Geophys. J. Int. (2001) **144,** 65–82

On the stability of magnetotelluric transfer function estimates and the reliability of their variances

Markus Eisel[1,*] and Gary D. Egbert[2]

[1] *Software and Systems AG, Eschborn, Germany. E-mail: Markus.Eisel@fra.sesa.de*
[2] *College of Oceanic and Atmospheric Sciences, Oregon State University, Corvallis, OR, USA*

Accepted 2000 July 24. Received 2000 July 24; in original form 1999 July 12

SUMMARY

Using data from a continuously operating two-station magnetotelluric (MT) array in central California we have computed robust remote reference MT transfer functions (TFs) for each day in the 2 yr period 1996–1997. Typical deviations of the daily estimates from the overall long-term average TF ranged from 2–3 per cent for periods of less than 300 s to about 10 per cent at a period of 2000 s. Day-to-day deviations were largely random, and exhibited little temporal correlation or long-term trend. There is some evidence for small-frequency independent variations in impedance amplitudes, suggestive of subtle slow changes in near-surface distortion. However, there was no clear seasonal component to this signal, as might be expected if hydrologic changes in the near surface were the cause. Comparison of estimated error bars to TF variability showed that for periods between 10 and 100 s (where coherent noise sometimes biased TF estimates) the standard asymptotic theory for the robust estimator yielded error bars that were too small by as much as a factor of two. At longer and shorter periods these standard error bars were consistent with the actual precision of the TF estimates. We also considered the reliability of error bars computed with two variants on the jackknife approach. For the first approximate scheme we computed the weights for the robust TF estimates once with all data, followed by application of the jackknife to the final weighted least-squares estimate. We show that for this 'fixed-weight' jackknife, variances can be given in closed form even for the remote reference case. Fixed-weight jackknife error bars were larger than those computed in the conventional fashion, but still significantly underestimated the true variability in the 10–100 s bias band, and were systematically too large at other periods. We also tried a subset deletion jackknife, applying the full robust procedure with contiguous subsets of data deleted. Provided large subsets (5 per cent, or approximately 1 hr) were deleted, this approach yielded significantly more realistic error bars in the bias band. However, error bars at periods outside the bias band now significantly overestimated the actual day-to-day variability of TF estimates. The jackknifed error bars were thus always more conservative, though not necessarily more reliable.

Key words: jackknife, magnetotellurics, seasonal variation, transfer functions.

1 INTRODUCTION

With the magnetotelluric (MT) method, naturally occurring variations of electromagnetic (EM) fields observed at the surface of the Earth are used to map variations in electrical conductivity within the Earth's crust and mantle. Under the fundamental MT assumption that the external sources are spatially uniform, the two horizontal components of the electric field variations can be linearly related to the two horizontal magnetic field components via a 2×2 impedance tensor or transfer function (TF), which can then be inverted for electrical conductivity. Estimation of these TFs from electric and magnetic field time-series is the first step in the interpretation of MT data in terms of the structure and physical state of the Earth's interior.

The earliest practitioners of MT used classical time-series methods for TF estimation (e.g. Bendat & Piersol 1971), applying a simple least-squares fitting procedure to a series of

* Previously at: GeoForschungsZentrum Potsdam, Germany and College of Oceanic and Atmospheric Sciences, Oregon State University, Corvallis, OR, USA.

Fourier-transformed data segments (e.g. Sims *et al.* 1971). For several reasons this classical approach often failed to yield physically reasonable or reproducible results for the MT problem. MT data quality can be highly variable, with both signal and noise power varying by orders of magnitude over time. Furthermore, there is typically noise in both the 'output' or predicted electric field channels and the 'input' or independent magnetic channels. Ignoring the noise in the input channels can lead to significant biases in TF estimates. In addition, variations of the correlation structure of both signal and noise can vary dramatically and contribute to the difficulty of estimating reliable TFs. Two developments over the past two decades have improved the situation considerably: the remote reference method, in which data from a simultaneously operating second MT site are used to cancel out local noise (Gamble *et al.* 1979b), and data-adaptive robust schemes that automatically down-weight or eliminate poor-quality data (e.g. Larsen 1980; Egbert & Booker 1986; Chave *et al.* 1987). With proper care in processing, modern remote reference MT data can routinely yield smooth and repeatable MT impedances. The comparison paper of Jones *et al.* (1989) illustrates the significant advantages of both the remote reference and robust schemes when applied to long-period MT data.

With a statistical approach to TF estimation, one also obtains estimation error variances that define the precision of the TF estimates, and of derived interpretation parameters such as apparent resistivities. These error bars are ultimately required at the modelling or inversion stage to assess the adequacy of fit of a derived model to the measured data, and thus play an important role in the overall interpretation of MT data. Chave & Thomson (1989) questioned the standard methods commonly used for calculating estimation variances and advocated use of the jackknife method (e.g. Efron 1982; Efron & Gong 1983). Their concerns stemmed from the fact that a number of assumptions commonly made about the statistical character of the data (e.g. Gaussian error distributions, stationarity, lack of serial correlation, homogeneity of variances) are often violated. As a non-parametric method, Chave & Thomson (1989) suggested that the jackknife should be much less sensitive to violations of these types of statistical assumptions.

Since 1995, researchers from the University of California, Berkeley, have recorded magnetotelluric (MT) time-series at two locations along the San Andreas Fault in Central California (Boyd *et al.* 1997). The purpose of this monitoring experiment is to investigate the possibility that anomalous variations in electromagnetic (EM) fields or Earth resistivity might occur as precursors to earthquakes (e.g. Park *et al.* 1993; Fraser-Smith *et al.* 1990).

In addition to allowing for investigation of possible EM earthquake precursors, this data set provides a rare opportunity to study the long-term stability of MT TFs, and to assess how well estimated error bars actually represent the precision of TF estimates. Our approach is to estimate TFs and error bars for each day separately, and then analyse the sequence of estimates to characterize systematic trends and random variability. In particular, by comparing daily deviations from long-term averages to estimates of error bars we can assess the reliability of these measures of TF precision. This sort of comparison complements simulation studies, in which the performance of estimators can be compared on synthetic data. Although there are some advantages to a simulation approach (for example, the true TF would be known and one could control the statistical character of signal and noise), there is also much value in a comparison of methods using real data. Although we do not know exactly the true TF in this case, and thus cannot be sure of the accuracy of our estimates, we can characterize precision. This precision is in fact appropriate to compare to statistical error bars, which should not be expected to account for bias effects due to miscalibration or slow drifts in instrument responses (or for temporal variations in TFs due to Earth conductivity). With this data set we can thus provide a valid test of error bar reliability using real data with all of the complications that this entails. Of course, there are specific noise conditions at the two sites considered here, and it must be stressed that the degree to which our results generalize is open to question. Similar experiments in varying environments would of course be desirable.

In this paper we focus exclusively on issues of TF stability and error bar accuracy. We do not explicitly consider the question of earthquake precursors, although our results do have implications for this problem as well. In particular, one sort of precursor that has been observed involves temporal changes in Earth resistivity prior to fault rupture (for a review see e.g. Park *et al.* 1993). To assess the significance of such observations we must be able to say how well an apparent resistivity curve should be reproduced when measurements are repeated at a later time. To do this properly it is essential to know that the estimated error bars accurately represent the true uncertainties in the TFs being compared. Application of our findings to the search for possible EM earthquake precursors near the San Andreas Fault will be presented elsewhere.

2 TIME-SERIES ANALYSIS

The two MT stations used for this study are located on the San Andreas Fault near Hollister, CA (site SAO, about 150 km southeast of San Francisco) and at Parkfield (site PKD, 150 km further southeast). A map of the area is shown in Fig. 1.

Each site is equipped with a five-channel MT1 system (EMI Inc) with three induction coils measuring the N–S (H_x), E–W (H_y) and vertical (H_z) components of the magnetic field, and two 100 m electrical dipoles in the N–S and E–W directions (E_x and E_y respectively). Quanterra data loggers are used to digitize and record the data at sampling frequencies of 1 and 40 Hz, with time synchronization between stations maintained by GPS. The data are then downloaded over telephone lines to the Berkeley Seismological Laboratory, where they are archived with other data from the Berkeley Digital Seismic Network (BDSN). Only the data recorded at 1 Hz are used in this study.

To estimate MT TFs we use a robust remote reference scheme, essentially the method described in Egbert & Booker (1986) with modifications for remote reference processing similar to that proposed by Chave & Thomson (1989). This procedure is based on the regression M-estimate (RME; see Huber 1981), which allows for heavy-tailed (non-Gaussian) error distributions. To set notation for our discussion of the error variance estimates in the following sections, we briefly summarize the procedure here.

Consider first the specific problem of estimating the MT TF between one component of the electric field (e.g. E_x) and the

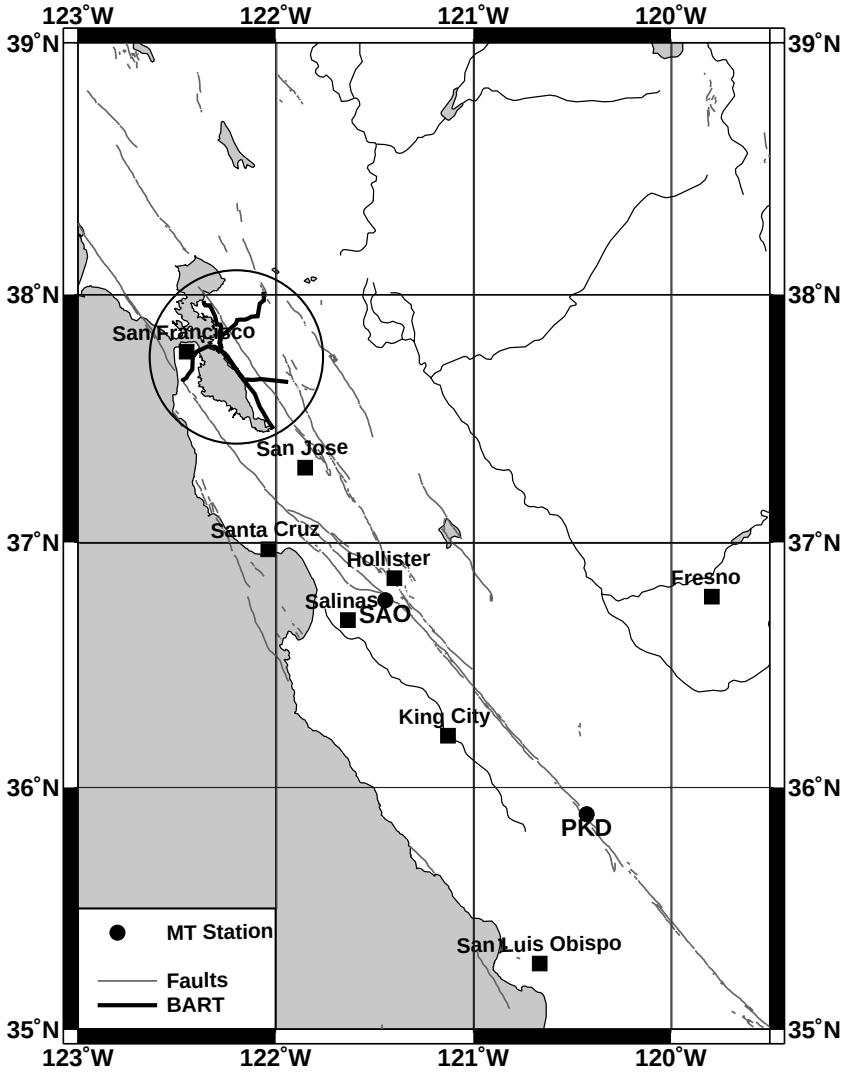

Figure 1. Map of central California showing the San Andreas Fault system, the San Francisco Bay Area with the tracks of the Bay Area Rapid Transport system (BART, encircled) and the locations of the two MT sites.

two horizontal magnetic field components H_x and H_y at a single fixed period T. The time-series is divided into M short time windows and Fourier transformed. Fourier coefficients for N periods in a band centred around T are used for the TF estimate, for a total of $I = MN$ complex data. Estimates $\hat{\mathbf{Z}}$ of the TF (i.e. impedance elements) are then chosen to minimize residuals in the equation

$$\begin{pmatrix} E_1 \\ \vdots \\ E_I \end{pmatrix} - \begin{pmatrix} H_{x1} & H_{y1} \\ \vdots & \vdots \\ H_{xI} & H_{yI} \end{pmatrix} \begin{pmatrix} \hat{Z}_1 \\ \hat{Z}_2 \end{pmatrix} = \begin{pmatrix} \eta_1 \\ \vdots \\ \eta_I \end{pmatrix}, \tag{1}$$

or in matrix notation,

$$\mathbf{E} - \mathbf{H}\hat{\mathbf{Z}} = \boldsymbol{\eta}. \tag{2}$$

With standard least squares (LS) this is accomplished by minimizing the sum of the squares of the residuals

$$\sum_i |\eta_i|^2 \rightarrow \min, \tag{3}$$

yielding

$$\hat{\mathbf{Z}} = (\mathbf{H}^\dagger \mathbf{H})^{-1} (\mathbf{H}^\dagger \mathbf{E}), \tag{4}$$

where the superscript † denotes the conjugate transpose of the complex matrix. Note that throughout this paper all LS problems will involve complex (frequency-domain) data. It is possible to recast everything as a real LS problem in terms of real and imaginary parts (as in Egbert & Booker 1986), but for our purposes it will be more convenient to retain complex notation. Note also that the complex 2-D column vector $\hat{\mathbf{Z}}$ defined in eq. (1) corresponds to the first *row* of the usual MT impedance tensor. All results apply equally to other impedance elements and to vertical-field transfer functions.

For the RME minimization of the quadratic loss functional, eq. (3) is replaced by the more general form

$$\sum_i \rho(|\eta_i|/\hat{\sigma}) \rightarrow \min, \tag{5}$$

where σ is some estimate of the scale of typical residuals. By choosing ρ to penalize large residuals less heavily than the quadratic (3) used for LS, the influence of outliers on the

estimate can be substantially reduced. Egbert & Booker (1986) used the loss function of Huber (1981):

$$\rho(\eta)=\begin{cases}\eta^2/2 & |\eta|<\eta_0\\ \eta_0|\eta|-\eta_0^2/2 & |\eta|>=\eta_0\end{cases}, \tag{6}$$

where $\eta_0=1.5$. To find the minimizer of (5) an iterative weighted LS procedure can be used. Let $\psi(r)=\rho'(r)$ be the derivative of the loss function (referred to as the influence function) and set $w(r)=\psi(r)/r$. The minimizer of (5) satisfies

$$\hat{\mathbf{Z}}=(\mathbf{H}^\dagger\mathbf{W}\mathbf{H})^{-1}(\mathbf{H}^\dagger\mathbf{W}\mathbf{E}), \tag{7}$$

where $\mathbf{W}=\mathrm{diag}(w_1,\ldots,w_I)=\mathrm{diag}[w(|\eta_1|),\ldots,w(|\eta_I|)]$ is a diagonal matrix of weights. The RME thus corresponds approximately to the weighted LS problem $\Sigma_i\, w(|\eta_i|)|\eta_i|^2\rightarrow\min$. However, the weights depend on the residuals η_i (and hence on the TF estimate), so an iterative procedure is required. Given an estimate of the TF (and of the error scale σ), residuals can be used to calculate weights, and the weighted LS problem can be solved for a new TF estimate. This procedure can be started from a standard LS estimate of the TF (and some robust estimate of error scale) and then repeated until convergence.

For convex loss functions (e.g. the Huber function of 6), convergence of this procedure to the unique minimizer of (5) is guaranteed (Huber 1981). For the loss function in (6) the weights are

$$w(\eta)=\begin{cases}1 & |\eta|<\eta_0\\ \eta_0/|\eta| & |\eta|>=\eta_0\end{cases}, \tag{8}$$

that is, data corresponding to large residual vectors are given smaller weights. By using a non-convex loss function based on the extreme value distribution (Thomson 1977; Egbert & Booker 1986) the weights are essentially 1 for residuals with magnitudes below a hard cut-off of η_1 and 0 above this. For the processing results discussed here we used the scheme of Egbert & Booker (1986), iterating with Huber weights to convergence, followed by a final step with a hard cut-off at $\eta_1=2.8$. Note that a similar procedure was used by Chave *et al.* (1987) and Chave & Thomson (1989).

For the case of remote reference data, the LS estimate of (4) is replaced by

$$\hat{\mathbf{Z}}=(\mathbf{R}^\dagger\mathbf{H})^{-1}(\mathbf{R}^\dagger\mathbf{E}), \tag{9}$$

where $\mathbf{R}$ is the magnetic field at the remote site (Gamble *et al.* 1979a). To generalize the RME to this remote reference case, Chave & Thomson (1989) proposed iterating the weighted analogues of eq. (9), i.e. $\hat{\mathbf{Z}}=(\mathbf{R}^\dagger\mathbf{W}\mathbf{H})^{-1}(\mathbf{R}^\dagger\mathbf{W}\mathbf{E})$, with the weights on the diagonal of $\mathbf{W}$ determined from the residual magnitudes, exactly as for the single-site robust estimator. We used a similar approach here, but with some further refinements. In particular, the weights w_i used at each iteration are chosen also to down-weight 'leverage points' (Chave & Thomson 1989), and to allow for outliers at the remote site. Additional details and discussion are provided in Egbert & Booker (1986), Chave *et al.* (1987), Chave & Thomson (1989), Egbert & Livelybrooks (1996) and Egbert (1997).

For all of the results discussed here, site PKD is used as the local site, and SAO serves as the remote reference. The continuous data stream was subdivided into segments of 24 hr duration, each starting at midnight UT. From the 2 yr period 1996–1997 there were instrumentation problems on 143 days. Results obtained for each of the remaining 588 days form the basis of our analysis. The 24 hr (86400 samples) time-series for each day were subdivided into sets of 128 points with 32 points overlap, pre-whitened and Fourier-transformed. This procedure was applied to three subsequent decimation steps (Egbert & Booker 1986) in which the original time-series were low-pass filtered and decimated by a factor of four, resulting in estimates of MT impedance tensors and vertical magnetic field TFs at 28 periods covering the range 2.3–2300 s.

3 STABILITY OF TRANSFER FUNCTIONS

Although the two stations are separated by almost 150 km, there was clear evidence for coherent noise and source inhomogeneity in the period range 10–100 s (Egbert *et al.* 2000). Interstation TFs (for example, relating magnetic fields at SAO to PKD) exhibited large systematic biases that varied rapidly with frequency and time of day. Egbert *et al.* (2000) showed that two sources contributed to these biases: the Bay Area Rapid Transit (BART) DC railway system (almost 300 km north of Parkfield) and inhomogeneous source fields generated by highly localized ionospheric resonance associated with Pc3 pulsations. Although not widely appreciated by MT researchers, the relatively short spatial scale of mid-latitude Pc3 pulsations is well documented (e.g. Andersen *et al.* 1976; Lanzerotti *et al.* 1981; Waters *et al.* 1991). Although these source heterogeneities have a very great effect on interstation TFs, effects on local MT impedances and vertical-field TFs at the distant Parkfield site are very subtle, as we shall show. In essence, external field length scales still always greatly exceed EM skin depths, so that a uniform source local impedance remains a good approximation (e.g. Dmitriev & Berdichevsky 1979), even though source heterogeneity is quite clear when comparing data from the two sites. The local PKD TFs thus remained generally stable throughout the 2 yr period of this study, in spite of the clear evidence for short-wavelength sources in the two station array.

To study temporal variations of the TFs we consider deviations of the daily estimates from the long-term average, which we compute as the median of the 588 daily averages and denote $\bar{\mathbf{Z}}(\omega)$. Fig. 2 shows the apparent resistivities $\bar{\varrho}_{xy}$, $\bar{\varrho}_{yx}$ and phases $\bar{\varphi}_{xy}$, $\bar{\varphi}_{yx}$ derived from the off-diagonal components of $\bar{\mathbf{Z}}$. As might be expected given the large amount of data, the curves are very smooth over the full period range. Relative variations in apparent resistivity $\Delta\varrho_{xy}=(\varrho_{xy}-\bar{\varrho}_{xy})/\bar{\varrho}_{xy}$ are plotted for all good days and all periods in Fig. 3(a). The normalization by $\bar{\rho}_{xy}$ allows for the large variation with period of ϱ, and makes $\Delta\varrho$ approximately proportional to the variation in the logarithm of apparent resistivity. In fact, it is readily verified that $\Delta\varrho_{xy}\approx 2\,\mathscr{R}e[\ln Z_{xy}-\ln \bar{Z}_{xy}]$, where ln here represents the complex natural logarithm. Variations of the phases $\Delta\varphi_{xy}=\varphi_{xy}-\bar{\varphi}_{xy}$ are plotted in Fig. 3(b). Since the phase variations are $\mathscr{I}m[\ln Z_{xy}-\ln \bar{Z}_{xy}]$, with appropriate scaling the two parts of the figure can be compared directly. Under the usual assumption that errors in the real and imaginary parts of the impedance are uncorrelated and of equal magnitude, a 1 per cent error in $\Delta\varrho$ should correspond to an error in $\Delta\varphi$ of 0.289°. We use this scaling factor in Fig. 3 (and subsequent figures) to make variations of apparent resistivity and phase directly comparable.

Note that we present all results here in the original measurement coordinate systems (with x geographic North). Evidence from more extensive MT surveys across the San Andreas Fault

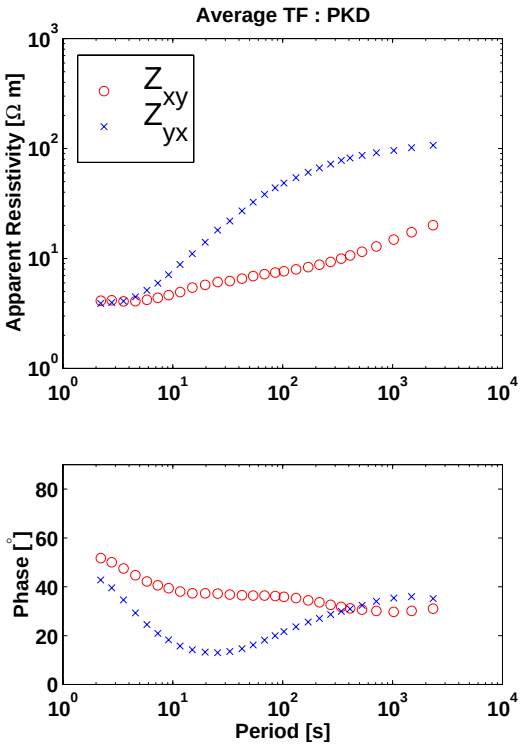

Figure 2. Apparent resistivity and phase curves for site PKD computed from the median of all daily impedance elements. Resistivities are computed in the measurement coordinate system, with x geographic north and y east. Error bars are smaller than symbol size.

near Parkfield suggest a geoelectric strike of about 45° (Unsworth *et al.* 1999). However, as the focus of this study is on TF variations and error bar reliability we have, for simplicity, chosen to keep estimated quantities in the original coordinate systems to avoid the need to consider the additional complications in error bar estimation that arise from rotation. See Appendix A for some further discussion.

Fig. 3 shows that the variations have a generally random appearance, and are of roughly similar magnitude for apparent resistivity and phase. Relative variations of ϱ_{xy} are typically less than 5 per cent, except at the longest periods. To characterize the temporal pattern of variations in impedance estimates better we calculated amplitude spectra for each period band. Averages of these spectra over broader (decade-wide) bands are plotted in Fig. 4. In all bands the variations have a nearly white spectrum for periods less than about 100 days. Amplitudes increase somewhat at longer periods, suggesting that there are some persistent (or temporally correlated) variations in apparent resistivities.

The longer-period variations are only barely discernible in Fig. 3. To emphasize these possibly systematic variations we applied a 2-D median filter to the array of residuals. The filter used has a length of 11 days along the time axis and a width of five bands along the period axis. The procedure is demonstrated using the xy-component of apparent resistivity. Starting from the relative residuals $\Delta\varrho(k, j)$, $k=1, K$; $j=i, J$, where K is the number of periods and J is the number of days, the 2-D median filter is calculated as

$$\tilde{\Delta}\varrho(k,j) = \text{median}[\Delta\varrho(k-2,j-5), \ldots, \Delta\varrho(k-2,j+5), \quad (10)$$

$$\Delta\varrho(k-1,j-5), \ldots, \Delta\varrho(k-1,j+5), \Delta\varrho(k,j-5), \ldots,$$

$$\Delta\varrho(k,j+5), \Delta\varrho(k+1,j-5), \ldots, \Delta\varrho(k+1,j+5),$$

$$\Delta\varrho(k+2,j-5)\ldots, \Delta\varrho(k+2,j+5)].$$

The filtered variations in resistivity are shown for the xy-mode in Fig. 5(a). This section clearly shows an increase in ϱ_{xy} from 1996 to 1997, a sharp decrease in ϱ_{xy} near day 80 of 1996, and some instances of persistent biases at the longest periods. There is a general tendency for the log apparent resistivity to shift up or down consistently across the full range of periods. Subtracting the smoothed part of the residuals yields high-passed day-to-day variations $\delta\varrho = \Delta\varrho - \tilde{\Delta}\varrho$, which have a completely random appearance (Fig. 5b). Calculation of

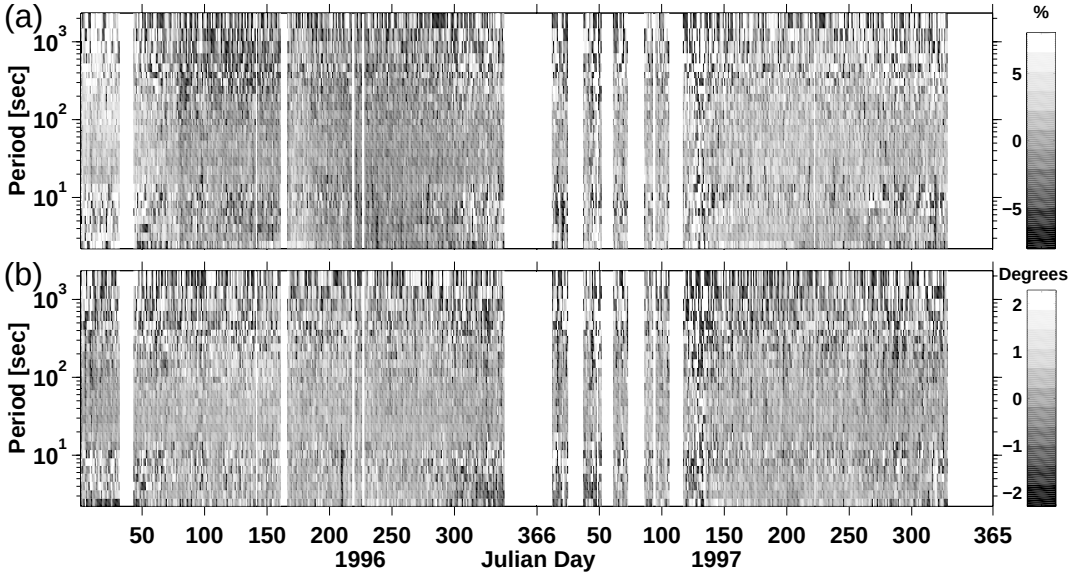

Figure 3. (a) Deviation of the daily estimates of the xy-mode of apparent resistivity from the long-term averages plotted in Fig. 2. Residuals are normalized by the long-term averages and are expressed in per cent. The figure thus gives variations about the average of log apparent resistivity. (b) Deviations of phase from long-term averages. As discussed in the text, plotting ranges for the variations in phase and log resistivity are compatible.

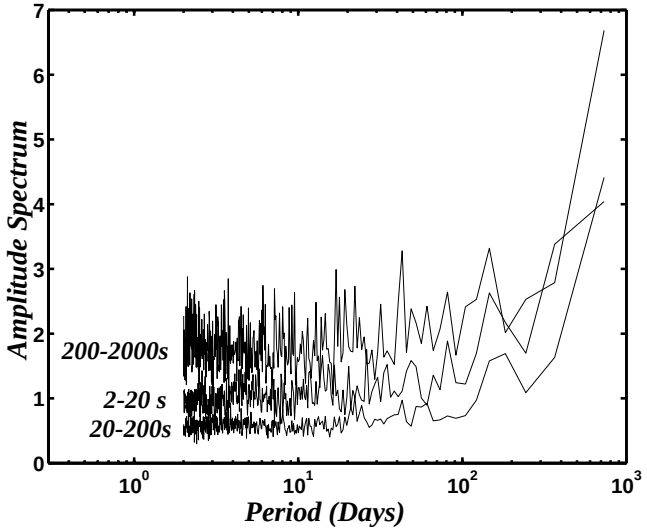

Figure 4. Amplitude spectra for day-to-day variations in ln ϱ_{xy} averaged over the broad period bands 2–20, 20–200 and 200–2000 s. For periods less than about 100 days the spectra are flat, indicating that except for some weak persistent slow variations, day-to-day variations are essentially random.

the autocorrelation of $\delta\varrho$ reveals no serial correlation for either mode, consistent with the flatness of the spectra of Fig. 4 for periods below 100 days. We also computed correlations of $\delta\varrho$ between period bands on fixed days. There is a weak positive correlation between residuals for adjacent bands (Fig. 6). Correlations are strongest (up to 0.2) for periods from 10–100 s.

The same analysis was applied to the phase variations and to the other TF components with similar results. Average amplitudes of the day-to-day variations of log apparent resistivity and phase are summarized for both modes in Fig. 7. Resistivities are most stable for periods of 20–200 s, with typical variations in this band being about 2 per cent. For periods beyond 200 s the amplitude of relative variations increases rapidly, exceeding 10 per cent at 2000 s. The increase in variability is approximately consistent with the decreasing number of degrees of freedom in the longer-period TF estimates. There is also a peak in variation amplitude centred at a period of around 10 s, where relative variations increase to 3–4 per cent. The pattern of day-to-day variations in phases generally tracks that of the apparent resistivities. Note that we have plotted the log resistivity and phase variations on a consistent scale (that is, 1 per cent variation corresponds to 0.289°). The fact that the phase variations are always slightly below the amplitude variations on this common scale implies that the phases are actually slightly more stable than the resistivities. This contradicts the usual folklore that phases are the more difficult quantity to determine accurately. However, a similar pattern has also been observed in strongly biased MT data collected in the vicinity of the German Deep Drilling site (Eisel *et al.* 2000).

To display the long-term variations of the TFs more compactly we subdivided the full period range into three sub-bands (2–20, 20–200 and 200–2000 s), averaged the residuals over these bands and low-passed the results with a median filter with a length of 25 days. Fig. 8 shows the time variations of the broad-band averaged residuals for apparent resistivities, phases and vertical magnetic TFs. The biggest variations in all components occurred during the first 170 days of 1996. Variations in the vertical magnetic field TFs during this period are especially great, with several clear offsets in the TF estimates. These offsets correspond to equipment changes, which are marked by the dashed vertical lines. These equipment changes also appear to coincide with large variations in apparent resistivities. Probably at least some of the variability in resistivity in the first quarter of the 2 yr period (and certainly the tears in vertical-field TFs) results from poorly calibrated variations in the system response.

Beyond day 170 of 1996, apparent resistivities exhibit the greatest fluctuation. Variations in both modes appear to be negatively correlated, and are very similar for all three

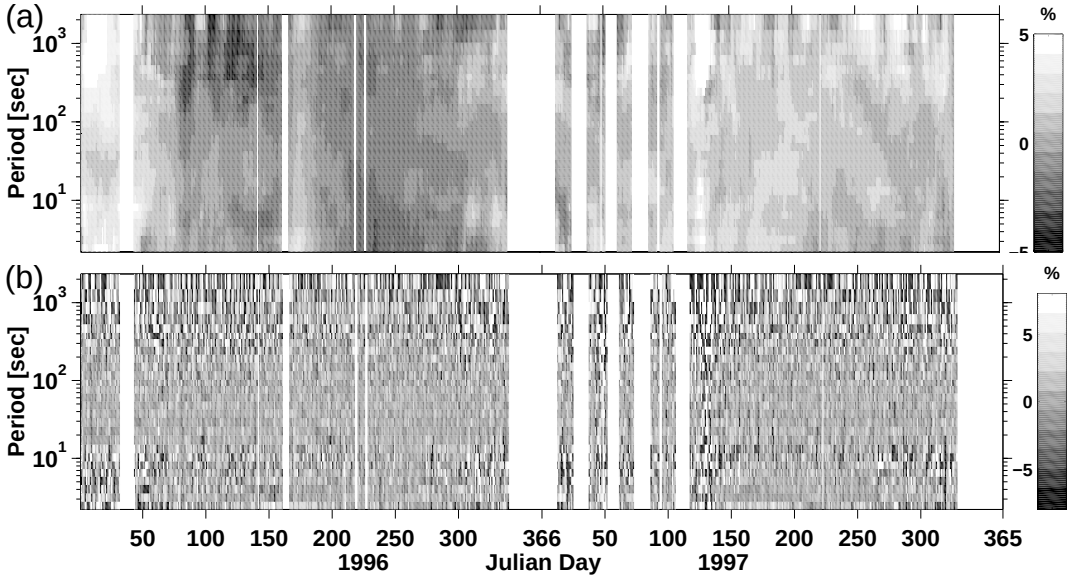

Figure 5. (a) Low-pass filtered variations in ln ϱ_{xy} expressed in per cent. A median filter of dimensions 5 (in the period direction) and 11 (in the time direction) was applied to the normalized residuals of Fig. 3. (b) Variations in ln ϱ_{xy} after subtraction of the low-pass-filtered (persistent) variations of (a).

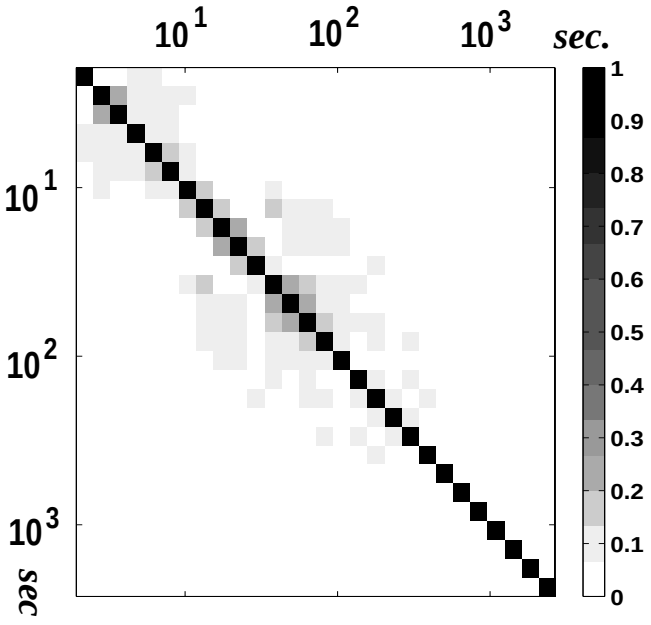

Figure 6. Correlation between period bands of variations of ln ϱ_{xy} averaged over time. There is a weak positive correlation between adjacent bands.

period bands. There were no significant instrumental changes at PKD during this period, so these variations in resistivity of about 1 per cent may be real. A possible explanation that would be consistent with the frequency-independent shift of log resistivity is that changes in the near-surface conditions result in fluctuations in static shifts. Changes in rotation of electric fields due to fluctuations in distortion twist (Groom & Bailey 1989) could possibly explain the negative correlation between variations in the two modes. However, since calibrations were not checked during this time, drifts in instrument response cannot be ruled out as an alternative explanation.

Long-term variations of phases are generally smaller, consistent with the generally smaller total variability of phases seen in Fig. 6. Again, note that we plot variations in phase and log

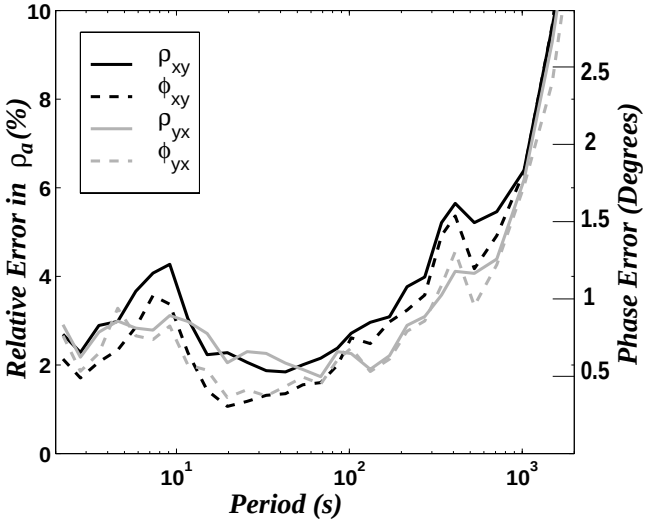

Figure 7. Average amplitude of variations of log apparent resistivities and phases as a function of period for *xy*- and *yx*-modes. Variations in ln ϱ_{xy} expressed in per cent.

apparent resistivity on consistent scales. Phase estimates are generally stable to within a few tenths of a degree. Except for the early problems with instrument calibration, the vertical-field TFs are also quite stable. There is no clear evidence of any seasonal variations in any of these components, as might be expected due to seasonal variations in near-surface hydrology. However, a tendency to noisier conditions and more missing data during winter months makes seasonal comparisons somewhat problematic.

We conclude that the MT TFs are stable to within a few per cent or better, at least when instrument configurations are unchanged. Unambiguous detection and quantification of systematic changes of this small magnitude will require extreme care in the installation and calibration of instrumentation. Some control over possible instrument drift would also be highly desirable. The inevitable need for replacement of instrument components makes monitoring variations of this level extremely challenging.

4 ESTIMATION OF VARIANCES AND TF VARIABILITY

In the previous section we showed that variations in TF estimates are small and predominantly random. Are the magnitudes of these random day-to-day variations consistent with the error bars computed for the TF estimates? In this section we consider this question with error bars for the RME calculated with the standard asymptotic theory. In the next section we address the question for several variants on the jackknife scheme proposed for MT TF error bar estimation by Chave & Thomson (1989). We first review the standard theory of error variances for LS estimates, along with the modifications needed for the RME and remote reference estimates.

4.1 Error bar calculations for robust remote reference estimates

The covariance of the linear LS TF estimate of (4) is readily derived from a simple linear propagation of errors:

$$\mathrm{Cov}(\hat{\mathbf{Z}}) = \hat{\sigma}^2(\mathbf{H}^{\dagger}\mathbf{H})^{-1}, \quad \hat{\sigma} = (I-2)^{-1}\sum_i |\eta_i|^2, \tag{11}$$

where η_i are again the residuals. As (11) is a linear equation relating two input variables to one output, $\hat{\mathbf{Z}}$ is a two-element vector and consequently $\mathrm{Cov}(\hat{\mathbf{Z}})$ is a 2×2 covariance matrix. Under the usual LS assumptions that all noise is in the predicted (electric) channels, that noise has zero mean and is independent of the signal, and that the noise variance σ^2 is a finite constant (independent of i), this result is exact. Note that no distributional assumptions on the errors (beyond the condition of finite second moments) are required for this variance formula to hold. Furthermore, as the number of data points I increases, the distribution of $\hat{\mathbf{Z}}$ will converge to a Gaussian (e.g. Graybill 1976). Thus, for large enough samples, confidence intervals for the TF estimates can be based on tables of the Gaussian distribution, independent of any distributional assumptions about the errors in the frequency domain FCs. Note, however, that the other assumptions about the errors stated above must still hold for these LS confidence intervals to be accurate. Furthermore, the definition of 'large enough' required for asymptotic normality to hold is often unclear in practice.

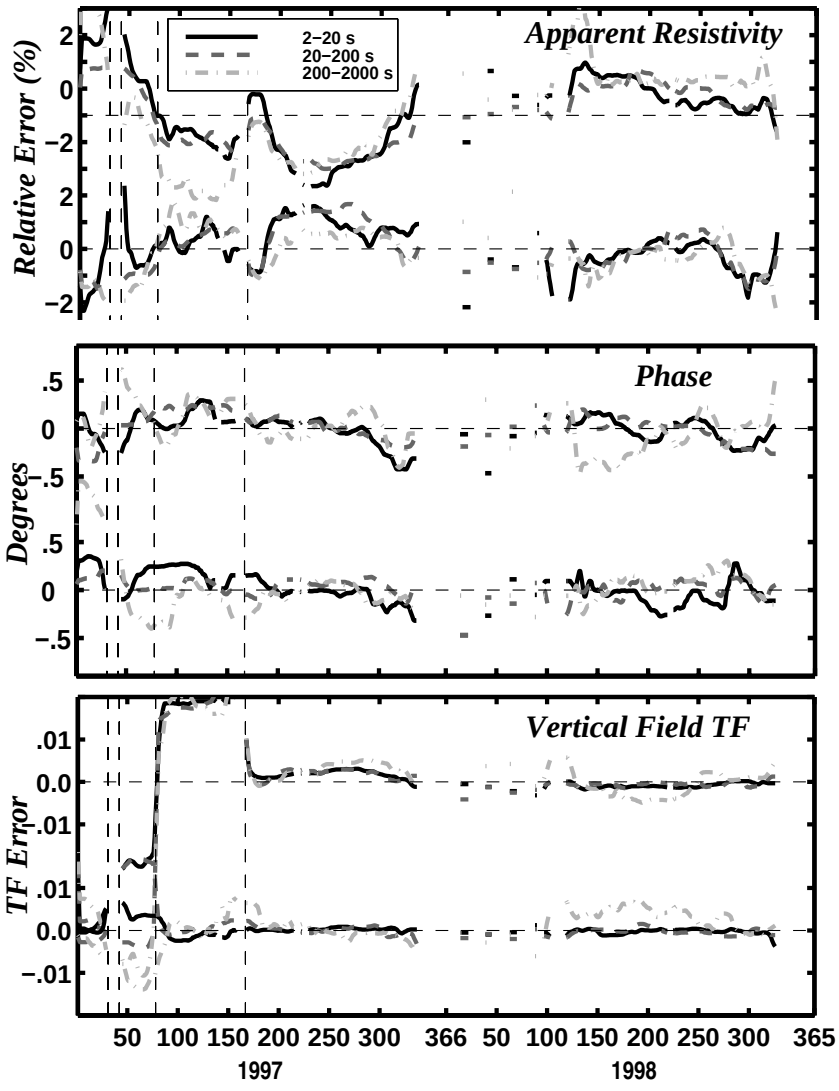

Figure 8. Band-averaged, low-pass-filtered residuals of apparent resistivities, phases and vertical magnetic field TFs. Equipment changes are marked by vertical dashed lines. The three bands are 2–20, 20–200 and 200–2000 s.

For the RME, variances are a bit more complicated because the estimate of the TF cannot be explicitly written as a linear combination of the data. Huber (1981) showed that the RME of the TF is asymptotically Gaussian with covariance

$$\mathrm{Cov}(\hat{\mathbf{Z}}) = \frac{\sigma^2 E(\psi^2)}{(E\psi')^2} (\mathbf{H}^\dagger \mathbf{H})^{-1}, \tag{12}$$

where E denotes the expectation and ψ is the influence function defined in Appendix B. Approximating σ^2, $E(\psi^2)$ and $E\psi'$ in the natural way leads to the estimate (Egbert & Booker 1986)

$$\mathrm{Cov}(\hat{\mathbf{Z}}) = \frac{(I-2)^{-1} \sum_i w_i^2 |\eta_i|^2}{\left[I^{-1} \sum_i \psi'(|\eta_i|/\hat{\sigma}) \right]^2} (\mathbf{H}^\dagger \mathbf{H})^{-1}, \tag{13}$$

where η_i, w_i, $i=1$, I and $\hat{\sigma}$ are the residuals, weights and error scale from the final iteration. Again, for large enough sample sizes, these variance estimates can be used with tables of the Gaussian distribution to construct confidence intervals, provided the other assumptions about the errors hold. Huber (1981, section 7.6) discussed some variants on this error variance estimate, but as discussed in Appendix A, eq. (13) offers some advantages for MT data processing, where multiple output components are processed simultaneously and later mixed together by impedance rotation.

The weighted LS algorithm for estimating the RME suggests a somewhat simpler approach to error bar calculations: fix the weights from the final iteration and apply the standard formulae (11) modified to account for the weights. Explicitly, the estimate of error covariance would then be

$$\mathrm{Cov}(\hat{\mathbf{Z}}) = (I-2)^{-1} \sum_i w_i |\eta_i|^2 (\mathbf{H}^\dagger \mathbf{W} \mathbf{H})^{-1}. \tag{14}$$

Comparison to eq. (13) reveals significant differences between these two expressions. In fact, eq. (14) is not in general correct, and Huber (1981, pp. 174–175) explicitly warned against use of this formula, especially for non-redescending influence functions ψ. For the case where a redescending ψ with a very sharp cut-off is used [as is the case for the final iterations in our processing, as well as in the scheme outlined by Chave *et al.* (1987)], the similarity between eqs (13) and (14) appears

to be greater. Now weights are essentially always 0 or 1, so $\Sigma_i w_i|\eta_i|^2 \approx \Sigma_i w_i^2|\eta_i|^2$. However, in general $(\mathbf{H}^\dagger\mathbf{W}\mathbf{H})^{-1}$ will converge to $(E\psi')^{-1}(\mathbf{H}^\dagger\mathbf{H})^{-1}$, so we should in general expect systematic biases when eq. (14) is used. Note also that the calculation of $E\psi'$ requires some care when a redescending influence curve with a sharp cut-off is used (see Appendix B).

For the remote reference case that allows for random errors in all data channels, error bars can be derived by linearized error propagation. The derivation given in Gamble *et al.* (1979b) can be extended to show that the full covariance matrix for the remote reference estimate can be approximated by

$$\mathrm{Cov}(\hat{\mathbf{Z}}) = \hat{\sigma}_\eta^2(\mathbf{R}^\dagger\mathbf{H})^{-1}(\mathbf{R}^\dagger\mathbf{R})(\mathbf{H}^\dagger\mathbf{R})^{-1} \tag{15}$$

$$\hat{\sigma}_\eta = (I-2)^{-1}\sum_i |\eta_i|^2 .$$

With the standard statistical assumptions about moments, it is straightforward to show that this is the correct large sample covariance of the remote reference estimate, which is again asymptotically Gaussian. Eqs (11) and (15) have the same form, with $(\mathbf{R}^\dagger\mathbf{H})^{-1}(\mathbf{R}^\dagger\mathbf{R})(\mathbf{H}^\dagger\mathbf{R})^{-1}$ in eq. (15) being equivalent to $(\mathbf{H}^\dagger\mathbf{H})^{-1}$ in eq. (11). This similarity suggests a simple general format for storing the full impedance covariances, allowing error bars to be calculated correctly in any rotated coordinate system. This is further discussed below and in Appendix A.

The similarity also suggests a simple generalization of eq. (12) to the robust remote reference case:

$$\mathrm{Cov}(\hat{\mathbf{Z}}) = \frac{(I-2)^{-1}\sum_i w_i^2|\eta_i|^2}{\left[I^{-1}\sum_i \psi'(|\eta_i|/\hat{\sigma})\right]^2}(\mathbf{R}^\dagger\mathbf{H})^1(\mathbf{R}^\dagger\mathbf{R})(\mathbf{H}^\dagger\mathbf{R})^{-1} . \tag{16}$$

There are possibly some theoretical issues that should be addressed with regard to this approximation. Here we assume that this approximation is adequate, and use eq. (16) for the error bar calculations for our robust remote reference TF estimates. We will sometimes refer to these as the 'standard' error bars to contrast with the jackknife approach we consider below. Note, however, that this 'standard' approach has probably not been widely implemented.

4.2 Comparison of standard error bars to TF estimate variability

We now compare the day-to-day random variability in the remote reference TF estimates to the error bars estimated using eq. (16). Variance estimates computed from a single day of data cannot account for systematic long-period variations due to seasonal trends or changes in instrumental response, so we base our analysis on the high-pass filtered residuals $\delta\varrho(k, j)$, $k=1, K; j=i, J$, where again K is the number of periods and J is the number of days. For each day and each period we form the ratio $R(k, j) = \delta\varrho_{xy}(k, j)/\sigma_{\varrho_{xy}}(k, j)$, where $\sigma_{\varrho_{xy}}$ are the standard errors computed using eq. (16).

The distribution of the ratios R provides a measure of how well the computed error bars represent the actual variability of the TF estimates. If the individual TF estimates can truly be treated as Gaussian with the specified error bars (as implied for large samples by the asymptotic theory, and as assumed for standard computations of confidence intervals), the distribution should approximate a standard Gaussian with zero mean and unit variance. In Fig. 9 we plot histograms of R separately for each period for ϱ_{xy}. There are clear deviations from this idealized form, especially in the period range 10–100 s, where a large number of outliers occur, with a significant number of deviations from the long-term average by four or more standard errors.

To give some idea of deviations of the distributions of normalized residuals from the idealized Gaussian form, we plot the percentage of residuals for each period that are less than

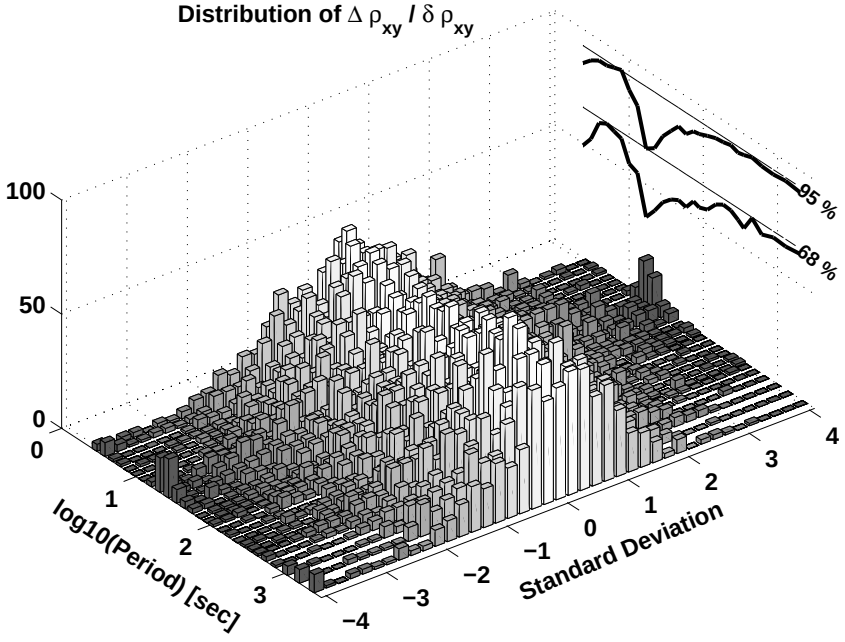

Figure 9. Distributions of the ratios between filtered, normalized residuals and the standard errors $R(\varrho) = \Delta\bar{\varrho}/\delta\varrho$ for each period. The sum of all contributions between ± 1 (respectively ± 2) standard deviations is shown as solid lines on the right-hand side of the plot. Thin lines indicate the expected 68 per cent (respectively 95 per cent) contributions. The z-axis is the number of samples in each column.

one and two standard errors. In the period range 10–100 s the fraction of residuals within one and two standard errors is significantly smaller than would be expected for the idealized Gaussian distribution (68 and 95 per cent respectively). The error bars clearly do not reflect the true variability of the TFs very well in this band. We hypothesize that these difficulties result from subtle temporally correlated variations in the MT impedance caused by source complications in this period range (Egbert *et al.* 2000). Since the impedance depends at least weakly on the external source wavelengths, variations in the spatial spectrum of sources can result in slight variations in the local MT impedance. In particular, the MT impedance resulting from cultural sources in the San Francisco Bay area (mostly BART) will differ slightly from the impedance due to natural ionospheric and magnetospheric sources. As the ratio of power in these components varies from day to day (probably mostly due to variations in natural source signal levels), the impedance will also vary. Note that these variations are small, and biases in the MT impedance are not particularly noticeable at these periods in plots of estimates from individual days. However, error bars in this frequency range are also quite small, and apparently do not adequately account for these variations in the impedance.

To summarize the distribution of normalized residuals for a more complete set of TF components we use a different display format. In Fig. 10 we plot the 68th and 95th percentiles of the distributions of normalized residuals for each period, for resistivities and phases of the off-diagonal elements of the impedance tensor, and for the real and imaginary parts of the vertical magnetic field TF. If the TF estimates have a truly Gaussian distribution and error estimates are correct, the curves should plot at ± 1 (for 68 per cent) and ± 2 (for 95 per cent). Larger values indicate that wider confidence intervals would be required to actually achieve coverage probabilities of 68 or 95 per cent. For example, Fig. 10 shows that at a period of 15 s a 95 per cent confidence interval for ϱ_{xy} would have to be four times the estimated standard error to truly contain the long-term mean 95 per cent of the time. For ϱ_{xy} the 95 per cent confidence interval would have to be three standard errors wide. For periods from 10 to 100 s the actual 68 and 95 per cent confidence limits are larger than implied by error bars computed with the standard asymptotic approach for all parameters. Again, we believe that this reflects temporally varying subtle biases due to finite spatial scale source effects in this period range. At longer and shorter periods confidence limits derived in the conventional manner using the standard asymptotic error bars are quite accurate.

5 THE JACKKNIFE METHOD

As the results from the previous section suggest, standard estimates of error bars for MT TF estimates may not always be completely satisfactory. As a possible remedy, Chave & Thomson (1989) proposed application of the non-parametric jackknife method for estimation of MT TF error bars. General discussions of the jackknife method are given in Efron (1982) and Efron & Gong (1983). These methods are discussed further in the context of spectra and coherence estimates by Thomson & Chave (1991).

5.1 Review of basic ideas

In its simplest form the jackknife procedure entails successively deleting each single line of data in eq. (1) and calculating the delete-one TF estimates $\hat{\mathbf{Z}}_{-i}$, $i=1, \ldots, I$ for each of the I data subsets. From these and the estimate based on all data ($\hat{\mathbf{Z}}$), the jackknife pseudo-values

$$\mathbf{P}_i = I\hat{\mathbf{Z}} - (I-1)\hat{\mathbf{Z}}_{-i} \tag{17}$$

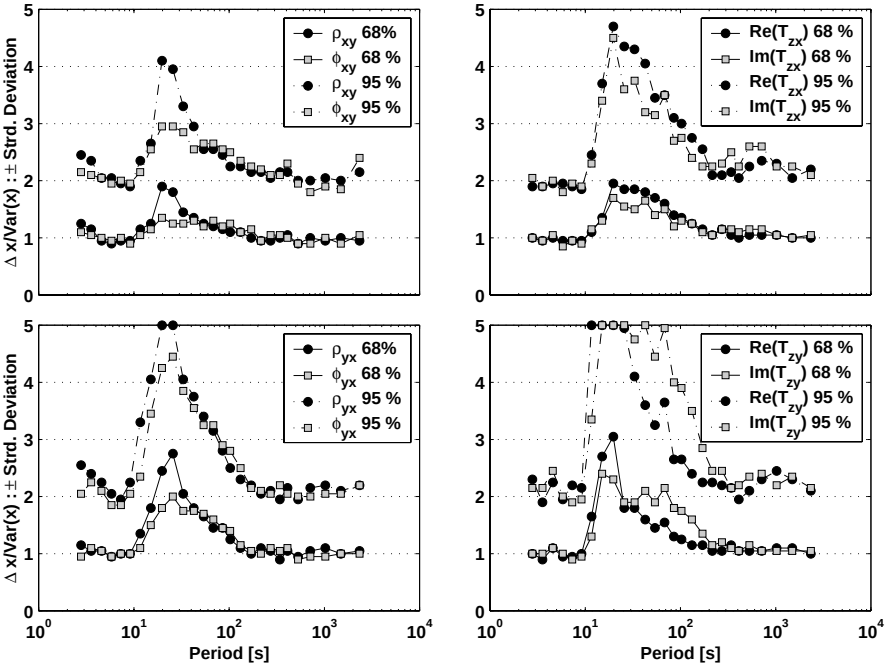

Figure 10. Standard deviation bounds that contain 68 per cent (respectively 95 per cent) of the ratios between TF residuals and their variances. Displayed are values for apparent resistivities, phases and real and imaginary parts of the vertical magnetic field TFs. Variances are calculated in the standard way.

are computed. The jackknife estimate of the TF is then given by the mean of the pseudo-values,

$$\tilde{\mathbf{Z}} = \frac{1}{I}\sum_{i=1}^{I} \mathbf{P}_i \,, \tag{18}$$

and the jackknife covariance estimate is computed as

$$\tilde{\mathbf{S}} = \frac{1}{I(I-2)}\sum_{i=1}^{I} (\mathbf{P}_i - \tilde{\mathbf{Z}})(\mathbf{P}_i - \tilde{\mathbf{Z}})^{\dagger} \,. \tag{19}$$

We are concerned here with the jackknife estimate of error covariance $\tilde{\mathbf{S}}$, and we will not consider the jackknifed TF estimate of eq. (18) further.

Use of the jackknife to calculate error bars is most clearly justified when an estimator is non-linear in the data. In this case some sort of linearization and asymptotic approximation is required, as discussed above for the RME error bars of eq. (12). The full estimation algorithm used for the MT TFs considered here is in fact quite complicated, with a number of *ad hoc* weighting or data selection steps. Initial time-series processing required to transform to the frequency domain adds further complications (e.g. serial correlation). To apply eq. (12), which was derived for a highly idealized regression model, we must make a number of additional approximations and assumptions. Some of these (for example, correction for serial correlation due to tapering before Fourier transforming the data) are discussed in Egbert & Booker (1986). These complexities suggest that the jackknife, which provides a relatively easy way to compute error bars of complex estimators without working out all of the details, might be usefully applied to MT TF error bar calculation. The jackknife might also be expected to be useful for error analysis of tensor decomposition schemes (Groom & Bailey 1989; Chave & Smith 1994) that involve non-linear transformations of the impedance tensor elements, leading to further complications in error distributions.

There are some other potential advantages to the jackknife. Hinkley (1977) showed that in a regression context an appropriately defined jackknife is robust to violations of the assumption of variance homogeneity; this is one of the assumptions implicit in the RME error bars of eq. (5), and is something we shall return to. Other evidence (Efron & Stein 1981) shows that the jackknife always produces larger error bars when applied to data that are not identically distributed. Some other claimed advantages of the jackknife are less clear. The suggestion that the jackknife avoids distributional assumptions (Chave & Thomson 1989) is of questionable relevance, since the large sample results of the previous section require essentially the same conditions on second moments as the jackknife (Miller 1974). This is in contrast to many of the examples considered in Thomson & Chave (1991) (e.g. estimation of variances for power spectral densities), where the asymptotic results depend more explicitly on distributional assumptions (or at least on higher-order moments). It is not obvious how the methods should compare for small samples, although Thomson & Chave (1991) reported some evidence (in other applications) that suggests that the jackknife should perform better on small samples than more conventional asymptotic methods. In any event it is worth noting that the jackknife is used for estimating variances of estimators and cannot be directly used for the computation of confidence limits (in contrast to the bootstrap method; see Efron 1982). Confidence intervals computed with jackknifed errors thus require use of an asymptotic result (of the same sort used for the standard RME) to translate variances into confidence limits.

For modern MT data the sample size I in eq. (17) can be extremely large, making application of the full jackknife procedure problematic. For one day of 1 Hz data, with TFs estimated in the usual fashion evenly spaced in log period with approximately eight bands per decade, I exceeds 10 000 for the shortest period bands. For five of the 28 estimation bands I exceeds 5000. A literal application of the jackknife requires completely redoing the full estimate once for each data point, leading to an increase in total processing time for each band by a factor of I. Since processing time is dominated by iterating the robust scheme for the shortest period bands, a literal application of the jackknife requires an enormous increase in overall processing time; for our case the increase is by more than a factor of 5000. This turns processing runs that might take 30 s or so on a modern workstation into calculations that take a few days. The time required for a literal application of the jackknife to error bar computation for robust MT TF estimation is thus at best marginally feasible. For the study reported here, where we needed to repeat the whole calculation 588 times (once for each day), a literal application of the jackknife was not practical.

We consider two possible ways to overcome this problem. First, rather than repeat the full iterative regression M-estimate with each data point omitted, we perform the robust iterations once with all data. Using weights determined from the final iteration, we apply the jackknife to the weighted LS problem. This is the approach described by Chave & Thomson (1989). Second, instead of deleting each point in turn we delete subsets of data. This subset deletion approach was used in some of the earliest discussions of jackknife methods (e.g. Quenouille 1956), and has been applied to the MT problem by Larsen *et al.* (1996). By keeping the number of subsets small (e.g. deleting 5 per cent of the data results in 20 subsets), the overall processing cost can be kept to a practical level, while allowing implementation of the jackknife on the full non-linear robust procedure.

5.2 The fixed-weights jackknife

With the weights $w_i = w(|\eta_i|/\hat{\sigma})$ fixed from the final iteration of the RME, computation of the transfer function is reduced to a standard LS problem by multiplying each of $\mathbf{H}$, $\mathbf{E}$ and $\mathbf{R}$ by the diagonal weight matrix $\mathbf{W}^{1/2}$. We assume that this rescaling has been done to all matrices, so we need only consider application of the unbalanced jackknife to the standard LS problem. We refer to this approach as the 'fixed-weights jackknife' to distinguish this from a full jackknife of the non-linear robust remote reference estimate.

The TF model (1) is a special case of the general regression problem, which is unbalanced in the sense that each data point is not weighted equally. In contrast, the balanced jackknife described above treats all observations equally. Hinkley (1977) thus proposed an unbalanced jackknife with pseudo-values of eq. (17) replaced by

$$\mathbf{P}_i = \hat{\mathbf{Z}} + I(1-h_i)(\hat{\mathbf{Z}} - \hat{\mathbf{Z}}_{-i}) \,. \tag{20}$$

The weights h_i are defined in terms of the diagonal elements of the 'hat matrix'; see Chave & Thomson (1989) for a good

discussion of the hat matrix. To be explicit, if $\mathbf{H}_i=(H_{x,i}H_{y,i})^{\mathrm{T}}$ is the transpose of the ith row of $\mathbf{H}$ defined in eq. (1), then the weights for the single-station case can be given explicitly as $h_i=\mathbf{H}_i^{\mathrm{T}}\mathbf{D}_0^{-1}\mathbf{H}_i^*$, where $\mathbf{D}_0=\mathbf{H}^\dagger\mathbf{H}$, and the asterisk denotes the complex conjugate. The unbalanced jackknife estimate and variance are then computed from eqs (18) and (19) using the weighted pseudo-values defined in eq. (20). The simplest generalization of this definition of weights to the remote reference case yields

$$h_i=\mathbf{H}_i^{\mathrm{T}}\mathbf{D}_0^{-1}\mathbf{R}_i^*\,,\qquad \mathbf{D}_0=\mathbf{R}^\dagger\mathbf{H}\,, \tag{21}$$

where $\mathbf{R}_i=(R_{x,i}R_{y,i})^{\mathrm{T}}$ is the transpose of the ith row of $\mathbf{R}$. For the single-station case the h_i are guaranteed to be real and positive, but this is only approximately true for eq. (21). To generalize, the unbalanced jackknife to the remote reference case (Chave & Thomson 1989) thus replaced the weights of eq. (20) with $|h_i|$. Other approximations are possible, for example, using only the remote reference field channels in eq. (21) (A. Chave, personal communication, 1999).

Hinkley (1977) showed that the delete-one values Z_{-i} in eq. (20) and the jackknifed covariance estimate of eq. (19) can be given in essentially closed form for the single-station case. We now derive a similar result for the remote reference case. Let $\mathbf{E}_{-i}$, $\mathbf{R}_{-i}$ and $\mathbf{H}_{-i}$ be the matrices in eq. (1) with line i deleted. The delete-one solution $\hat{Z}_{-i}$ of the remote reference estimation problem (9) can then be written as

$$\begin{aligned}\hat{\mathbf{Z}}_{-i}&=(\mathbf{R}_{-i}^\dagger\mathbf{H}_{-i})^{-1}(\mathbf{R}_{-i}^\dagger\mathbf{E}_{-i})\\&=(\mathbf{R}^\dagger\mathbf{H}-\mathbf{R}_i^*\mathbf{H}_i^{\mathrm{T}})^{-1}(\mathbf{R}^\dagger\mathbf{E}-\mathbf{R}_i^*E_i)\,.\end{aligned} \tag{22}$$

Using the Sherman–Morrison–Woodbury formula for an invertible matrix $\mathbf{A}$ and vectors $\mathbf{u}$ and $\mathbf{v}$ (Golub & van Loan 1989),

$$(\mathbf{A}+\mathbf{u}\mathbf{v}^{\mathrm{T}})^{-1}=\mathbf{A}^{-1}-\frac{(\mathbf{A}^{-1}\mathbf{u})\mathbf{v}^{\mathrm{T}}\mathbf{A}^{-1}}{1+\mathbf{v}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{u}}\,, \tag{23}$$

and the definition of $\mathbf{D}_0$ in eq. (21), the delete-one estimate (22) can be written

$$\begin{aligned}\hat{\mathbf{Z}}_{-i}&=\left[\mathbf{D}_0^{-1}+\frac{(\mathbf{D}_0^{-1}\mathbf{R}_i^*)\mathbf{H}_i^{\mathrm{T}}\mathbf{D}_0^{-1}}{1-\mathbf{H}_i^{\mathrm{T}}\mathbf{D}_0^{-1}\mathbf{R}_i^*)}\right](\mathbf{R}^\dagger\mathbf{E}-\mathbf{R}_i^*E_i)\\&=\hat{Z}-\mathbf{D}_0^{-1}\mathbf{R}_i^*\mathbf{E}_i+\frac{\mathbf{D}_0^{-1}\mathbf{R}_i^*\mathbf{H}_i^{\mathrm{T}}\mathbf{D}_0^{-1}\mathbf{R}^\dagger\mathbf{E}}{1-h_i}\\&\quad-\frac{\mathbf{D}_0^{-1}\mathbf{R}_i^*\mathbf{H}_i^{\mathrm{T}}\mathbf{D}_0^{-1}\mathbf{R}_i^*E_i}{1-h_i}\\&=\hat{\mathbf{Z}}-\frac{\mathbf{D}_0^{-1}\mathbf{R}_i^*(E_i-\mathbf{H}_i^{\mathrm{T}}\hat{\mathbf{Z}})}{1-h_i}\\&=\hat{\mathbf{Z}}-\frac{\mathbf{D}_0^{-1}\mathbf{R}_i^*\eta_i}{1-h_i}\,,\end{aligned} \tag{24}$$

where η_i is the residual for data point i.

Eq. (24) allows us to calculate $\hat{\mathbf{Z}}_{-i}$ from the all-data estimate $\hat{\mathbf{Z}}$ without recomputing the full LS solution. Using eq. (24), together with eqs (18), (19), (21) and (17), the (fixed-weights) jackknife error estimates can be computed at negligible cost. Note that the validity of this efficient approach does not depend on how the weights for the balanced jackknife are chosen. In particular, it would work for the choice $|h_i|$ used by Chave & Thomson (1989), or for hat matrix elements based on the reference channels only.

Using the h_i defined in eq. (21) as weights in eq. (20), further simplification of the jackknifed covariance $\tilde{\mathbf{S}}$ is possible. In this case, using eq. (24), eq. (17) can be rewritten as

$$\mathbf{P}_i=\hat{\mathbf{Z}}+I(1-h_i)(\hat{\mathbf{Z}}-\hat{\mathbf{Z}}_{-i}) \tag{25}$$

$$=\hat{\mathbf{Z}}+I(1-h_i)\left(\hat{\mathbf{Z}}-\hat{\mathbf{Z}}+\frac{\mathbf{D}_0^{-1}\mathbf{R}_i^*\eta_i}{1-h_i}\right) \tag{26}$$

$$=\hat{\mathbf{Z}}+I\mathbf{D}_0^{-1}\mathbf{R}_i^*\eta_i\,. \tag{27}$$

Then,

$$\begin{aligned}\tilde{\mathbf{Z}}&=\hat{\mathbf{Z}}+\sum_{i=1}^{I}\mathbf{D}_0^{-1}\mathbf{R}_i^*\eta_i\\&=\hat{\mathbf{Z}}+\mathbf{D}_0^{-1}\sum_{i=1}^{I}\mathbf{R}_i^*(E_i-\mathbf{H}_i^{\mathrm{T}}\hat{\mathbf{Z}})=\hat{\mathbf{Z}}\,,\end{aligned} \tag{28}$$

so the jackknife estimate $\tilde{\mathbf{Z}}$ equals the usual remote reference estimate $\hat{\mathbf{Z}}$. The weighted jackknife covariance thus reduces to

$$\tilde{\mathbf{S}}=\frac{1}{I(I-2)}\sum_{i=1}^{I}(\mathbf{P}_i-\hat{\mathbf{Z}})(\mathbf{P}_i-\hat{\mathbf{Z}})^\dagger \tag{29}$$

$$=\frac{I}{I-2}\mathbf{D}_0^{-1}\left(\sum_{i=1}^{I}|\eta_i|^2\mathbf{R}_i^*\mathbf{R}_i^{\mathrm{T}}\right)(\mathbf{D}_0{}^{1})^\dagger\,.$$

Now let

$$\bar{\Sigma}_{\mathbf{R}}^{\mathrm{T}}=\frac{1}{I}\sum_i\mathbf{R}_i^*\mathbf{R}_i^{\mathrm{T}}=I^{-1}\mathbf{R}^\dagger\mathbf{R}\,,\quad \hat{\sigma}_\eta^2=\frac{1}{I}\sum_i|\eta_i|^2\,, \tag{30}$$

so that $\bar{\Sigma}_{\mathbf{R}}$ is the average power spectral density matrix for the reference components, and $\bar{\sigma}_\eta^2$ is the average residual power in the predicted E. We then have

$$\sum_{i=1}^{I}|\eta_i|^2\mathbf{R}_i^*\mathbf{R}_i^{\mathrm{T}}=\hat{\sigma}_\eta^2\mathbf{R}^\dagger\mathbf{R}+\sum_{i=1}^{I}(|\eta_i|^2-\hat{\sigma}_\eta^2)(\mathbf{R}_i^*\mathbf{R}_i^{\mathrm{T}}-\bar{\Sigma}_{\mathbf{R}}^{\mathrm{T}})\,, \tag{31}$$

so that eq. (29) can be written

$$\begin{aligned}\tilde{\mathbf{S}}=\frac{I}{I-2}\Bigg[&\hat{\sigma}_\eta^2(\mathbf{R}^\dagger\mathbf{H})^{-1}(\mathbf{R}^\dagger\mathbf{R})(\mathbf{R}^\dagger\mathbf{H})^{-\dagger}\\&+\sum_{i=1}^{I}(|\eta_i|^2-\hat{\sigma}_\eta^2)(\mathbf{R}_i^*\mathbf{R}_i^{\mathrm{T}}-\bar{\Sigma}_{\mathbf{R}}^{\mathrm{T}})\Bigg]\,.\end{aligned} \tag{32}$$

The first term in eq. (32) is identical to the standard remote reference error covariance of eq. (15) [with the additional factor $I/(I-2)$ correcting the degrees of freedom in the definition of the residual variance of eq. (30) in the usual way]. The second term in eq. (32) gives the covariance between variations in the squared residual magnitude $|\eta_i|^2$ and variations in the reference site signal power. Since we assume that the reference site is highly coherent with the MT signal at the local site, this correction factor will only be significant when there is a systematic correlation between local signal power and noise power. When there is a positive correlation between signal and noise powers (as reported for long-period vertical-field TFs by Egbert & Booker 1986), the fixed-weights jackknife would be

expected to yield larger error bars. As the analysis of Hinkley (1977) shows, these increased error bars should also more accurately reflect the true variability in the estimates.

Note that eq. (32) is an exact result. If there is no significant correlation between residual magnitude and signal power magnitude, there should be no consistent or systematic difference between error bar magnitudes computed with the fixed-weights jackknife approach and more standard LS theory, treating the weights from the final iteration as fixed. However, as discussed in Section 4.1, this approach to calculating error bars for the RME is not asymptotically correct. In contrast to the full jackknife, the fixed-weights jackknife thus cannot be expected to be exactly correct asymptotically either. However, in situations where there is significant correlation between signal and noise power, the fixed-weights jackknife might still yield more reasonable error bars. The second term in eq. (32) might also be added to the asymptotically correct form eq. (16), but this issue needs further study.

As with the standard approach we can compare the variance estimates obtained from the fixed-weights jackknife approach with the observed random day-to-day variability in the estimated TFs. As for Fig. 10 we compute the ratios as above, now dividing the TF residuals by the corresponding fixed-weights jackknife standard errors. Note that the jackknife and standard error calculations provided error bars for TF estimates. Error bars for the apparent resistivities and phases (which are simple non-linear functions of the TFs) are computed from these by linearization, under the assumption that the errors are a small perturbation. In this case this assumption is quite reasonable. Note that for the jackknife it would be possible to calculate error bars directly for apparent resistivity and phases, and thus avoid this linearization step.

The 68th and 95th percentiles of the ratio R are displayed in Fig. 11 for the same transfer function components as in Fig. 10. Error bars obtained with the fixed-weights jackknife approach are slightly more conservative than the standard ones. The analysis given above suggests that this probably reflects a small correlation between signal and noise power magnitudes. This is not too surprising, since at least some of the signal and noise observed at PKD result from the same source (that is, cultural noise from the San Francisco Bay Area). However, variances in the period range 10–100 s, where the bias problem is significant, remain much too small. Furthermore, at long periods the error bars for the MT impedance components are systematically too large (although only slightly). As the goal is to obtain an accurate assessment of true uncertainty in the TF estimates, this is as much of a problem as systematic underestimation of error magnitudes. Overall, the fixed-weights jackknife produces results that are very similar to a careful application of more classical asymptotic methods (i.e. the 'standard approach').

5.3 The subset deletion jackknife

The fixed-weights scheme considered above does not really make use of the main advantage of the jackknife—the ability to calculate error bars for complex non-linear estimators. To make application of the full jackknife practical we consider a simple (and very old: Quenouille 1956) variant on the jackknife in which a fixed fraction (e.g. 5 per cent) of the data is omitted for computation of each pseudo-value. For each subset of data we perform the full robust processing procedure. Pseudo-values, TF estimates and variances are calculated following eqs (17), (18) and (19), where I is now the number of subsets (for example, $I=20$ when the fraction of omitted data is 5 per cent), and the index i refers to subset number.

The only significant modification required of the standard jackknife procedure is in the definition of the weights h_i required for calculation of the pseudo-values in eq. (20). We calculate these using the sum over the diagonal elements of the

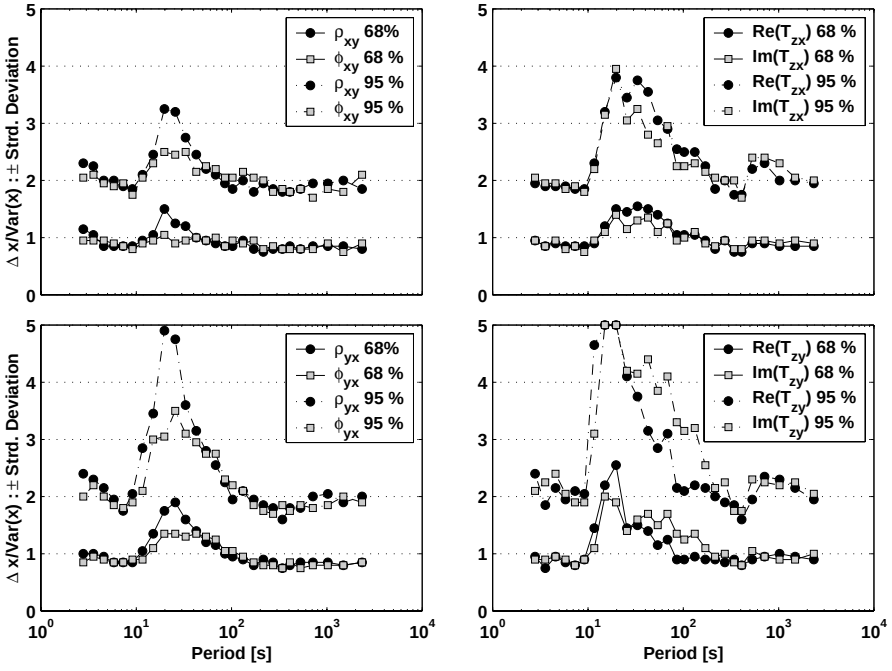

Figure 11. The same graphs as in Fig. 10 but using variance estimates from the first Jackknife approach together with the standard TF estimates.

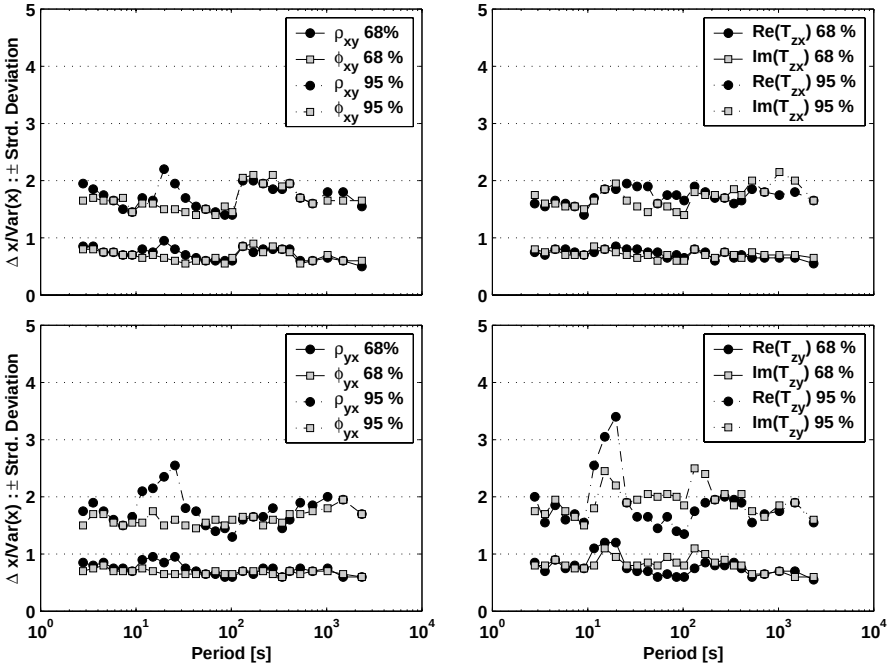

Figure 12. The same graphs as in Fig. 10 but using variance estimates from the second jackknife approach and omitting 5 per cent fractions in each subset. TF estimates are calculated in the standard way.

omitted data of the hat matrix (computed from the weighted data obtained from the final iteration of the robust scheme using all data).

To help assess the effect of the size of deleted subsets we applied this modified jackknife scheme using two different subset fractions, 5 and 0.5 per cent. Subsets consisting of 5 per cent of the continuous data samples correspond roughly to deleting a 1 hr window in time, while 0.5 per cent subsets correspond to about 7 min. For longer periods the shorter time windows must be longer than 7 min to allow deletion of even one point. At longer periods, for the 0.5 per cent subsets we are thus actually performing a standard delete-one jackknife. The results are shown in Fig. 12 for subsets of size 5 per cent and in Fig. 13 for subsets of size 0.5 per cent. The ratios between TF residuals and the 5 per cent deletion jackknife standard errors (Fig. 12) show significant improvements in the critical period band 10–100 s. However, error bars remain underestimated for some components (in particular for vertical-field TFs) in the

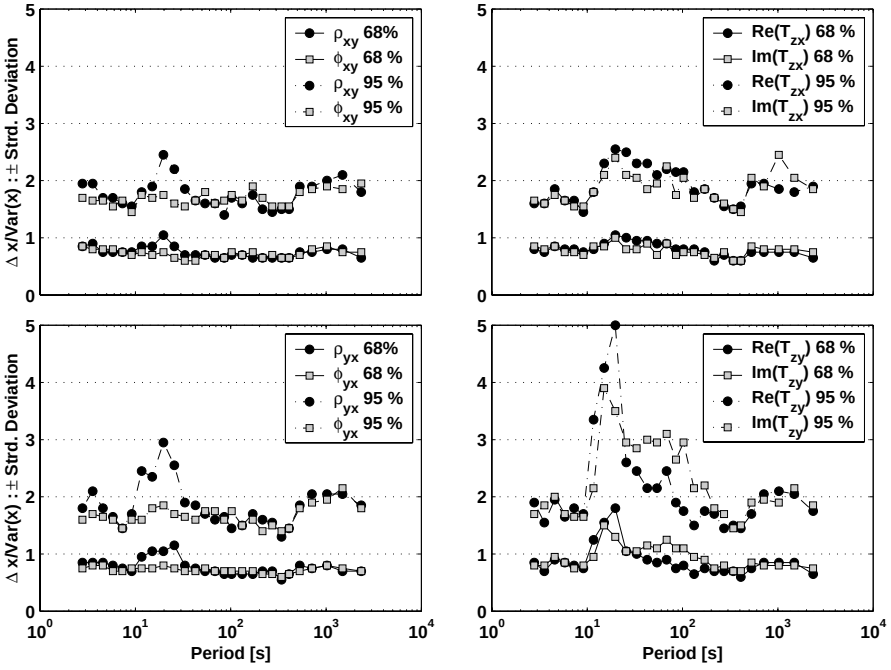

Figure 13. The same graphs as in Fig. 12 but omitting 0.5 per cent of data in each subset.

bias band. Furthermore, error bars computed with the subset deletion jackknife are often significantly larger than they should be.

Reducing the amount of omitted data in each subset to 0.5 per cent (that is, approaching the delete-one jackknife) tends to decrease the estimated variances across the full period band, making the percentiles of the normalized residual distributions larger. For this case the error bars in the 10–100 s bias band are again systematically too small. An explanation for the difference observed between the 5 and 0.5 per cent results might be that the coherent noise responsible for the biases varies consistently with time of day (Egbert *et al.* 2000). Omitting large sections of data thus results in great variability in the pseudo-values used to calculate the jackknife covariances. Deleting longer sections might be advantageous whenever there is time-varying coherent noise. On the other hand, the variances from the 5 per cent deletion case are often systematically too large, which is also undesirable.

6 CONCLUSIONS

The long-term MT time-series recordings from the PKD/SAO earthquake monitoring array offer the rare opportunity to study the long-term stability of MT TFs. With one day of data reliable estimates, with typical errors (determined as the deviation from long-term averages) of 2–5 per cent, were obtained up to a period of 1000 s. None of the standard MT TF components—apparent resistivity, impedance phase or vertical magnetic field TF—exhibited long-term systematic variations that were unambiguously above random background noise levels. Apparent resistivities varied the most, with slow frequency-independent variations of the order of 1–2 per cent that tended to be anti-correlated between modes. This pattern of variations is suggestive of slow changes in near-surface distortion effects. There was no obvious seasonal component to these variations, as might be expected if changes in near-surface hydrology were the cause. Variations in phases tended to be smaller. Possible drifts in instrument response and uncertainties in calibration cannot be ruled out, and generally complicate detection of such small long-term variations. Changes in TFs coincident with instrumentation changes were indeed significant, and clearly demonstrate the importance (and difficulty) of precise instrument calibration for this sort of monitoring experiment.

This data set also provides the opportunity for a detailed empirical study of the reliability of error bars for the impedance estimates. For periods from 10–100 s, error bars calculated using the standard asymptotic theory for the RME were too small by as much as a factor of two. There is evidence in this period range for coherent noise from a DC train system (BART) in the San Francisco Bay area 300 km to the northwest, and short spatial scales associated with natural Pc3 hydromagnetic oscillations (Egbert *et al.* 2000). These source complications apparently cause subtle temporal variations or biases in MT impedances that are not adequately accounted for by the statistical error bars. Outside of this 'bias band' these standard error bars provided a very accurate estimate of TF precision.

We also tested two variants on the jackknife method advocated by Chave & Thomson (1989) for MT TF error bar calculation. For the approximate fixed-weights jackknife (essentially what is described in Chave & Thomson 1989) we showed that the error estimate can be given in closed form. If properly implemented, additional computations required for this scheme are thus negligible. The explicit expression for the fixed-weight jackknife also clarifies the difference between this scheme and more conventional approaches. The fixed-weight jackknife allows for possible correlation between the magnitude of signal and noise, but otherwise should be similar to an asymptotically incorrect error estimate for the RME (that is, the weighted LS error bar estimate, with the weights fixed at the final iteration). For the Parkfield data the fixed-weight jackknife yielded error bars that were almost always more conservative, but not necessarily more accurate overall. This scheme also failed to account for the actual day-to-day TF estimate variability in the bias band of 10–100 s.

The subset deletion jackknife performed significantly better in the bias band, especially when larger fractions of data were deleted. However, variances outside this band were again systematically too large, a result that is just as undesirable as underestimated error bars. Furthermore, the computational cost of the subset deletion approach is still considerable: for the 5 per cent subsets processing time increases by a factor of 20; for the 0.5 per cent subsets the increase is by a factor of 200. Typical run times on a workstation for one day of 1 Hz data approach a few hours in this case. These daily segments are in fact relatively short time-series compared to those routinely collected in modern wide-band MT surveys, so the issue of computational efficiency cannot really be ignored.

We remind the reader of the obvious: our study has been limited to a single pair of sites, and our conclusions must necessarily be rather tentative. Under different circumstances the performance of any or all of the error bar estimates might be expected to vary. For example, the clear advantage of the subset deletion jackknife appears to be limited to the band contaminated by coherent noise. Outside this band the standard asymptotic error bars were in fact more accurate. Extrapolation of these results suggests that if coherent noise and source effects were not a significant problem, there would be no advantage to using the subset deletion jackknife. On the other hand, at higher latitudes and longer periods, where source effects (in some sense similar to the complications seen in our data set for periods of 10–100 s) are most likely, the advantage of the subset deletion jackknife might be greater. Similarly, the fixed-weight jackknife might sometimes improve the accuracy of error bar estimates more significantly than observed here. Again, this might be particularly true at longer periods and higher latitudes, where source effects would be expected to lead to positive correlation between signal and noise amplitudes. Further tests under a variety of circumstances would be required to verify or contradict these suggestions.

One other issue that has so far received relatively little attention is the calculation of error bars for impedance elements after rotation. Rotation of the impedance tensor as part of an analysis of distortion and dimensionality is now standard practice in MT interpretation. Even if error bars for the impedance estimates are correct in the measurement coordinate system, they will not in general be correct after rotation unless the full covariance structure of the impedance estimates is accounted for. This is quite simple to do (see Appendix A), but is nonetheless seldom done in practice. Indeed, the standard EDI impedance format (Wight 1991) does not allow for storage of the full impedance error covariance. For the Parkfield MT site, ignoring covariances between errors of impedance components can lead to error bars in the rotated coordinate systems

that are off by 50 per cent or more. This is as large as the differences between error bar estimates obtained by the various methods considered here. This suggests that estimation and reporting of the full error covariance is probably as important an issue as the way in which error bars are computed.

All of the jackknife error estimates were more conservative, and one could argue that this is desirable to avoid possible overinterpretation of noisy data. However, it does not seem really necessary to perform all of the calculations required by the jackknife to obtain systematically conservative error bars. Multiplication of error bars by a factor of 1.5 or 2 would certainly accomplish the same thing. Error bars are ultimately used to assess the adequacy of the fit of estimated impedances to model predictions. At this stage the error bars must reflect inaccuracies implicit in instrument calibration and the modelling approach used, as well as statistical errors in the impedance estimates. It makes no sense to demand agreement with data to 1 per cent unless the modelling scheme can really achieve such accuracy (and instruments are calibrated this well). If 2-D modelling is used for the interpretation (as is commonly the case), agreement to within even a few per cent of the estimated impedances will seldom be justified. Even with 3-D modelling, limitations in grid resolution, numerical stability and accuracy probably limit the degree to which we should expect data to agree with a numerical model. In these sorts of calibrations, 'geological' and numerical noise are much harder to account for than the statistical precision of the impedance estimates. Common practice is thus to assume an error floor, often as large as 5–10 per cent, to allow for these uncertainties. Given the need to make these sorts of *ad hoc* adjustments to error bars, the use of a computationally intensive procedure to refine estimates of error bars seems difficult to justify for the routine processing of MT survey data. For our particular example the jackknife performed significantly better in the bias band. In fact, the actual variability of the MT TFs was only a few per cent in this band (see Fig. 7), so even with a small error floor the underestimation of error bars by the standard approach would most probably have no effect on actual inversion or interpretation of this data.

For specialized applications use of the jackknife might be more justifiable. For our example here, where we would like to assess the significance of subtle changes with time of MT impedances, we require highly accurate error bars. Unfortunately, the improvement in accuracy of the subset deletion jackknife is accompanied by systematic overestimation of error bars at other periods. In fact, in applications such as this, the actual day-to-day variability (which we can estimate directly) is probably the best way to estimate TF precision.

ACKNOWLEDGMENTS

H. Frank Morrison, Sierra Boyd and the Berkeley Seismological Laboratory provided data for this study. Comments and hints from Dr Alan Chave and a second, anonymous referee helped to give the manuscript its final shape. This work was partially supported by USGS NEHRP grant 1434-HQ-97-GR-03026 and NSF grant 9614411-EAR to GDE and by grants from Deutscher Akademischer Austauschdienst (DAAD, Gemeinsames Hochschulsonderprogramm III von Bund und Ländern) and Deutsche Forschungsgemeinschaft (DFG, Ei 370/4-1) to ME.

REFERENCES

Andersen, C.W., Lanzerotti, L.J. & Maclennan, C.G., 1976. Local time variation of induction vectors as indicators of internal and external current systems, *Geophys. Res. Lett.,* **3,** 495–498.

Bendat, J.S. & Piersol, A.G., 1971. *Random Data: Analysis and Measurment Procedures,* John Wiley and Sons, New York.

Boyd, O.S., Egbert, G.D., Eisel, M. & Morrison, H.F., 1997. A preliminary analysis of em field monitoring network, *EOS, Trans. Am. geophys. Un.,* **78,** 459.

Chave, A.D. & Smith, J.T., 1994. On electric and magnetic galvanic distortion tensor decompositions, *J. geophys. Res.,* **99(B3),** 4669–4682.

Chave, A.D. & Thomson, D.J., 1989. Some comments on magnetotelluric response function estimation, *J. geophys. Res.,* **94,** 4 215–14 226.

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

Dmitriev, V.I. & Berdichevsky, M.N., 1979. The fundamental model of magnetotelluric sounding, *Proc. IEEE,* **67,** 1033–1044.

Efron, B., 1982. *The Jackknife, the Bootstrap and Other Resampling,* Society for Industrial and Applied Mathematics, Philadelphia.

Efron, B. & Gong, G., 1983. A leisurely look at the bootstrap, the jackknife, and cross-validation, *Am. Stat.,* **37,** 36–48.

Efron, B. & Stein, C., 1981. The jackknife estimate of variance, *Am. Stat.,* **9,** 586–596.

Egbert, G. & Booker, J., 1989. Multivariate analysis of geomagnetic array data i: the response space, *J. geophys. Res.,* **94,** 14 227–14 248.

Egbert, G.D., 1997. Robust multiple-station magnetotelluric data processing, *Geophys. J. Int.,* **130,** 475–496.

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

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

Egbert, G.D., Eisel, M., Boyd, O.S. & Morrison, H.F., 2000. Pc3s: source effects in mid-latitude geomagnetic transfer functions, *Geophys. Res. Lett.,* **124,** 25–28.

Eisel, M., Haak, V., Pek, Y. & Čero, V., 2000. A magnetotelluric profile across the KTB surrounding: 2-D and 3-D modelling results, *J. geophys. Res.,* in press.

Fraser-Smith, A.C., Bernardi, A., McGill, P.R., Ladd, M.E., Helliwell, R.A. & Villard, O.G., *et al.*, 1990. Low-frequency magnetic field measurements near the epicenter of the m_s 7.1 Loma Prieta earthquake, *Geophys. Res. Lett.,* **19,** 1465–1468.

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

Gamble, T., Goubau, W. & Clarke, J., 1979b. Error analysis for remote reference magnetotellurics, *Geophysics,* **44,** 959–968.

Golub, G.H. & van Loan, C.F., 1989. *Matrix Computations,* John Hopkins University Press, Baltimore.

Graybill, F.A., 1976. *Theory and Application of the Linear Model,* Wadsworth Publishing Inc, Pacific Grove, CA.

Groom, R. & Bailey, R., 1989. Some effects of multiple inhomogeneities in magnetotellurics, *Geophys. Prospect.,* **37,** 697–712.

Groom, R.W. & Bahr, K., 1992. Corrections for near surface effects: decomposition of the magnetotelluric impedance tensor and scaling corrections for regional resistivities: a tutorial, *Surv. Geophys.,* **13,** 341–379.

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

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

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

Lanzerotti, L.J., Medford, L.V., Maclennan, C.G., Hasegawa, T., Acuna, M.H. & Dolce, S.R., 1981. Polarization characteristics of hydromagnetic waves at low geomagnetic latitudes, *J. geophys. Res.,* **86,** 5500–5506.

Larsen, J.C., 1980. Electromagnetic response functions from interrupted and noisy data, *J. Geomag. Geoelectr.,* **32,** SI89–SI103.
Larsen, J.C., Mackie, R.L., Manzella, A., Fiordelisi, A. & Rieven, S., 1996. Robust smooth magnetotelluric transfer functions, *Geophys. J. Int.,* **27,** 801–819.
Miller, R.G., 1974. The jackknife—a review, *Biometrika,* **61,** 1–15.
Park, S.K., Johnston, M.J.S., Madden, T.R., Morgan, F.D. & Morrison, H.F., 1993. Electromagnetic precursors to earthquakes in the ulf band; a review of observations and mechanisms, *Rev. Geophys.,* **31,** 117–132.
Quenouille, M.H., 1956. Notes on bias in estimation, *Biometrika,* **43,** 353–360.
Sims, W., Bostick, F. & Smith, H., 1971. The estimation of magntotelluric impedance tensor elements from measured data, *Geophysics,* **36,** 938–942.
Smith, J.T., 1995. Understanding telluric distortion matrices, *Geophys. J. Int.,* **122,** 219–226.
Thomson, D. & Chave, A., 1991. Jackknifed error estimates for spectra, coherences, and transfer function, in *Advances in Spectrum Analysis and Array Processing,* Haykin, S., Prentice Hall, Englewood Cliffs, NJ.
Thomson, D.J., 1977. Spectrum estimation techniques for characterization of the wt4 waveguide, i., *Bell Syst. Tech. J.,* **56,** 1769–1815.
Unsworth, M., Egbert, G.D. & Booker, J.R., 1999. High-resolution electromagnetic imaging of the San Andreas fault in Central California, *J. geophys. Res.,* **104(B1),** 1131–1150.
Waters, C., Menk, F. & Fraser, B., 1991. The resonance structure of low latitude pc3 geomagnetic pulsations, *Geophys. Res. Lett.,* **18,** 2293–2296.
Wight, D., 1991. Mt/emap data interchange standard, technical standards committee report, *Tech. Rept,* SEG, Tulsa, OK.

APPENDIX A: STORING MT IMPEDANCE COVARIANCES

The MT impedance tensor is a 2×2 matrix with tensor properties, namely it can be rotated into any coordinate system. More generally, the electric and magnetic fields can be translated into different coordinate systems. Both sorts of coordinate transformations are now routinely used to estimate the geoelectric strike, and to account for surface distortion effects (e.g. Groom & Bahr 1992; Smith 1995). Computation and storage of the full impedance tensor is thus essential for modern MT interpretation. To compute error bars correctly in a transformed coordinate system, the full covariances of the four elements of the impedance tensor are required. We consider here a general and efficient scheme for storing this full covariance information along with the impedance elements.

In general, the full error covariance for the elements of the impedance tensor $\mathbf{Z}$ is given by

$$\mathbf{Cov}[Z_{ij}Z_{i'j'}]=N_{ii'}S_{jj'}\,,\quad j,j'=1,2\,,\quad i,i'=1,2\,. \tag{A1}$$

Here $\mathbf{N}$ is the 2×2 covariance matrix of the residuals $E\eta\eta^{\dagger}$, and $\mathbf{S}$ is the inverse signal power matrix. For the single-station case, $\mathbf{S}=(\mathbf{H}^{\dagger}\mathbf{H})^{-1}$, while for the remote reference case $\mathbf{S}=(\mathbf{R}^{\dagger}\mathbf{H})^{-1}(\mathbf{R}^{\dagger}\mathbf{R})((\mathbf{R}^{\dagger}\mathbf{H})^{-1})^{\dagger}$. As Egbert & Booker (1989) and Egbert (1997) showed, the full error covariance for multiple station estimates based on a multivariate errors-in-variables model has a similar form, with all covariance elements expressed as a product of a noise covariance term (N_{ii}) and an inverse signal power term (S_{jj}). In all cases we thus need two 2×2 matrices to keep the full error covariance: $\mathbf{N}$, which provides estimates of the noise power, and $\mathbf{S}$, which provides estimates of signal power. Note that eq. (A1) easily generalizes to allow for additional predicted components (for example, including vertical magnetic TFs or allowing for multiple electric dipoles as in an electromagnetic profiling experiment). In this case the size of the matrix $\mathbf{N}$ is increased to the total number of output or predicted channels.

To estimate variances of any element of the impedance, only the diagonal elements of $\mathbf{N}$ and $\mathbf{S}$ are required. However, to change coordinate systems the off-diagonal elements of these matrices are required. In particular, if we transform the predictor ($\mathbf{H}$) channels by multiplying by the matrix $\mathbf{U}_H$ and the predicted ($\mathbf{E}$) channels by the matrix $\mathbf{U}_E$, the impedance is transformed to

$$\mathbf{Z}'=\mathbf{U}_E\mathbf{Z}\mathbf{U}_H^{-1}\,. \tag{A2}$$

The error covariance of the impedance estimate is still given by eq. (A1) with $\mathbf{N}$ and $\mathbf{S}$ transformed to

$$\mathbf{N}'=\mathbf{U}_E\mathbf{N}\mathbf{U}_E^{\dagger}\,,\quad \mathbf{S}'=(\mathbf{U}_H^{-1})^{\dagger}\mathbf{S}\mathbf{U}_H^{-1}\,. \tag{A3}$$

In the usual case $\mathbf{U}_E$ and $\mathbf{U}_H$ are rotation matrices, so that $\mathbf{U}_H^{\dagger}=\mathbf{U}_H^{-1}$ (and so $(\mathbf{U}_H^{-1})^{\dagger}=\mathbf{U}$). Also, for simple coordinate rotations both $\mathbf{H}$ and $\mathbf{E}$ are rotated by the same angle so $\mathbf{U}_H=\mathbf{U}_E$.

To allow a correct transformation of error bars we thus compute and store the full residual covariance $\mathbf{N}$ along with the usual inverse signal cross-product matrix $\mathbf{S}$. Correction factors required for the asymptotic RME (such as estimates of $E\psi'$) are incorporated into the saved matrices in the natural way.

A major advantage of this storage scheme is that output file formats are identical for single-station, remote reference and multiple-station processing. This makes the merging of results from different acquisition bands and sites into a unified database of common format easy, and simplifies subsequent processing steps such as rotation of coordinate systems. This has proved useful for rapid initial in-field processing, where initial results are often computed at the site with single-station processing. Remote reference processing can then be used for the most severely biased bands to produce initial relatively bias-free estimates quickly.

A more standard way to store processing results that also allows for correct rotation of coordinate systems is to keep the full cross-product matrix (SDM) for each period. However, this format is not particularly well suited to robust schemes. Furthermore, the output file format and the steps that must be taken to compute the actual impedance estimates depend on whether the data is remote reference or single site. This complicates subsequent usage of the processed data.

APPENDIX B: ASYMPTOTIC COVARIANCES FOR REDESCENDING INFLUENCE FUNCTIONS

In this appendix we consider computation of the asymptotic covariance matrix for the regression M-estimate using eqs (12) and (13) (or for remote reference eq. 16) in the case where the loss function ρ is non-convex. This corresponds to the case of the so-called redescending influence function, which we use for the final iterations of the robust estimator. Following Thomson (1977) and Egbert & Booker (1986) we use a double exponential form for the influence function,

$$\psi(\eta)=\eta\exp[-\exp(\eta_1(\eta-\eta_1))]\,, \tag{B1}$$

with $\eta_1=2.5$–3. To a good approximation, $\psi(\eta)=\eta$ for $\eta<\eta_1$, $\psi(\eta)=0$ for $\eta>\eta_1$, and the weights used in the final iteration of

the robust estimate are essentially either 1 or 0. However, ψ is a continuous function (as it must be for eq. 12 to hold), so $\psi(\eta)$ has to go down (redescend) in the neighbourhood of η_1. Thus, ψ', which appears in the asymptotic formula (12) is nearly 1 for small residuals, nearly 0 for large residuals, and negative in between. Although the transition zone is narrow, it still effects the expectation $E(\psi'(\eta))$, which is most obviously approximated by

$$I^{-1} \sum \psi'(\eta_i/\hat{\sigma})^2 . \tag{B2}$$

By making the interval over which ψ' is negative very small (as the double exponential form does), we might expect that residuals will seldom actually land in this narrow interval, so that ψ' will be 0 or 1, and we will find $E(\psi') \approx 1-q$, where q is the fraction of data downweighted in the final iteration. However, by making the interval small, we have to make the negative slope of ψ (i.e. ψ') large. Usually no residuals will land in the narrow zone of negative slope, but when one or more does, the sample approximation (B2) will be effected severely, possibly even coming out negative. As a result, the estimate of $E\psi'$ given by eq. (B2) will be unstable and have a large variance. We have thus implemented an approximate, but more stable, estimate of $E(\psi')$.

Letting $p(\eta)$ be the probability density function of the absolute value of the (normalized) residuals, and assuming that ψ' is approximately 1 or 0 outside a narrow zone of width 2η centred on η_1, we have

$$E(\psi') = \int_0^\infty p(\eta)\psi'(\eta)d\eta \approx \int_0^{\eta_1-\varepsilon} p(\eta)d\eta + \int_{\eta_1-\varepsilon}^{\eta_1+\varepsilon} p(\eta)\psi'(\eta)d\eta$$

$$= 1-q-p(\eta_1) . \tag{B3}$$

Here we have assumed that $p(\eta)$ is nearly constant across the transition interval of width 2η, and then used the fact that

$$\int_{\eta_1-\varepsilon}^{\eta_1+\varepsilon} \psi'(\eta)d\eta \approx -1 . \tag{B4}$$

Allowing for the negative slope of the redescending influence curve thus approximately increases the error variances by a factor of $[(1-q)/(1-q-p(\eta_1))]^2$, compared to doing the calculations with discontinuous (exactly 0/1) weights.

Of course, it is difficult to estimate $p(\eta_1)$ stably if few residuals fall into the narrow transition zone. We thus make the further assumption that the tails of the distribution fall off exponentially. Under this assumption it is simple to show that $p(\eta_1) = q$. This leads to the readily computed approximation $E(\psi') \approx 1-2q$, where q is the fraction of data given (essentially) zero weight on the final iteration. If the tails of the residual distribution fall off more slowly than exponential, then q would slightly overestimate $p(\eta)$, so this approximation could be a bit conservative. Using this approximation on the monitoring data considered here typically increased error variances by about 10–15 per cent relative to the simpler approximation $E\psi' \approx 1-q$.