DOE/ER/14057-1
Distribution Category UC-60

DOE/ER/14057--1

DE90 016188

NEW APPROACHES TO ESTIMATION OF MAGNETOTELLURIC PARAMETERS

ANNUAL REPORT
FOR THE PERIOD
1 AUGUST, 1989 - 31 July, 1990

DR. GARY D. EGBERT

College of Oceanography
Oregon State University
Oceanography Admin Bldg 104
Corvallis, OR 97331-5503

DR. JOHN R. BOOKER

Geophysics Program, AK-50
University of Washington
Seattle, WA 98195

PREPARED FOR THE
U.S. DEPARTMENT OF ENERGY
OFFICE OF BASIC ENERGY SCIENCES
UNDER GRANT DE-FG06-89ER14057

Technical Progress Report for DOE grant DE-FG06-89ER14057
New Approaches to Magnetotelluric Parameter Estimation

Introduction

We proposed the development and application of some new statistical techniques for improving the collection and analysis of wide-band magnetotelluric (MT) data. The principle goal of our work is to develop and implement fully automatic single station and remote reference impedance estimation schemes which are robust, unbiased and statistically efficient. In the initial proposal, we suggested several extensions to the regression M-estimates to better allow for non-stationary and non-Gaussian noise in both electric and magnetic field channels (measured at one or more simultaneous stations). A second goal of the proposal was to develop formal, reliable procedures for estimating undistorted 2-d strike directions and to develop statistics for assessing the validity of the 2-d assumption that are unaffected by near surface static distortion effects. To test and validate our methods we proposed working with data selected from a series of over 200 wide-band MT sites collected by Dr. Harvey Waff at the University of Oregon. The grant included a subcontract to Dr. John Booker at the University of Washington. Booker is funded by a separate DOE grant for research on rapid multi-dimensional inversion of MT data, and the primary purpose of the subcontract is to enhance contact between developments in initial processing, and subsequent inversion of this data. For the current budget period, we proposed setting up a data base, and completing development and initial testing of the single station and remote reference methods outlined in the proposal.

Summary of Progress

In this budget period (the first for this grant) the funding received essentially covers 4 months of 100% effort by the PI. Because of the timing of this report, only 2 months of effort have been spent on this project so far, and only a portion of the work proposed for this budget period has been completed. As such the results reported here should be considered preliminary.

The principle accomplishments to date include the following:

(1) Two large regional data sets have been obtained from the University of Oregon MT group. Existing processing and plotting programs have been modified to allow for differences in data format, frequency ranges, filter corrections etc.

(2) Several simple, but necessary, enhancements to the robust code have been made.

(3) We have begun testing the robust MT code on wide-band data. Although the robust methods yield improvements in estimates (which are spectacular, at times), we have found that there are some special problems in processing wide band MT which require novel solutions. The principle difficulty is that noise, and particularly outliers, in the horizontal magnetic fields can be much more of a problem with wide band MT (particularly in the "dead band" of 0.1–1.0 *hz*) than it was for the long period MT data we have previously worked with. We have begun developing and testing methods for dealing with this problem.

(4) We have developed methods for converting frequency domain transfer functions into a time domain impulse response. Our goal is to use the impulse response as a prediction filter for identifying badly contaminated data in the time domain.

(5) Booker has made substantial progress on adapting inversion programs to handle real data. A major finding is that the smooth 2-d inversion code often fails to produce reasonable results when the MT impedances are contaminated by a small number of outliers. Booker is developing robust features in the inversion itself to improve stability.

Further details are given in the following sections.

Data

We have obtained wide-band MT data from Dr. Harvey Waff and his group at the University of Oregon. Initially we are concentrating on two large regional data sets. The first, consisting of 20 sites, was collected by the UO group in Northern California, near the Mendocino triple junction. This data was collected by UO in collaboration with one of us (Booker) for an NSF funded research project. A second data set that we have obtained consists of nearly 50 wide-band MT sites collected in 4 transects across the Oregon High Cascades. This data was collected by the UO group on contract for Cal Energy Inc. We have ordered and installed the hard disk drive requested in the proposal, and a data base has been organized.

Some initial effort has been required to adapt our existing plotting and processing programs to accommodate peculiarities of the wide-band data available. For instance: To efficiently handle the many relatively short data segments (with possible gain changes) which are stored in separate files, we have adapted initial time domain cleaning programs to merge data into a common file with a fixed gain. Changes in plotting programs have been made to accommodate the very different spectral characteristics of the wide-band data. Filter correction features in processing programs had to be adapted to include the UO system response. All of these changes represent relatively minor, but necessary, initial steps in our work. We note that we allotted 1 month in the proposed work plan for this; this schedule has been met.

Enhancements to Robust Processing Code

We have also completed several necessary enhancements to the processing code described in Egbert and Booker (1986) and Jones *et al* (1989). First, we have added code to allow for adaptive autoregressive pre-whitening of the time series before Fourier analysis. This is a relatively standard technique, and certainly does not represent a fundamental contribution to the state-of-the-art in MT processing on our part. However, we felt that this enhancement was necessary for wide-band processing. For long period EM data (i.e., periods greater than 10 s.) the spectrum is very red, and we have found that, while pre-whitening is necessary, a simple first difference is completely adequate. For wide-band data the spectrum is more complicated, and more general pre-whitening methods are required.

We have also added code to allow for a simple robust remote reference estimate of the sort discussed by Chave and Thomson (1989) and Jones *et al.* (1989). It is our intent to develop a substantially refined approach to robust remote reference (e.g., Chave and Thomson do not explicitly consider the possibility of outliers in the local or remote reference channels). We have added this code to give us a baseline for comparison when we consider the more complicated robust remote reference approaches suggested in the proposal.

A potential difficulty with even the robust impedance estimates (discussed in Chave and Thomson (1989); see also Egbert and Booker, (1986)) is that the presence of a small number of very large amplitude sections can result in the complete breakdown of estimates. In the terminology of linear statistical models, Fourier coefficients with unusually large magnetic (or more generally, reference) channel amplitudes are *leverage points*, i.e., individual data points which have an inordinate influence (leverage) on the estimate. Even if the large amplitude is real signal, this can result in failure of the robust estimates (Huber, 1981). The situation is even worse if the large amplitude sections result from contamination of the magnetic channel and do not correspond to signal. For long period magnetovariational data we controlled this problem (to a great extent at least) by using a power dependent weighting scheme (Egbert and Booker, 1986). We could do this because we knew the spectral characteristics of our signal and noise very well, and could thus use a fixed set of empirically determined weights. For wide-band MT the situation is not so simple, since both signal and noise amplitudes are highly variable in space and time.

To deal with this problem in an adaptive fashion which is more appropriate to wide band MT processing, we have implemented code which automatically downweights leverage points. This is accomplished as follows:

(1) A robust estimate of the horizontal magnetic field spectral density matrix (SDM) $\hat{\mathbf{H}}$ is computed by an iterative process. To do this we start with the horizontal field SDM computed from the averaged cross products of all data:

$$\hat{\mathbf{H}}_0 = N^{-1} \sum_{i=1}^{N} \mathbf{h}_i \mathbf{h}_i^{\dagger} \qquad (1)$$

Then the effective influence of a single point on the estimate can be computed as

$$e_i = \mathbf{h}_i^{\dagger} \hat{\mathbf{H}}_0^{-1} \mathbf{h}_i \tag{2}$$

If all horizontal field segments are of similar amplitude, $e_i \approx 1$ for each i. If e_i is much larger than 1, then the ith data segment has a relatively large influence on the estimate. In fact it is readily seen that $\sum_{i=1}^{N} e_i = N$, and that e_i/N gives an estimate of the fraction of the total influence of the ith data segment on the estimate. We set a limit on e_i (e.g. 10), omit every point which exceeds this limit, and recompute $\hat{\mathbf{H}}$. We then recompute the influence values e_i with the new estimate $\hat{\mathbf{H}}_1$. This process is iterated until all e_i are less than the specified upper limit. This effectively gives an estimate of the scale of the horizontal magnetic fields which excludes all the unusually large values.

(2) Using the estimate of $\hat{\mathbf{H}}$ from the final iteration of step (1), we compute e_i (defined in equation (2)) for all data points. Before further processing data is multiplied by a weight of the form

$$w(e_i) = \begin{cases} 1 & \text{if} \quad e_i < p_1 \\ (p_1/e_i)^{1/2} & \text{if} \quad p_1 < e_i < p_2 \\ 0 & \text{if} \quad p_2 < e_i \end{cases} \tag{3}$$

where p_1 and p_2 are parameters which depend on the sample sizes. For the results reported here, we have used

$$p_1 = 2N^{1/2} \qquad\qquad p_2 = 20N^{1/2} .$$

This scheme has two effects. First, data sections of extremely high power are assumed to be garbage, and are thrown out. For instance if $N = 100$, data segments with power 200 times larger than the typical power are omitted completely. Second, sections of moderately high power are downweighted, so that the maximum influence any single point can have on the results is bounded. With $N = 100$ this bound is 20 times the influence of a typical data point. Note that the scale used to define a "typical" data point is determined by omitting all large amplitude sections. We will refer to the method described here as a generalized or weighted M-estimate. It is very similar to a general class of leverage resistant M-estimators described by Huber (1981).

As an initial example where this approach makes a big difference, we consider apparent resistivity estimates from one of the EMSLAB long period MT stations (figure 1). The horizontal magnetic fields were contaminated by a single section of very large amplitude noise (readily visible in plots of the time series). Standard LS estimates (dashed line, solid dots) are nearly worthless. The robust results (dotted line, asterisks) are much better, but are still very erratic around a period of 200 seconds. Results obtained using the method described above (solid line, triangles) are much improved. When leverage points are adaptively downweighted, the estimates are completely insensitive to the presence or absence of the contaminated data. Note also that at shorter periods the estimate which omits leverage points is also systematically above the standard robust estimate. This reflects the bias effect of the noise in the horizontal components, which the robust scheme cannot completely eliminate.

Some Initial Results for Wide-band MT

As a first step in implementing the new approaches discussed in the proposal, we have begun testing the enhanced robust code on wide-band MT data sets. A principle finding is that noise (and significant outliers) in the magnetic fields are often a major problem in the wide-band data. The reason for this lies in the very different spectral characteristics of the signal and noise processes. The signal spectrum has a minimum near 0.1–1.0 *hz*. In contrast, cultural noise can be very large at these frequencies. Thus the signal can be overwhelmed by noise in the magnetic fields. A particularly bad example of this difficulty can be seen in the section of time series plotted in figure (2). This sort of noise can create problems, even for the robust processing scheme, since the data that is weighted most heavily in the initial estimate is the most contaminated.

We illustrate the effect of the magnetic field noise seen in figure (2) on impedance estimates in figures (3)-(7). The standard least squares (LS) estimates (figure 3) are essentially useless over most of the frequency range. The robust estimates (figure 4) produce more reasonable results, but they are still erratic and biased to low amplitudes between 0.3–0.05 *hz*. In figure (5) we plot estimates obtained after omitting the leverage points with the approach outlined above. While substantial biases are still evident near 0.1 *hz*, the estimates are noticeably improved. In the remainder of our discussion, "robust estimates" will refer to generalized regression M-estimates, with adaptive downweighting of leverage points.

In figures (6) and (7) we plot LS and robust results computed from time series from which obviously contaminated sections (determined by visual examination) have been omitted. The LS estimates are still very poor. In contrast, the robust estimator now performs reasonably well, although there still are some problems evident near 0.1 *hz*, where the error bars are greatly increased and the apparent resistivity curve is biased downwards. It is probable that this bias results from a persistent poor signal to noise ratio in the magnetic fields. This problem should be corrected by application of the remote reference method, although there is some evidence to suggest that simple robust remote reference estimates do not work well when the intrinsic signal to noise ratio is low (Jones *et al.*, 1989). So far, however, we have considered only single station estimates, with the goals of minimizing the bias problem and automating the visual screening.

We have tried a simple approach: a preliminary coherency presort, to eliminate low coherence data segments before robust stacking (figure 8). By itself, sorting by coherence is not a new idea (e.g., Stodt, 1983; Jones and Jodicke, 1984). What is novel here is to use this as a preliminary sort, followed by robust processing of the high coherence segments. Results of preliminary test of this are given in figures (9) and (10). When applied to the raw unscreened data (figure 9) the method dramatically improves the preformance of the estimator in the dead band near a period of 10 sec. When applied to data which has been visually pre-screened, the results are even better (figure 10). Note that by itself coherency sorting (as we have implemented it here) does not produce estimates comparable to the robust scheme (figure 11).

We conclude that coherency pre-sorting combined with the regression M-estimate preforms considerably better than either one alone, at least in the dead band. Together, the two approaches are capable of producing very good results from very noisy data, even when no effort is made to remove the obviously contaminated data. However, an initial visual screening of the data to remove the obviously contaminated sections further improves results. While we would always advocate looking at the data and omitting any obvious junk, it is desirable to automate the preliminary screening as much as possible. We are thus now trying to develop methods for automated screening of data in the time domain. First steps in this direction are discussed in the next section.

Time domain Impulse Response

The data plotted in figure (2) suggest that outliers from some sorts of cultural noise may be fairly well localized in time. In an effort to see if we can automatically detect and remove these bad data sections we have begun considering some time domain methods. With the approach envisioned we would predict the electric fields from the magnetic fields (or vice-versa) as a convolution in the time domain. As a first step we have begun experimenting with methods for converting the frequency domain impedance estimates into an impulse response which relates the electric and magnetic fields in the time domain:

$$\mathbf{e}_t = \sum_{t'=-\infty}^{\infty} \mathbf{I}_{t'} \mathbf{h}_{t-t'} \quad . \tag{4}$$

The impedance estimates $\hat{\mathbf{Z}}$ at frequency ω are related to the impulse response tensor $\mathbf{I}_t$ via the discrete Fourier transform

$$\hat{\mathbf{Z}}(\omega) = \sum_{t=-\infty}^{\infty} \mathbf{I}_t e^{2\pi i \omega t} + \varepsilon \, . \tag{5}$$

We thus seek a discretely sampled (and hence band limited) impulse response which agrees, over the frequency band of interest, with the estimated frequency domain impedance tensors. Note that this discrete impulse response need not be causal, even in cases (e.g. 1-d conductivity; Weidelt, 1972) where the continuous time impulse response (which is not band limited) is causal (Egbert, ms in preparation). To compute $\hat{\mathbf{I}}_t$, $t = -\infty, \infty$ from the observed impedances we use a regularized inverse approach, and seek $\hat{\mathbf{I}}_t$ which is consistent with the estimated transfer functions (within the errors) and minimizes a norm of the form:

$$\|\hat{\mathbf{I}}\|^2 = \sum_{t=-\infty}^{\infty} t^{2n} \|\hat{\mathbf{I}}_t\|^2 \tag{6}$$

For $n = 2$ this is equivalent to finding the impulse response $\hat{\mathbf{I}}_t$ whose Fourier transform minimizes

$$\int_0^{\omega_{Nyq}} d\omega \, r \| \frac{d^2 \mathbf{Z}}{d\omega^2} r \|^2$$

subject to fitting the observed impedances adequately. $\hat{\mathbf{I}}$ computed in this way is thus the impulse response corresponding to a spline smoother (in the frequency domain) of the observed impedances. In practice, it is necessary to truncate the impulse response to a finite time interval. In fact, the smoothness criteria (6) tends to localize the impulse response in the vicinity of the origin, so a relatively small number of non-zero coefficients are required to adequately approximate the estimated impedance in the time domain. This is illustrated in figure (12) where the impulse response for one one impedance component (for a very long period MT data set) is plotted.

Again the results reported here are extremely preliminary. A more detailed discussion of the use of time domain methods must await further development and testing.

Inversion of Real Data

Using a very fast iterative algorithm developed under separate DOE funding, Booker has been inverting large MT data sets. In the first year of this work, he has looked at a 35 site profile from Canada, A 48 site profile in a mining district in Nevada and is presently working on a 100 site profile across the Wind River Overthrust in Wyoming. To stabilize these inversions, he has applied Huber robust down-weighting to outlying MT data. This has worked very well for noisy data, but has lead to difficulties for more accurate data. It appears clear why this is: If the starting model is a long way from the correct model, the Huber down-weighting may treat some of the important data points as outliers. Booker has therefore developed an alternative robustification which works in the model space rather than the data space. If the inversion tries to put in rough structure, which may be associated with noise, then the algorithm partially roles back the step size in the direction of this roughness. This scheme works very well for early iterations for noisy or accurate data and continues to work well for all iterations for low-noise data. An example application of the inversion is given in figure (13). The next modification will be a combination of the two in which the Huber down-weighting in the data domain remains weak until the large scale model structure has been built up.

Future Work

In broad outline our work plan remains as originally proposed, although our initial encounters with real data, and our perception of current trends in induction research, suggest some minor reorderings of priorities. Here is a brief summary of our current research plan.

(1) We believe that most MT data will be collected with a remote reference in the future. As a consequence, we intend to concentrate our effort, both in the remainder of the current budget period, and in the next budget period, on robust remote reference estimation methods. We will begin work in this area soon, and we expect to make substantial progress on this problem before the end of the current budget period. Our plans in this regard remain much as they were when the original proposal was written: we will focus on adapting the multiple station methods of Egbert and Booker (1989) to the remote reference problem.

(2) In the next budget period we proposed developing statistical methods for treating the distortion problem. This general problem has received a lot of attention recently, and some of the ideas for parameter estimation suggested in the proposal have already been implemented by other researchers. Rather than duplicate these efforts, we will concentrate more on the development and testing of rigorous statistical methods for assessing the validity of the 2-d assumption.

(3) One new idea has been added to those discussed in the proposal: the use of time domain prediction filters to look for sections of bad data. This idea, which is in the initial stages of development, is discussed more fully in the technical progress report.

(4) Booker will continue work on 2-d inversion of real data. Further enhancements of the robust features in the inversion are planned. In addition, the two large data sets that will be processed in their entirety by Egbert will be inverted by Booker. Tests to see how much difference the robust methods make in the final interpretation will be conducted.

(5) In the time since the proposal was written researchers in the academic MT community have become increasingly aware of the potential value of robust methods. A large number of these researchers have expressed interest in obtaining copies of our processing code. We have released versions of our programs, but not without some trepidation. Our code, while reasonably portable, has been developed as an experimental tool for testing new processing methods. As such, it is not well adapted to routine use. The code is much more complicated than necessary, and many of its features would never be used for processing wide-band data. We have decided that it would be useful to put some of our effort in the second year into streamlining our processing code, improving its user interface, and cleaning up documentation. It is our intent that the final report for this project will include a substantially streamlined copy of the source code for our programs, together with a users manual.

(6) In addition to the final technical report, we expect to submit an article to Geophysics, comparing the various methods that we are developing with more conventional methods used in processing of wide band MT data. We will also present our results at the 1991 SEG meeting. More immediately, we have submitted an abstract for a talk summarizing preliminary results which will be given this summer at the 10th induction workshop in Ensenada, Mexico.

Conclusion

We are in the early stages of a research project whose goal is to develop improved procedures for the initial processing and reduction of wide-band MT data. We have obtained data and modified programs, and testing of robust methods and extensions has begun. Initial results are very encouraging - the robust approach seems to work substantially better than simple LS estimates. However, the low signal to noise ratio in the dead band centered around .1 hz creates some special difficulties which require some modification and extension of the standard robust methods which have been primarily developed and tested on long period data. We have started work on several approaches to this problem.

References

Chave, A.D., and D.J. Thomson, Some comments on magnetotelluric response function estimation, *J. Geophys. Res., 94,* 14,219-14226, 1989.

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

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

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

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

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

Stodt, J.A. Noise analysis for conventional and remote reference magnetotellurics, Ph.D. thesis, Univ. of Utah, 220 pp., 1983.

Wiedelt, The inverse problem of geomagnetic induction. *Z. Geophys., 38,* 257-289, 1972.

fig. 1

Figure 1: Apparent resistivities obtained from three estimators from long period MT data taken during the EMSLAB experiment. Dashed line and solid dots: standard least squares. Dotted line and asterisks: regression M-estimate. Solid line and triangles: generalized regression M-estimate, adapted to down-weight leverage points. Severe contamination of horizontal fields in several short sections leads to the breakdown of the standard regression M-estimate. The generalized M-estimate performs well even in the presence of severe outliers in the reference magnetic field channels.

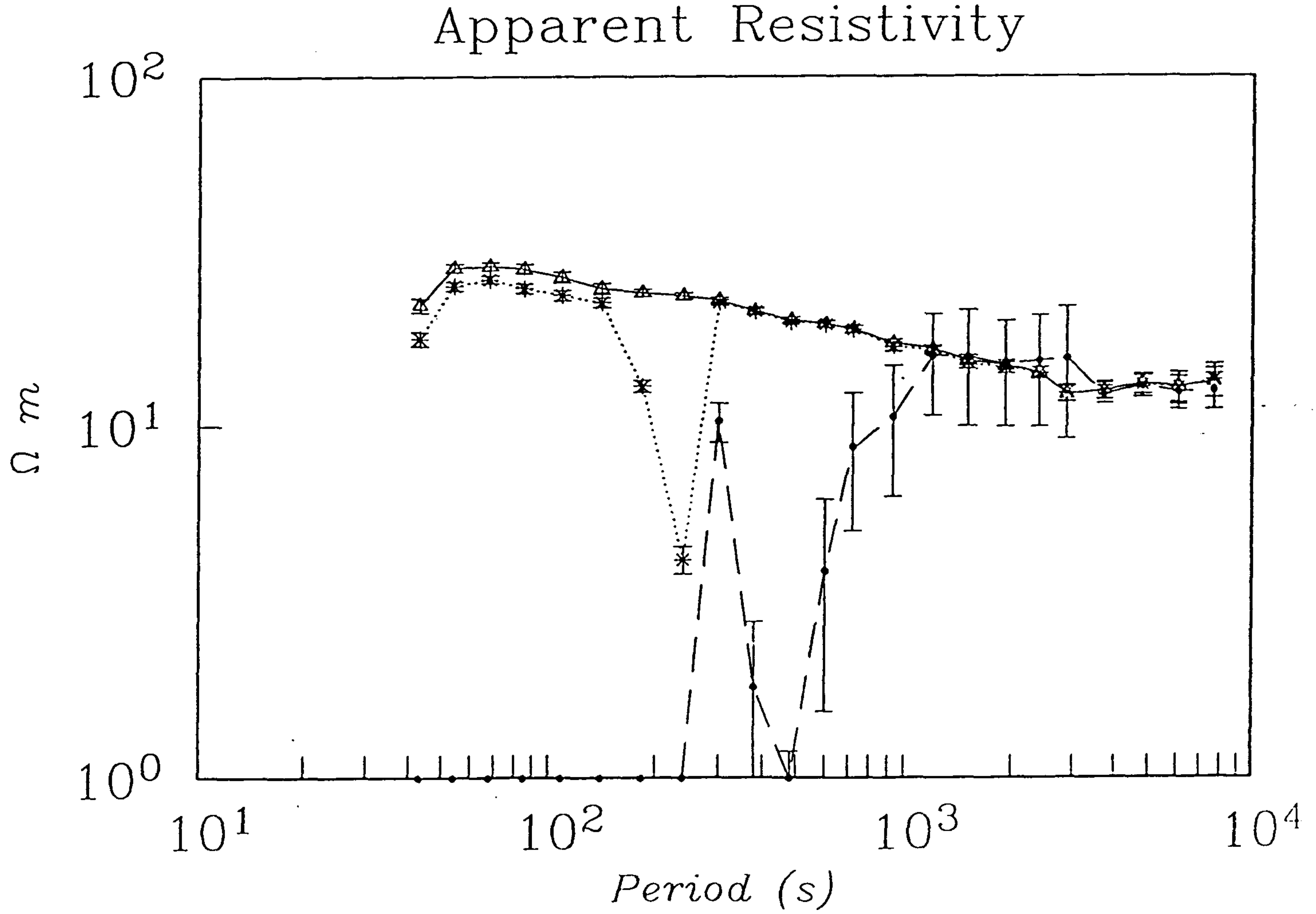

fig. 2

Figure 2: Example wide band MT time series, showing contamination of horizontal magnetic fields for wide band site mslb75.

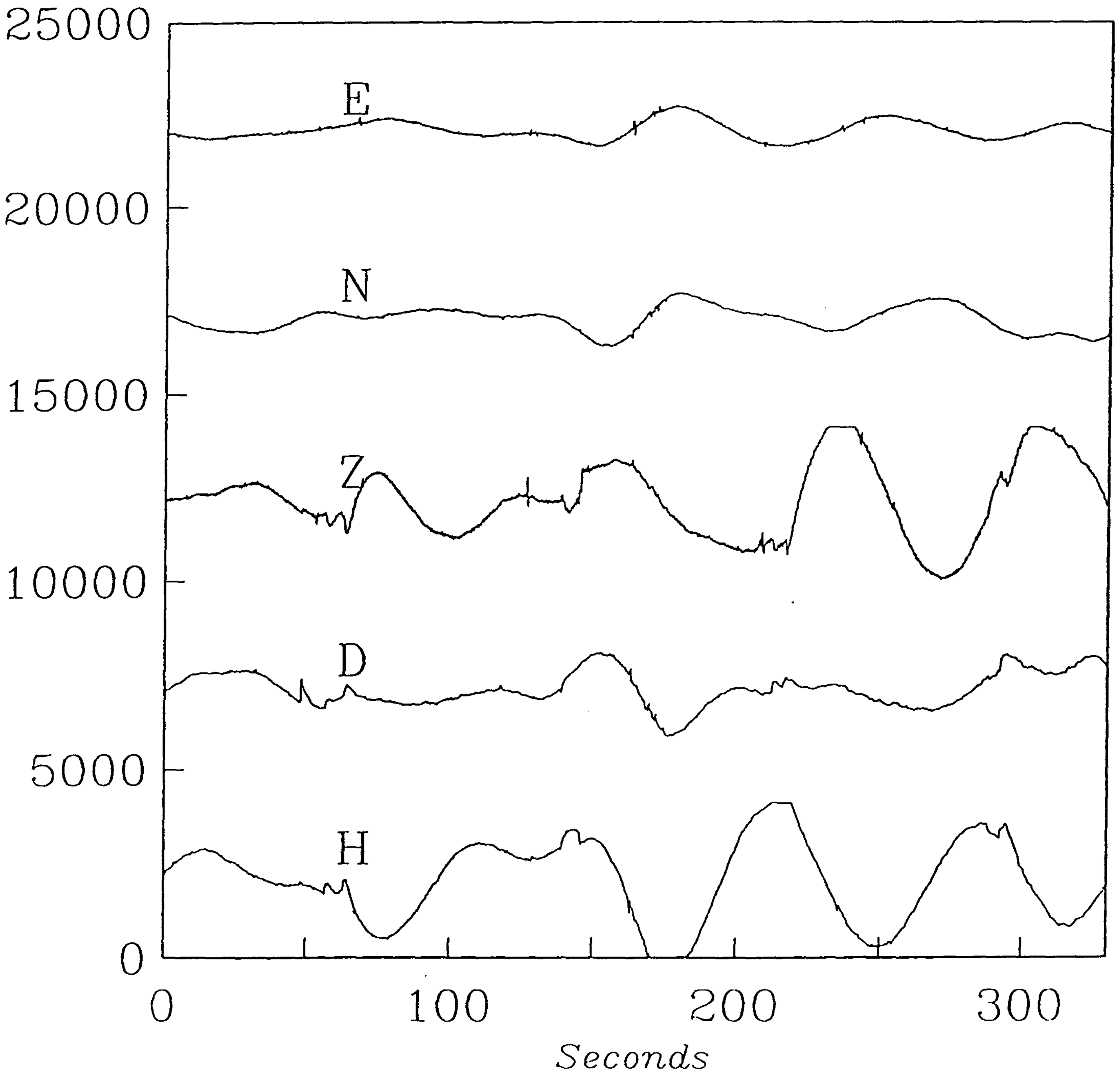

fig. 3

Figure 3: Least squares (LS) estimates of apparent resistivities and phases for mid-band at mslb75.

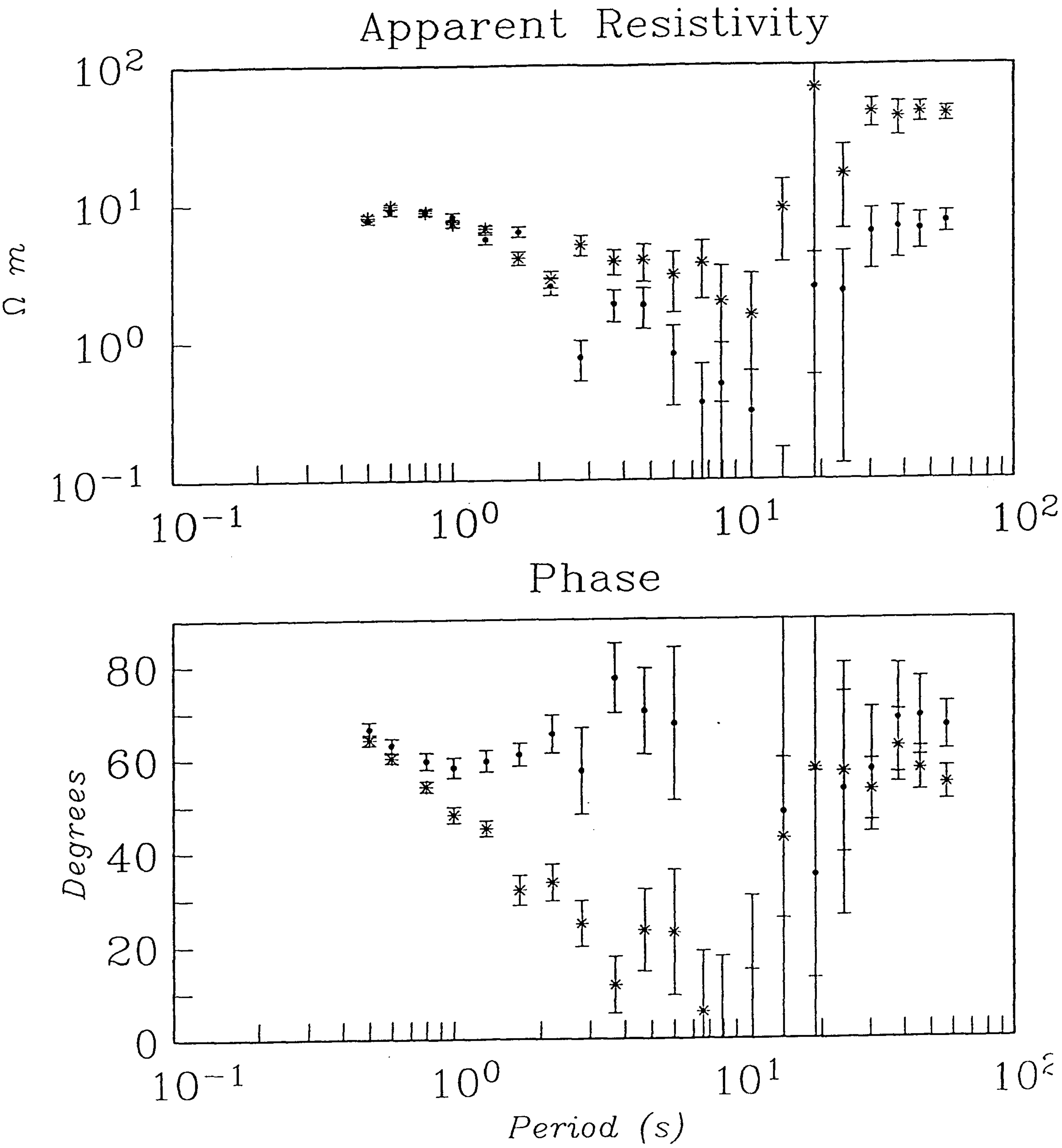

fig. 4

Figure 4: As in figure 3, but robust regression M-estimates. Although results are substantially improved, the estimates are severely biased near 10 second periods. Note also the significant difference between robust and LS estimates at shorter periods.

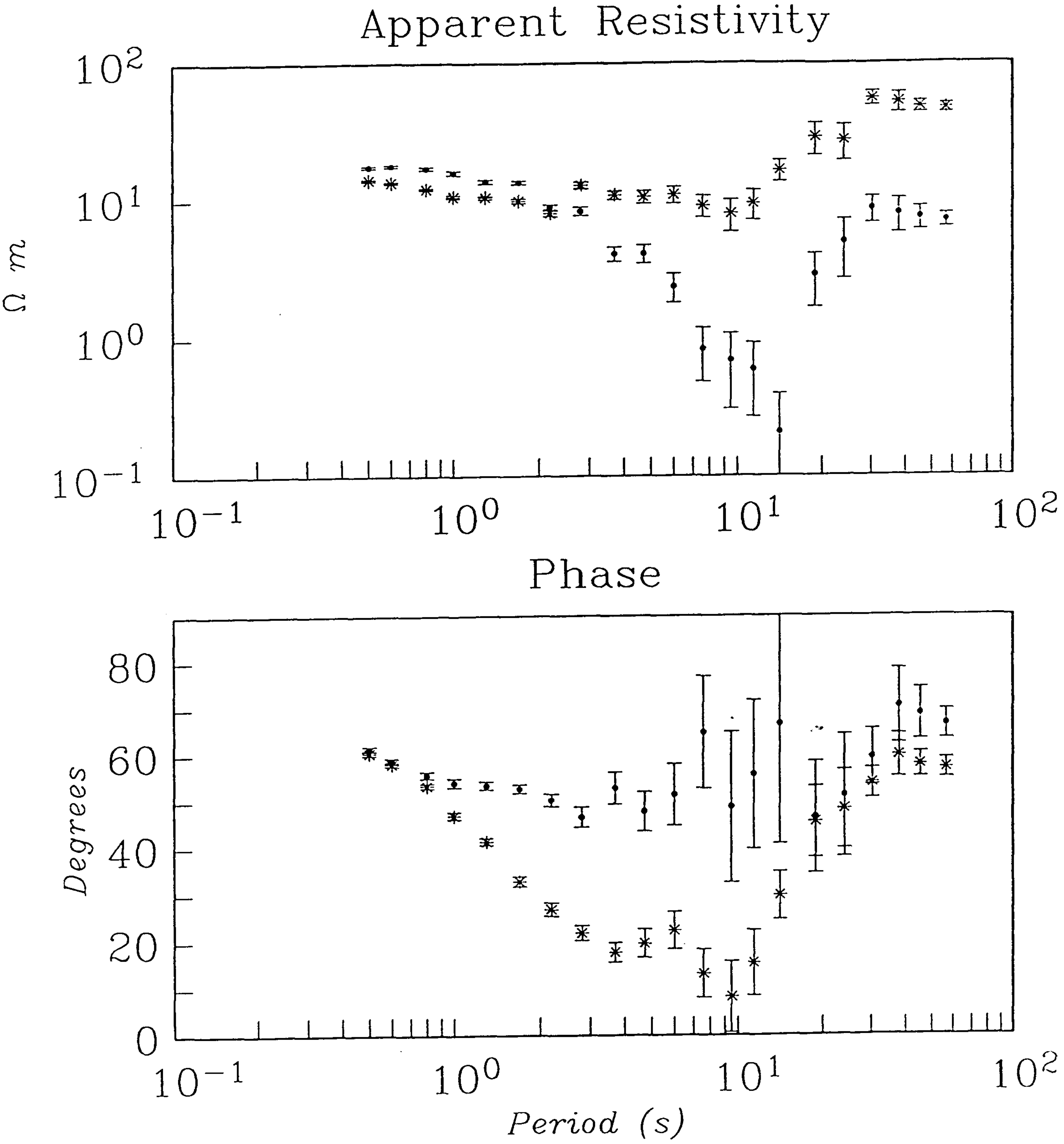

fig. 6

Figure 5: As in figure 4, but estimates were computed with the generalized M-estimate which automatically downweights leverage points. Although the results are improved further, there are still substantial problems in the dead band.

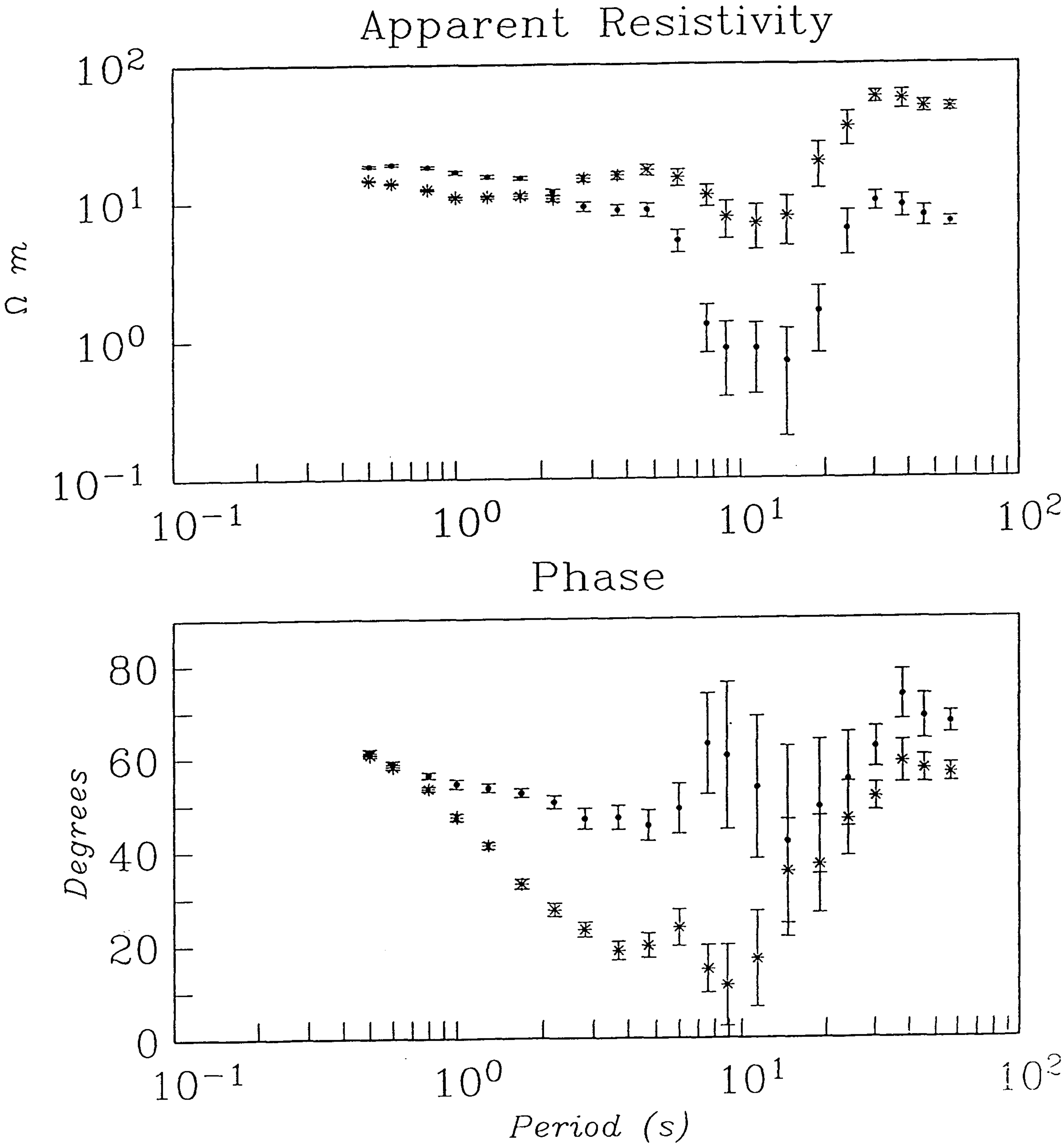

fig. 6

Figure 6: As in Figure 3, but the time series was visually screened to remove the most obviously contaminated sections (such as those visible in figure 2). The least squares estimates are still very poor.

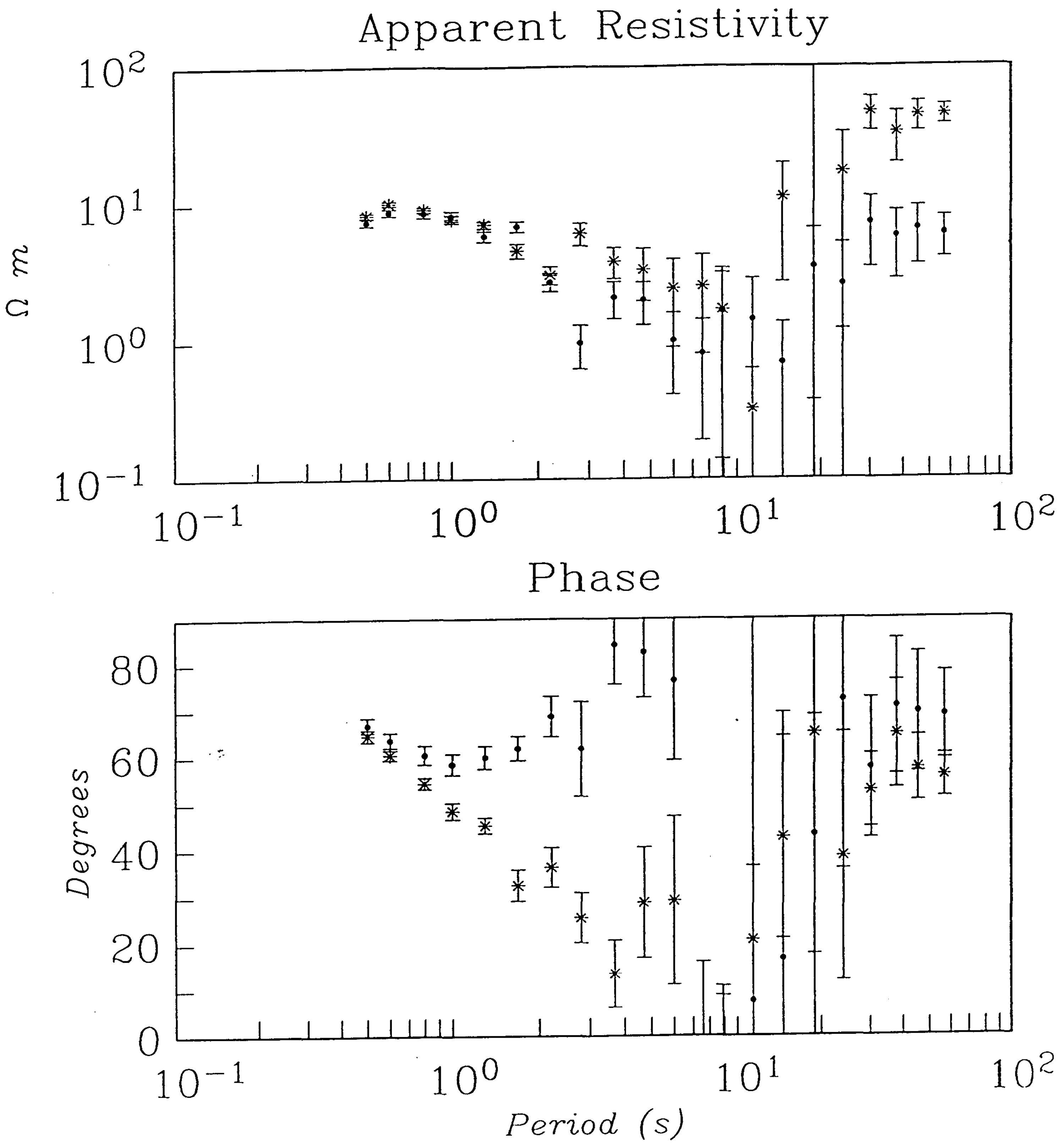

fig. 7

Figure 7: Visually screened as in figure 6, but estimates computed with generalized M-estimate. Estimates are smooth and the bias in the dead band is significantly reduced.

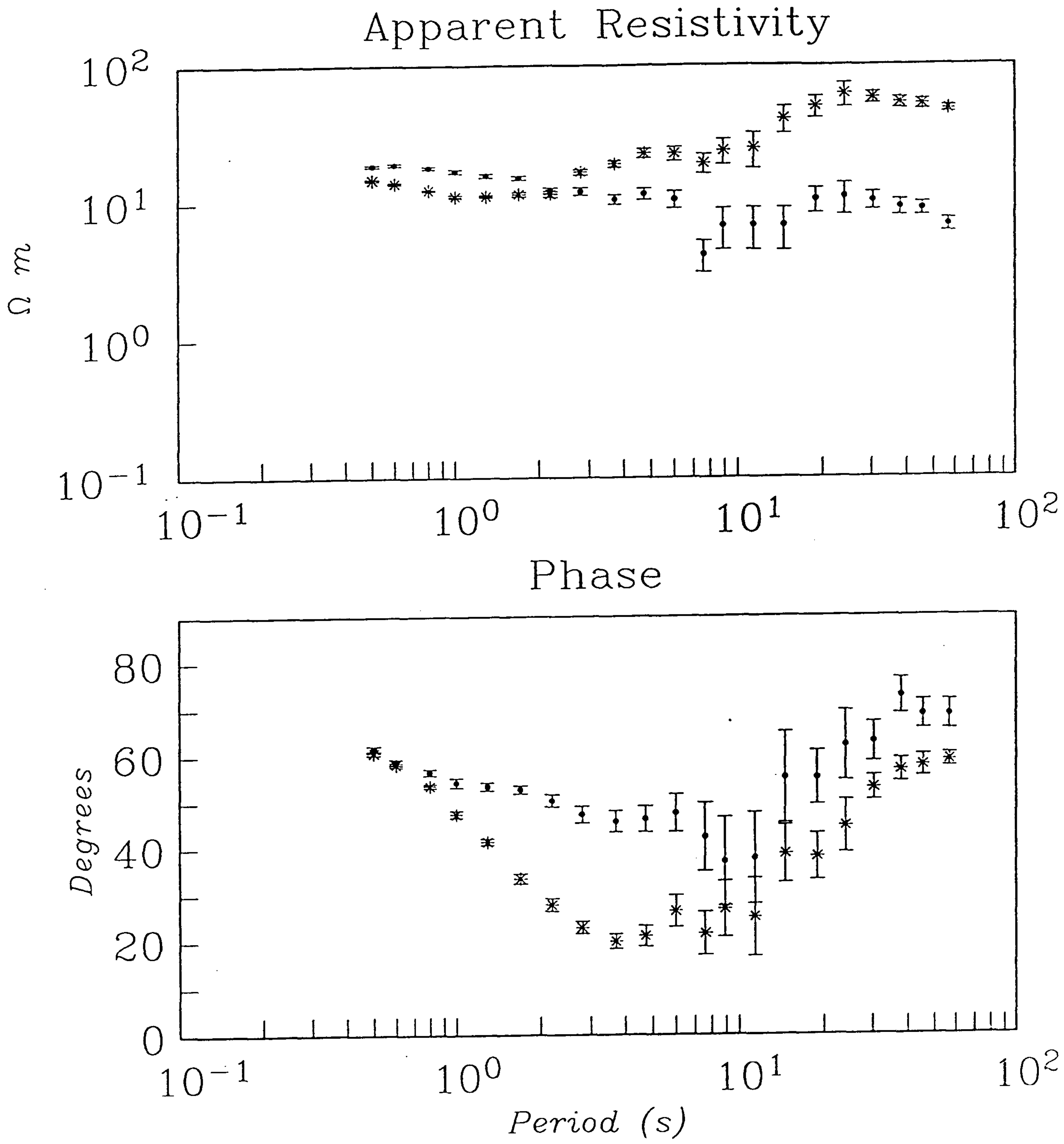

Figure 8

Figure 8: Multiple squared coherence as a function of set number for one component of the electric field from site mslb75. With the coherence pre-sorting scheme data sets with low coherence are omitted from further processing.

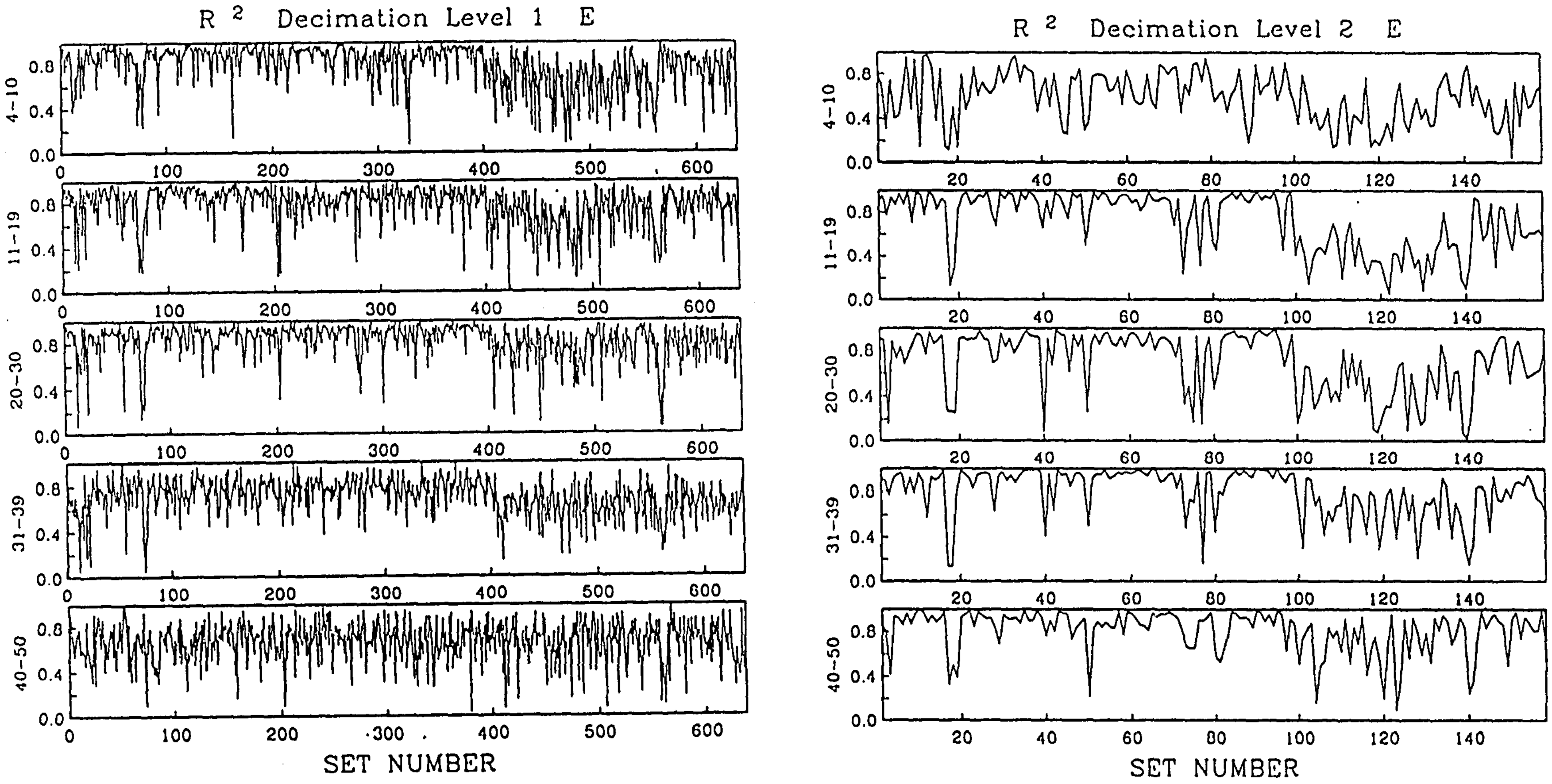

figure 9

Figure 9: As in figure 5, but with coherence pre-sorting. Results are at least as good as with the visual screening.

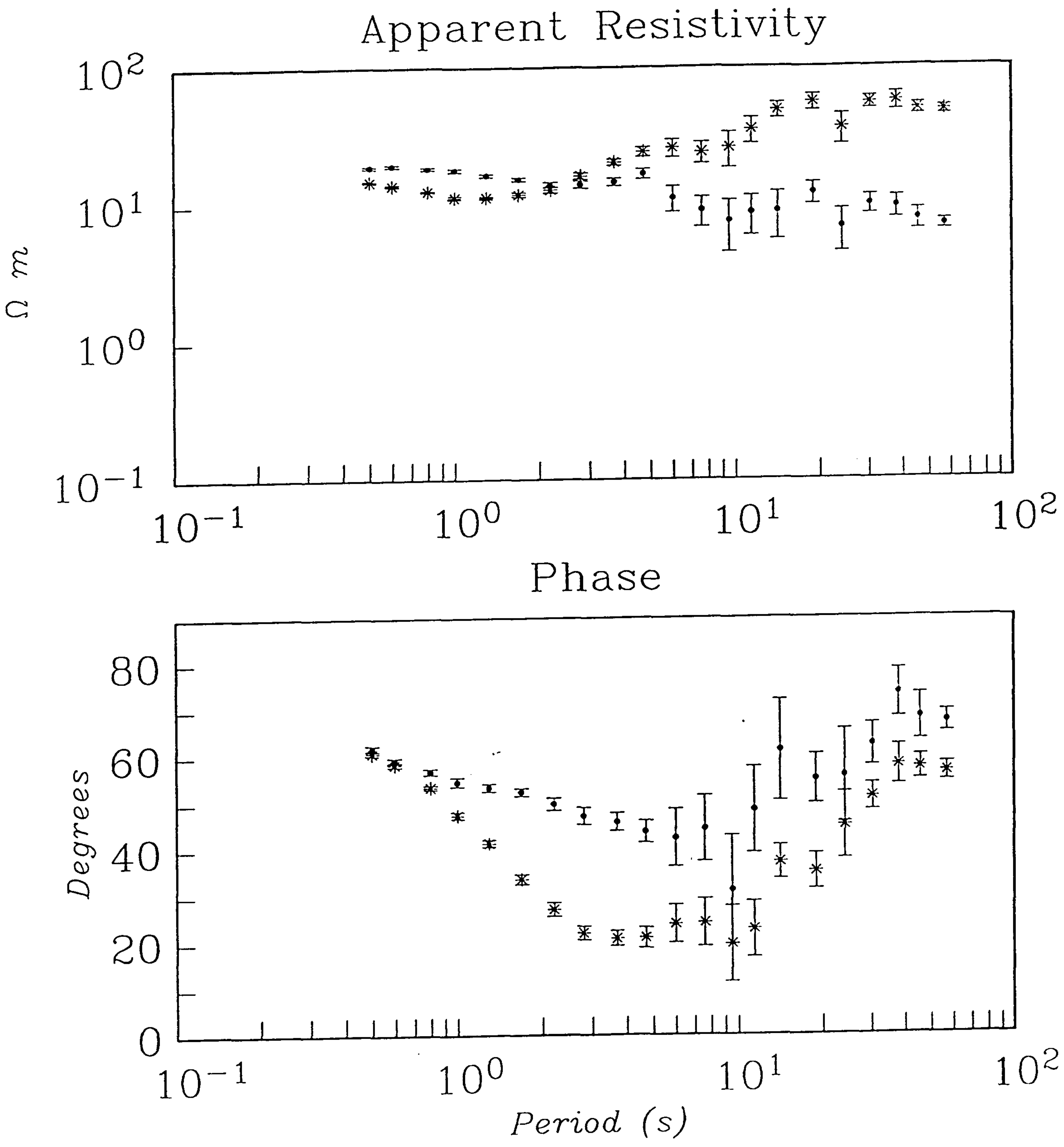

figure 10

Figure 10: As in figure 9, but with visual screening also. The bias in the dead band is almost eliminated, and the estimates vary smoothly with frequency.

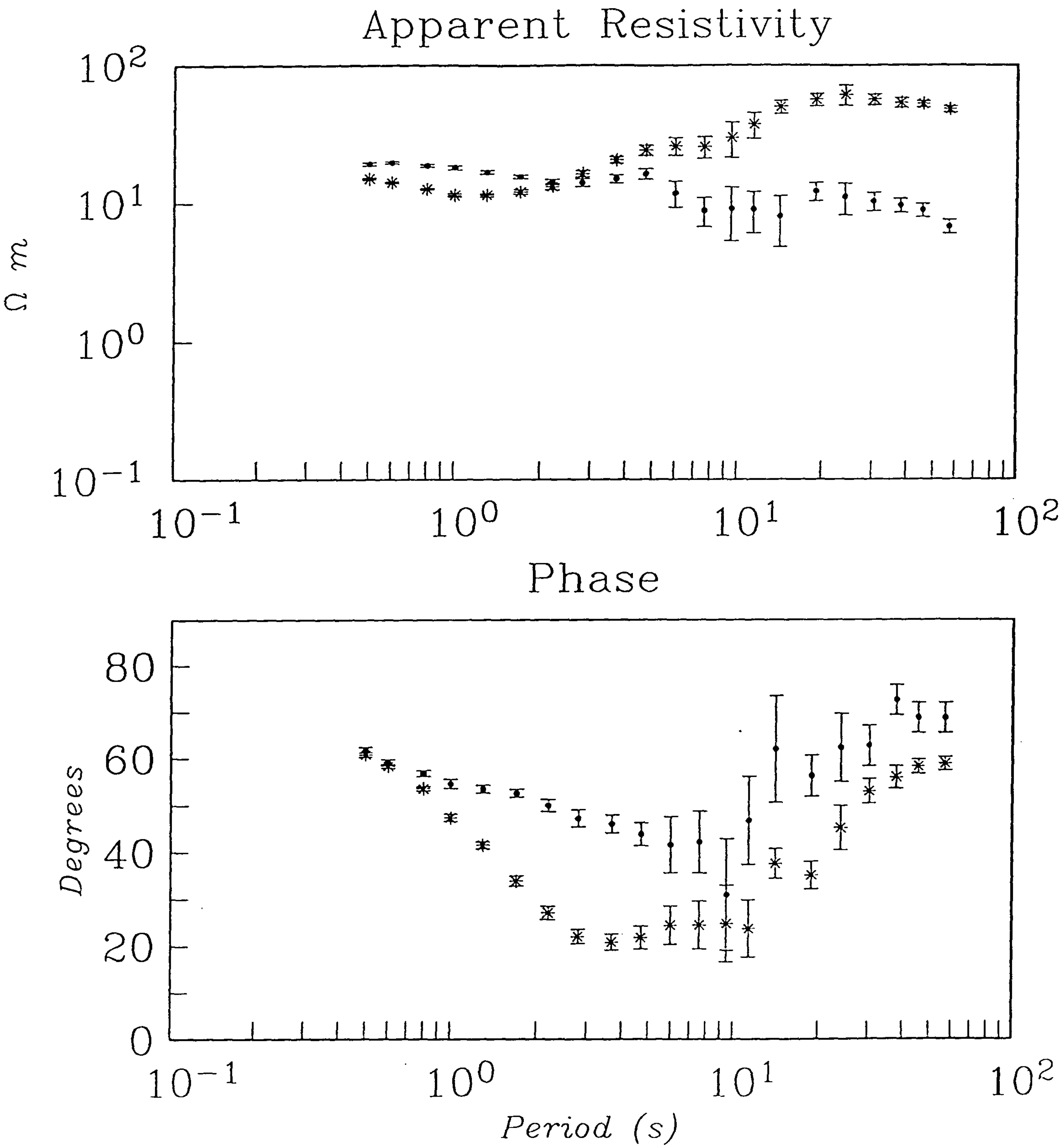

Figure 11

Figure 11: Standard LS with coherence sorting. Without the robust final stacking the coherence sorting does little to improve the estimates.

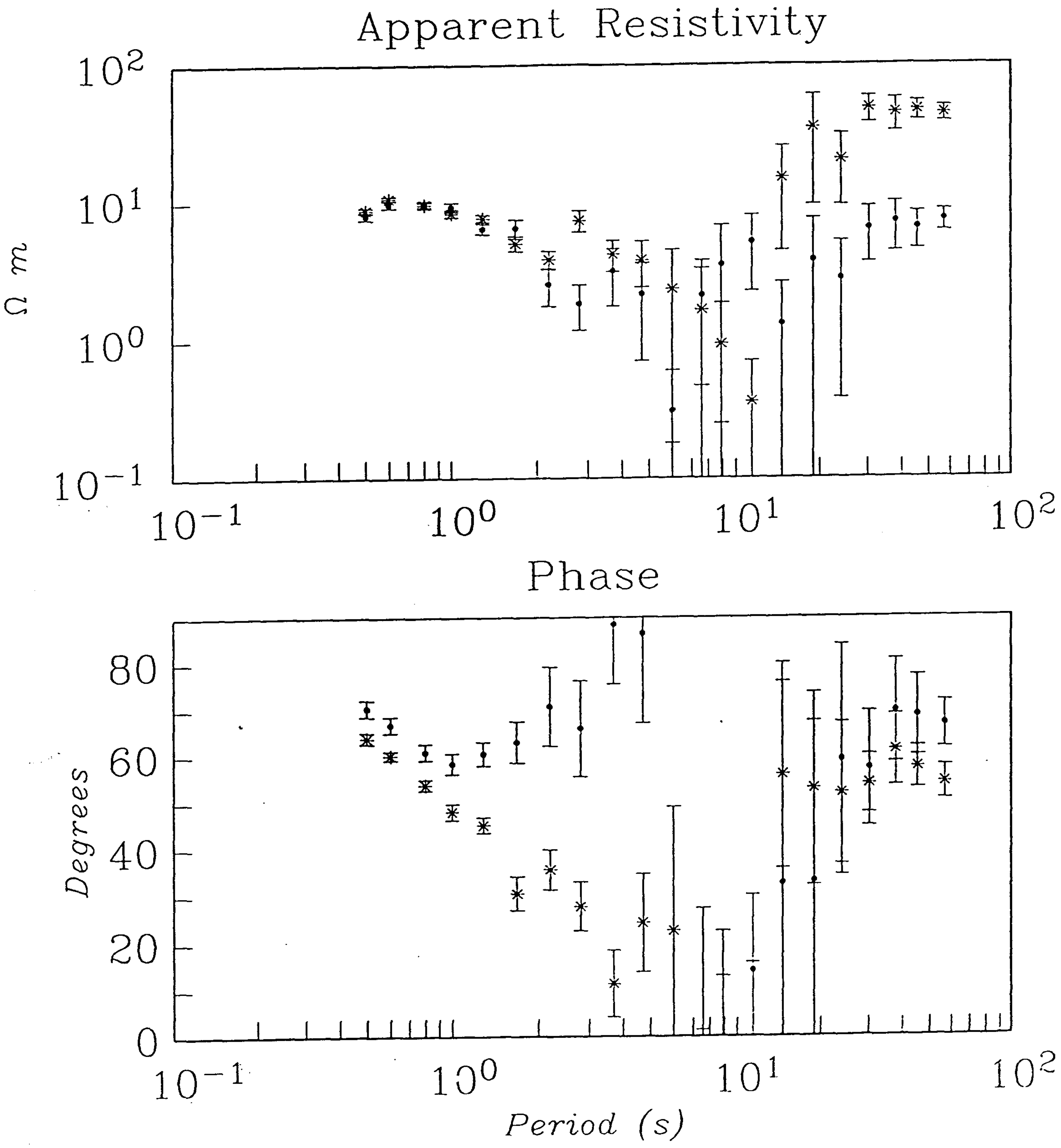

figure 12

Figure 12: (a) Impulse response coefficients corresponding to one component of the impedance tensor estimated for long period MT data. Note that the impulse response need not be causal, because it is band limited. The coefficients go to zero very rapidly (12b), so the the impulse response can be well approximated with a small number of non-zero coefficients. We are developing time domain prediction methods in an effort to detect and automatically remove noise which is well localized in time (e.g., as in figure 2).

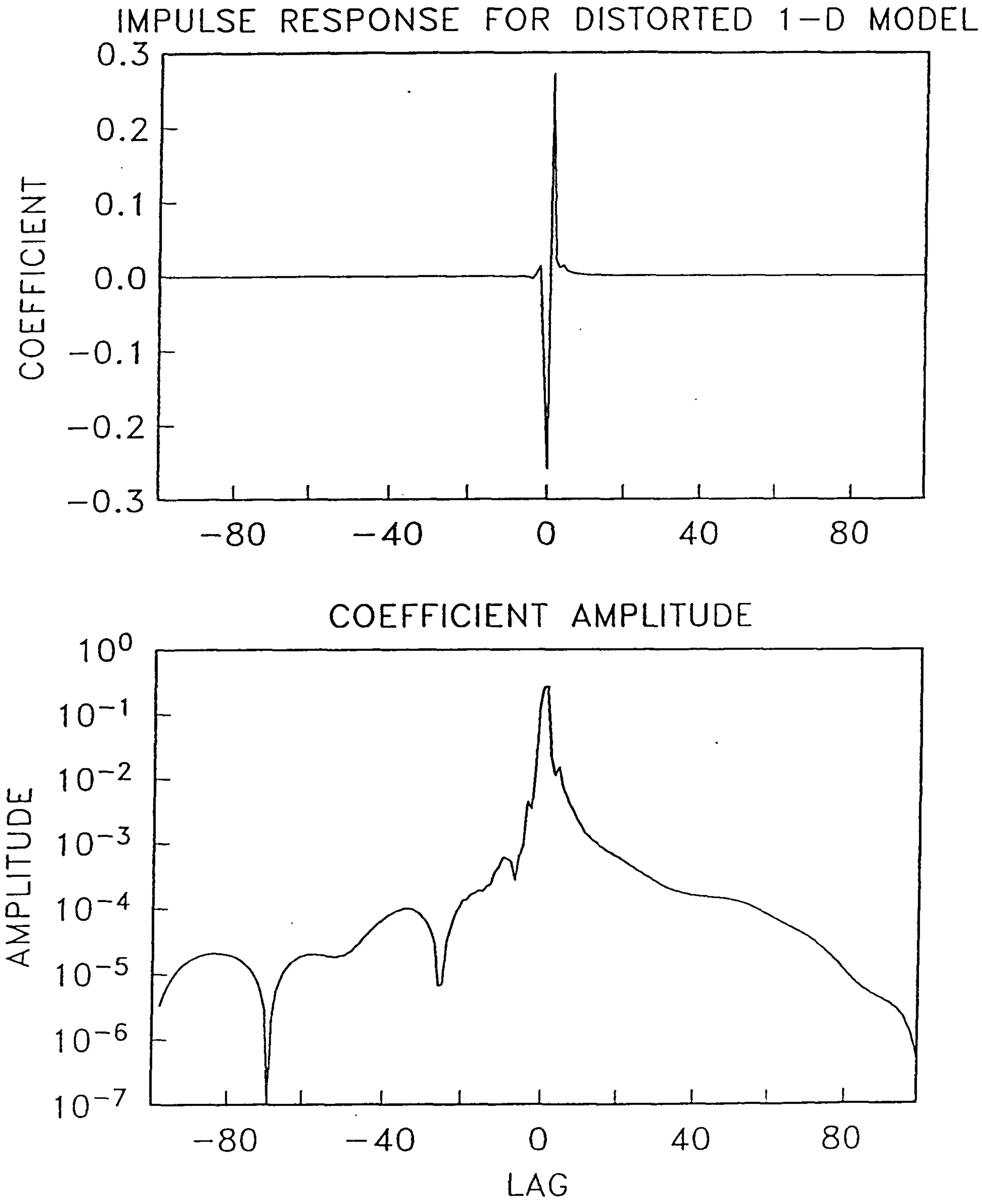

Figure 13 The top panel shows an east-west crossection through the crust of the electrical resistivity in southeastern Saskatechewan, Canada. The resolution of this image almost certainly exceeds that possible with seismic methods other than reflection, and clearly demonstrates the potential of the magnetotelluric method with high quality data and the new processing techniques. The panel just below the model shows the apparent resistivity and phase data predicted by this model as a function of position and period from 10.7 to 455 seconds. The bottom panel shows the measured data in the same pseudosection form. The fit between observed and modelled data is obviously extremely good (about 3% rms). The major conductive body just west of 0 km is known as the North American Central Plains (NACP) anomaly. The other structure extending west from about +200 km has been called the Thomson Belt (TOBE) anomaly. It lies southward and on strike to a major mineralized zone with that name that crops out about 500 km to the north. This name may be incorrect, because there is evidence to the north of this profile that the TOBE maintains its approximate spatial relation to the NACP while both shift west. Furthermore, the TOBE appears to dip west and have a very similar scale and dip as the NACP. An alternative explantion is that both anomalies represent mineralization or berhaps fluids on large scale imbricated thrust faults associated with the trans-Hudson Orogen. Note that the dark area at the east end of the profile is very resistive. Except for the NACP and TOBE structures and a very thin surface layer, the model is everywhere quite resistive. The apparantly high conductivity below the NACP body results from the very high conductivity of this body, which shields the information about the deeper resistive crust.

COPROD II

Rapid Relaxation Inverse - University of Washington