channel and the $K-1$ others; or the five channels at one site and the $5(J-1)$ remaining channels). We write (7) as

$$\mathbf{X}_{1i} = \mathbf{W}_1 \boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}_{1i}, \tag{A1a}$$

$$\mathbf{X}_{2i} = \mathbf{W}_2 \boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}_{2i}. \tag{A1b}$$

In general, $\mathbf{X}_{1i}$ and $\mathbf{X}_{2i}$ are P- and $(Q = K - P)$-dimensional complex data vectors for the two groups of channels; $\mathbf{W}_1$ is $P \times M$ and $\mathbf{W}_2$ is $Q \times M$. All random vectors in (A1) are complex and assumed to have zero mean. Initially we make the following assumptions about covariances of signal and noise:

$$\mathbf{W}_2^* \mathbf{W}_2 = \mathbf{I}_Q, \tag{A2}$$

$$E(\boldsymbol{\alpha}_i \mathbf{a}_i^*) = \boldsymbol{\Sigma}_\alpha = \mathrm{diag}[\lambda_1^2, \ldots, \lambda_M^2], \tag{A3}$$

$$E(\boldsymbol{\varepsilon}_{1i} \boldsymbol{\varepsilon}_{2i}^*) = \mathbf{0}, \tag{A4}$$

$$E(\boldsymbol{\varepsilon}_{2i} \boldsymbol{\varepsilon}_{2i}^*) = \boldsymbol{\Sigma}_{N2} = \mathbf{I}_Q, \tag{A5}$$

$$E(\boldsymbol{\varepsilon}_{1i} \boldsymbol{\varepsilon}_{1i}^*) = \boldsymbol{\Sigma}_{N1} = \mathrm{diag}[\sigma_1^2, \ldots, \sigma_P^2]. \tag{A6}$$

Assumptions (A2) and (A3) represent no restriction, but merely define our resolution of the ambiguity in the over-parametrization of (A1b). Eqs (A4)–(A6) reflect our assumption that the noise is incoherent. The assumption that $\boldsymbol{\Sigma}_{N1}$ is diagonal may be relaxed, although we will not explicitly address this extension here. We adopt the restrictive form for $\boldsymbol{\Sigma}_{N2}$ in (A5) to simplify initial derivations. This assumption

effectively means that we have normalized the components of $\mathbf{X}_2$ by the incoherent-noise scales for the second group of data channels. We will ultimately relax this restriction, and allow for a general diagonal error covariance.

First, note that in the absence of noise

$$\mathbf{X}_1 = \mathbf{W}_1\boldsymbol{\alpha} = \mathbf{W}_1\mathbf{W}_2^*\mathbf{W}_2\boldsymbol{\alpha} = \mathbf{W}_1\mathbf{W}_2^*\mathbf{X}_2, \tag{A7}$$

so $\mathbf{T}_1 = \mathbf{W}_1\mathbf{W}_2^*$ is the idealized (noise-free) transfer function for predicting $\mathbf{X}_1$ from $\mathbf{X}_2$. Now let

$$\mathbf{r}_i = \mathbf{X}_{1i} - \hat{\mathbf{T}}_1\mathbf{X}_{2i}, \tag{A8}$$

where $\hat{\mathbf{T}}_1$ is the least-squares estimate of $\mathbf{T}_1$ resulting from regressing $\mathbf{X}_1$ on $\mathbf{X}_2$, and let

$$\mathbf{R} = (I-Q)^{-1}\sum_{i=1}^{I}\mathbf{r}_i\mathbf{r}_i^* \tag{A9}$$

be the covariance matrix of the residuals r_i. If there were no noise added to $\mathbf{X}_2$, the diagonal elements R_{pp} would be unbiased estimates of σ_p^2, $p = 1, P$. Our first goal is to give an expression for the expected value of the residual covariance $E(\mathbf{R})$ valid for conditions (A2)–(A6). Using this expression, we derive a refined 'method-of-moments' estimator for the incoherent-noise variances for this case.

The full data vector

$$\mathbf{X} = \begin{pmatrix}\mathbf{X}_1\\ \mathbf{X}_2\end{pmatrix}$$

has covariance matrix

$$E\mathbf{X}\mathbf{X}^* = \begin{pmatrix}E\mathbf{X}_1\mathbf{X}_1^* & E\mathbf{X}_1\mathbf{X}_2^*\\ E\mathbf{X}_2\mathbf{X}_1^* & E\mathbf{X}_2\mathbf{X}_2^*\end{pmatrix} = \begin{pmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{21}\\ \boldsymbol{\Sigma}_{12} & \boldsymbol{\Sigma}_{22}\end{pmatrix}$$
$$= \begin{pmatrix}\mathbf{W}_1\boldsymbol{\Sigma}_\alpha\mathbf{W}_1^* + \boldsymbol{\Sigma}_{N1} & \mathbf{W}_1\boldsymbol{\Sigma}_\alpha\mathbf{W}_2^*\\ \mathbf{W}_2\boldsymbol{\Sigma}_\alpha\mathbf{W}_1^* & \mathbf{W}_2\boldsymbol{\Sigma}_\alpha\mathbf{W}_2^* + \mathbf{I}_Q\end{pmatrix}. \tag{A10}$$

The expectation of the residual covariance can then be given in terms of conditional covariances (e.g. Rao 1973):

$$E(\mathbf{R}) = \mathrm{Cov}(\mathbf{X}_1|\mathbf{X}_2) = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$$
$$= \boldsymbol{\Sigma}_{N1} + \mathbf{W}_1\boldsymbol{\Sigma}_\alpha\mathbf{W}_1^* - \mathbf{W}_1\{\boldsymbol{\Sigma}_\alpha\mathbf{W}_2^*[\mathbf{W}_2\boldsymbol{\Sigma}_\alpha\mathbf{W}_2^* + \mathbf{I}_Q]^{-1}\mathbf{W}_2\boldsymbol{\Sigma}_\alpha\}\mathbf{W}_2^*. \tag{A11}$$

Since the K columns of $\mathbf{W}_2$ are orthonormal we can find an $M \times (Q-M)$ orthonormal matrix $\mathbf{W}_{2\perp}$ so that $\tilde{\mathbf{W}}_2 = (\mathbf{W}_2\mathbf{W}_{2\perp})$ is a unitary matrix. Similarly, append a matrix of zeros to $\boldsymbol{\Sigma}_\alpha$ so that $\tilde{\boldsymbol{\Sigma}}_\alpha = (\boldsymbol{\Sigma}_\alpha\mathbf{0})$ is $M \times Q$ and

$$\tilde{\tilde{\boldsymbol{\Sigma}}}_\alpha = \begin{pmatrix}\boldsymbol{\Sigma}_\alpha & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}$$

is $Q \times Q$. Then it is readily seen that the expression in brackets in (A11) is equal to

$$\tilde{\boldsymbol{\Sigma}}_\alpha\tilde{\mathbf{W}}_2^*[\tilde{\mathbf{W}}_2\tilde{\tilde{\boldsymbol{\Sigma}}}_\alpha\tilde{\mathbf{W}}_2^* + \mathbf{I}_Q]^{-1}\tilde{\mathbf{W}}_2\tilde{\boldsymbol{\Sigma}}_\alpha^*$$
$$= \tilde{\boldsymbol{\Sigma}}_\alpha[\tilde{\tilde{\boldsymbol{\Sigma}}}_\alpha + \mathbf{I}_Q]^{-1}\tilde{\boldsymbol{\Sigma}}_\alpha^* = \mathrm{diag}[\lambda_1^4/(\lambda_1^2+1), \ldots, \lambda_M^4/(\lambda_M^2+1)]. \tag{A12}$$

Substituting back into (A11) and simplifying we find

$$E(\mathbf{R}) = \boldsymbol{\Sigma}_{N1} + \mathbf{W}_1\mathbf{D}\mathbf{W}_1^*, \tag{A13}$$

where

$$\mathbf{D} = \mathrm{diag}[\lambda_1^2/(\lambda_1^2+1), \ldots, \lambda_M^2/(\lambda_M^2+1)] \approx \mathbf{I}_M \quad \text{if } \lambda_1^2, \ldots, \lambda_M^2 \gg 1. \tag{A14}$$

Note that the components of $\mathbf{X}_2$ each have unit noise variance, while λ_m^2, $m = 1, M$, can be interpreted as the *total* signal power (summed over all K channels in the array) in mode m. This final approximation in (A14) should thus be valid provided signal-to-noise-ratios are at least 2 or so (or if K is large). For the remainder of our discussion we adopt this approximation and write

$$E[\mathbf{R}] \approx \boldsymbol{\Sigma}_{N1} + \mathbf{W}_1\mathbf{W}_1^* = \boldsymbol{\Sigma}_{N1} + \mathbf{T}_1\mathbf{T}_1^* \tag{A15}$$

(recall that $\mathbf{T}_1 = \mathbf{W}_1\mathbf{W}_2^*$, so $\mathbf{T}_1\mathbf{T}_2^* = \mathbf{W}_1\mathbf{W}_2^*\mathbf{W}_2\mathbf{W}_1^* = \mathbf{W}_1\mathbf{W}_1^*$). The simplest plausible way to estimate σ_p^2 would be to use the corresponding residual variance R_{pp}. Eq. (A15) gives an expression for the bias in this estimate:

$$ER_{pp} \approx \sigma_p^2 + \sum_{q=1}^{Q}|T_{pq}|^2. \tag{A16}$$

We now relax assumptions (A5), allowing a general diagonal covariance

$$E[\boldsymbol{\varepsilon}_{2i}\boldsymbol{\varepsilon}_{2i}^*] = \boldsymbol{\Sigma}_{N2} = \mathrm{diag}[\sigma_{p+1}^2, \ldots, \sigma_K^2]. \tag{A17}$$

Then $\mathbf{X}_2' = \boldsymbol{\Sigma}_{N2}^{-1/2}\mathbf{X}_2$ satisfies our original assumptions. A simple calculation shows that $\mathrm{Cov}(\mathbf{X}_1|\mathbf{X}_2') = \mathrm{Cov}(\mathbf{X}_1|\mathbf{X}_2)$ so (A15) still holds, provided $\mathbf{T}_1'$ replaces $\mathbf{T}_1$ (which we still take to be the transfer function relating $\mathbf{X}_1$ to $\mathbf{X}_2$). Now $\mathbf{T}_1' = \mathbf{T}_1\boldsymbol{\Sigma}_{N2}^{1/2}$, so (A15) becomes

$$E[\mathbf{R}] \approx \boldsymbol{\Sigma}_{N1} + \mathbf{T}_1\boldsymbol{\Sigma}_{N2}\mathbf{T}_1^*. \tag{A18}$$

Let $\rho_p = E[R_{pp}]$, $p = 1, P$, and index the components of the transfer function so that

$$\mathbf{T}_1 = \begin{pmatrix}T_{1,P+1} & \cdots & T_{1,K}\\ \vdots & & \vdots\\ T_{P,P+1} & \cdots & T_{P,K}\end{pmatrix}. \tag{A19}$$

Then (A18) implies

$$\rho_p = \sigma_p^2 + \sum_{l=P+1}^{K}|T_{pl}|^2\sigma_l^2, \quad p = 1, P. \tag{A20}$$

This gives us P equations in $K = P+Q$ unknowns (σ_l^2, $l = 1, K$).

For definiteness, we consider the simple case, where $P = 1$, $Q = K-1$. For each of the K channels we obtain one equation. Changing the index notation slightly we can express all of these equations as a system of K equations in K unknowns:

$$\rho_k = \sigma_k^2 + \sum_{l=1}^{K}|T_{kl}|^2\sigma_l^2, \quad k = 1, K. \tag{A21}$$

A similar reorganization of (A20) into (A21) can be achieved for other possible groupings of data channels.

The basic idea behind our estimate of incoherent-noise scales is to replace the expectation ρ_k by the actual observed residual variance R_{kk}, and the actual transfer function coefficients by estimates, and then solve (A21) for σ_l^2, $l = 1, K$. In detail, there are of course some further complications. Since the Q components of $\mathbf{X}_2$ are determined by only M independent signal components (with noise added), it would be unwise to do the regression of $\mathbf{X}_1$ on $\mathbf{X}_2$ directly. The high degree of collinearity among the components of $\mathbf{X}_2$ will lead to large statistical errors in the estimated transfer function components, and to possible numerical difficulties. A better approach is to use regression on principal components. The basic idea is that $\mathbf{T}_1 = \mathbf{W}_1\mathbf{W}_2^*$ is known to be of deficient rank; using this information, we can significantly improve estimates of the

components of $\mathbf{T}_1$. To do this we (1) estimate $\mathbf{W}_2$, using the M principal components of $\mathbf{X}_2$; (2) estimate $\hat{\boldsymbol{\alpha}} = \mathbf{W}_2^* \mathbf{X}_2$ using the estimates of $\mathbf{W}_2$; (3) estimate $\mathbf{W}_1$ by regressing $\mathbf{X}_1$ on $\hat{\boldsymbol{\alpha}}$; and (4) compute $\hat{\mathbf{T}}_1 = \hat{\mathbf{W}}_1 \mathbf{W}_2^*$.

Note that it is difficult to determine the correct number of principal components to use in this scheme (i.e., M) until we have a good estimate of the incoherent noisier scales. However, as (A21) shows we can obtain approximately unbiased estimates of σ_k^2 using the full vector $\mathbf{X}_2$. We thus begin our estimation scheme by using as many principal components as possible (consistent only with numerical stability). After obtaining preliminary estimates of the noise scales, we refine our estimate of M, and use fewer principal components for the regression.

There is one final issue to be addressed. Eq. (A21) (with ρ_k replaced by R_{kk}) can be written in matrix notation as

$$\mathbf{r} = (\mathbf{I}_K + \mathbf{B})\boldsymbol{\sigma}^2, \tag{A22}$$

where $\boldsymbol{\sigma}^2 = (\sigma_k^2, k = 1, K)$, $\mathbf{r} = (R_{kk}, k = 1, K)$, $B_{kl} = |T_{kl}|^2$. A possible difficulty with the estimate $\hat{\boldsymbol{\sigma}}^2 = (\mathbf{I}_K + \mathbf{B})^{-1}\mathbf{r}$ suggested above is that the component variances might not be positive. To ensure positivity of variances we have thus modified this estimate to

$$\hat{\boldsymbol{\sigma}}^2(\mu) = (\mathbf{I}_K + \mu\mathbf{B})^{-1}\boldsymbol{\rho}, \tag{A23}$$

where $0 < \mu \leq 1$ is chosen so that all components of $\hat{\boldsymbol{\sigma}}^2(\mu)$ are 'sufficiently' positive. Note that, as μ ranges from zero to one, $\hat{\boldsymbol{\sigma}}^2(\mu)$ varies smoothly from the 'raw' uncorrected residual variances $\mathbf{r}$ to the solution to (A21). Since all components of $\mathbf{r}$ are positive, there is guaranteed to be a range of values for μ for which all components of $\hat{\boldsymbol{\sigma}}(\mu)$ are also positive. Note also that all elements of the matrix $\mathbf{B}$ are also positive, so any positive solution vector $\hat{\boldsymbol{\sigma}}(\mu)$ must be (component-wise) less than $\mathbf{r}$. Thus the effect of this procedure is to adjust the raw residual variance estimates downwards so that they are more nearly unbiased.

To determine μ in practice, we start with $\mu = 1$ and solve (A23) with successively smaller values of μ until the condition $\hat{\sigma}_k(\mu)^2 > f_{\min} R_{kk}$ is satisfied for $k = 1, K$. The parameter $f_{\min}$ determines the maximum allowable factor that the residual variances will be reduced by in the incoherent-noise variance estimates. We have used $f_{\min} = 0.2$ in our work so far. Tests with synthetic array data with known error variances show that the procedure does quite well in recovering the correct variances.

APPENDIX B: SEPARATION OF SIGNAL AND COHERENT NOISE

In this appendix we discuss separation of coherent noise from the desired MT signal, assuming some prior information about signal (or noise) processes can be provided. We adopt the model of eq. (7) with the errors $\boldsymbol{\varepsilon}_i$ assumed to satisfy $\mathrm{Cov}(\boldsymbol{\varepsilon}_i) = \mathbf{I}$. Throughout our discussion we assume that something is known about the quasi-uniform MT part of the signal (represented by $\mathbf{U}\boldsymbol{\beta}_i$). However, the model (B1) is symmetric in $\mathbf{U}\boldsymbol{\beta}_i$ and $\mathbf{V}\boldsymbol{\gamma}_i$, so similar knowledge about noise processes could be used by switching the roles of $\mathbf{U}$ and $\mathbf{V}$ throughout. We consider two distinct cases, requiring different sorts of prior information.

(I) We know (or have a good estimate of) $\mathbf{U}$.

(II) We know, or have a surrogate for $\boldsymbol{\beta}_i$. In general, we assume that we have random vectors $\mathbf{b}_i$ independent of $\boldsymbol{\gamma}_i$ satisfying

$$\mathbf{b}_i = \mathbf{A}\boldsymbol{\beta}_i + \boldsymbol{\eta}_i, \qquad E[\boldsymbol{\eta}_i \boldsymbol{\varepsilon}_i^*] = \boldsymbol{\Sigma}_{\eta\varepsilon}, \qquad \mathrm{Cov}(\boldsymbol{\eta}_i) = \mathbf{I}, \tag{B1}$$

where the 2×2 matrix $\mathbf{A}$ is assumed to be invertible, and the matrix defining the covariance between $\boldsymbol{\eta}$ and $\boldsymbol{\varepsilon}$ is assumed known. Two examples of this case are: (1) the $\mathbf{b}_i$ are the horizontal components of the magnetic field at a quiet base site; and (2) the $\mathbf{b}_i$ correspond to the predicted magnetic fields at a quiet base site, i.e. the $\hat{\boldsymbol{\beta}}_i$ defined in (31). Note that for both of these examples we can readily compute $\boldsymbol{\Sigma}_{\eta\varepsilon}$, using the fact that all incoherent-noise variances are normalized to one. In the notation of (29), for the first example $\boldsymbol{\Sigma}_{\eta\varepsilon} = \mathbf{B}^*$, while for the second $\boldsymbol{\Sigma}_{\eta\varepsilon} = \mathbf{B}^*\mathbf{W}\mathbf{W}^*$.

Case I. Since the span of the columns of $\mathbf{U}$ is contained in the span of the columns of $\mathbf{W}$, we may write $\mathbf{W} = (\mathbf{U}\mathbf{U}_\perp)$, where $\mathbf{U}_\perp$ is orthonormal, and orthogonal to $\mathbf{U}$. In practice, given $\mathbf{U}$ we can easily estimate $\mathbf{U}_\perp$. Let $\mathbf{Q} = \mathbf{I} - \mathbf{U}\mathbf{U}^*$ be the projection onto the orthogonal complement of the MT signal space. Then the number of significant eigenvalues of $\mathbf{QSQ}$ should be $N = M - 2$. The corresponding eigenvectors are orthogonal to $\mathbf{U}$, and (except for the effects of random noise) they are contained in the span of $\mathbf{W}$, and can be used to estimate $\mathbf{U}_\perp$.

The coherent noise vectors $\mathbf{V}$ are not necessarily orthogonal to the MT signal vectors $\mathbf{U}$, but in general satisfy

$$\mathbf{V} = \mathbf{U}\mathbf{C} + \mathbf{U}_\perp \mathbf{C}'. \tag{B2}$$

As we are only really interested in the span of $\mathbf{V}$ we may take $\mathbf{C}'$ in (B4) to be the identity. To estimate $\mathbf{V}$ we thus need only estimate $\mathbf{C}$.

Substituting (B2) in (B1) and multiplying both sides of this equation by $\mathbf{U}^*$ and $\mathbf{U}_\perp^*$ we obtain

$$\mathbf{U}^*\mathbf{X}_i = \boldsymbol{\beta}_i + \mathbf{C}\boldsymbol{\gamma}_i + \mathbf{U}^*\boldsymbol{\varepsilon}_i = [\boldsymbol{\beta}_i + \mathbf{U}^*\boldsymbol{\varepsilon}_i], \tag{B3}$$

$$\mathbf{U}_\perp^*\mathbf{X}_i = \boldsymbol{\gamma}_i + \mathbf{U}_\perp^*\boldsymbol{\varepsilon}_i. \tag{B4}$$

Note that the errors $\mathbf{U}^*\boldsymbol{\varepsilon}_i$ and $\mathbf{U}_\perp^*\boldsymbol{\varepsilon}_i$ on the right-hand side of (B3) and (B4) are still uncorrelated, and of unit variance. To estimate $\mathbf{C}$ we treat all of $\boldsymbol{\beta}_i + \mathbf{U}^*\boldsymbol{\varepsilon}_i$ as error, ignore the errors on the right-hand side of (B4), and regress $\mathbf{U}^*\mathbf{X}_i$ on $\mathbf{U}_\perp^*\mathbf{X}_i$. The resulting estimate for $\mathbf{C}$ can then be expressed in terms of the SDM computed from the vectors $\mathbf{X}_i$ as

$$\hat{\mathbf{C}} = \mathbf{U}^*\mathbf{S}\mathbf{U}_\perp[(\mathbf{U}_\perp^*\mathbf{S}\mathbf{U}_\perp)^{-1}]. \tag{B5}$$

Because of the errors in $\mathbf{U}_\perp^*\mathbf{X}_i$, this estimate of $\mathbf{C}$ will be biased. This bias can be removed by replacing $\mathbf{U}_\perp^*\mathbf{S}\mathbf{U}_\perp$ by $\mathbf{U}_\perp^*\mathbf{S}\mathbf{U}_\perp - \mathbf{I}$ in (B5), thus estimating $\mathbf{V}$ by

$$\hat{\mathbf{V}} = \mathbf{U}_\perp + \mathbf{U}[\mathbf{U}^*\mathbf{S}\mathbf{U}_\perp][\mathbf{U}_\perp^* - \mathbf{I}]^{-1}. \tag{B6}$$

Case II. This case is most simply worked out by considering the joint sample covariance matrix of $\mathbf{b}$ and $\mathbf{X}$:

$$\mathbf{S} = (1/I)\sum_{i=1}^{I} \begin{pmatrix} \mathbf{b}_i \\ \mathbf{X}_i \end{pmatrix} (\mathbf{b}_i^* \mathbf{X}_i^*), \tag{B7}$$

which has expectation

$$E(\mathbf{S}) = \begin{pmatrix} \mathbf{A}\boldsymbol{\Sigma}_\beta \mathbf{A}^* + \mathbf{I} & \mathbf{A}\boldsymbol{\Sigma}_\beta \mathbf{U}^* + \boldsymbol{\Sigma}_{\eta\varepsilon} \\ \mathbf{U}\boldsymbol{\Sigma}_\beta \mathbf{A}^* + \boldsymbol{\Sigma}_{\eta\varepsilon}^* & \mathbf{U}\boldsymbol{\Sigma}_\beta \mathbf{U}^* + \mathbf{V}\boldsymbol{\Sigma}_\gamma \mathbf{V}^* + \mathbf{I} \end{pmatrix}. \tag{B8}$$

From (B8) we see that

$$\hat{\mathbf{U}}' = (\mathbf{S}_{21} - \boldsymbol{\Sigma}^*_{\eta\varepsilon})\mathbf{S}_{11}^{-1} \tag{B9}$$

is an asymptotically unbiased estimate of $\mathbf{U}' = \mathbf{U}\mathbf{A}'^{-1}$, where $\mathbf{A}' = \mathbf{A} + \mathbf{A}^{-1}\boldsymbol{\Sigma}_\beta^{-1}$. Since $\mathbf{U}$ is determined only up to multiplication by an arbitrary invertible 2×2 matrix, we can obtain an estimate of $\mathbf{U}$ by orthonormalizing $\hat{\mathbf{U}}'$.

To estimate $\mathbf{V}$ we form the matrix

$$\mathbf{S}_V = \mathbf{S}_{22} - (\mathbf{S}_{21} - \boldsymbol{\Sigma}^*_{\eta\varepsilon})(\mathbf{S}_{11} - \mathbf{I})^{-1}(\mathbf{S}_{12}\boldsymbol{\Sigma}_{\eta\varepsilon}). \tag{B10}$$

Using (B8) and (B9) it can be verified that (asymptotically) the expected value of $\mathbf{S}_V$ is $\mathbf{V}\boldsymbol{\Sigma}_\gamma\mathbf{V}^* + I$. An eigenvector decomposition of $\mathbf{S}_V$ can thus be used to compute estimates of $\mathbf{V}$.