Long Term Electromagnetic Monitoring at Parkfield, CA

by

Karl Neil Kappler

B.Sc. (University of Victoria) 2002
M.S. (University of California at Berkeley) 2005

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Engineering - Civil and Environmental

in the

Graduate Division
of the
University of California at Berkeley

Committee in charge:

Professor Frank Morrison, Co-chair
Professor Jamie Rector, Co-chair
Professor David Brillinger
Professor Gary Egbert

Spring 2008

The dissertation of Karl Neil Kappler is approved:

Co-chair ____ Date ____________________

Co-chair ____ Date ____________________

____ Date ____________________

____ Date ____________________

University of California, Berkeley

Spring 2008

Long Term Electromagnetic Monitoring at Parkfield, CA

by

Karl Neil Kappler

Abstract

Long Term Electromagnetic Monitoring at Parkfield, CA

by

Karl Neil Kappler

Doctor of Philosophy in Engineering - Civil and Environmental

University of California at Berkeley

Professor Frank Morrison, Co-chair

Professor James W. Rector, Co-chair

Electric and magnetic fields in the (10^{-4}-1.0) Hz band were monitored at two sites adjacent to the San Andreas Fault near Parkfield and Hollister, California. Observed fields typically comprise natural magnetotelluric fields, with cultural and instrument noise. A data window [2002-2005], enclosing the September 28, 2004 M6 Parkfield earthquake, was analyzed to determine if anomalous electric or magnetic fields, or changes in ground conductivity, occurred before the earthquake. The data were edited, removing intervals of instrument malfunction, leaving 875 days left in the four-year period. Frequent, local spike-like disturbances were removed. The distribution of these spikes was not biased around the time of the earthquake. Signal to noise ratios, estimated via magnetotelluric processing techniques, provided an index of data quality. Plots of signal and noise amplitude spectra, showed the behavior of the ULF fields to be remarkably constant over the period of analysis. From these first-order plots, it is clear that most of the recorded energy is coherent over the spatial extent of the array. Three main statistical techniques were employed to separate local anomalous electrical or magnetic fields from the dominant coherent natural fields: transfer function estimates between components at each site were employed to subtract the dominant field, and look deeper at the ’residual’ fields; the data were decomposed into principal components to identify linear combinations of array channels, which are maximally uncorrelated; the technique of canonical coherences was employed to distinguish anomalous fields which are spatially broad from anomalies which occur at a single site only, and furthermore to distinguish anomalies which are present in both the electric and magnetic fields form those which are present in only one field type. Standard remote reference apparent resistivity estimates were generated daily at Parkfield. Most of the variation was observed to be seasonal, and frequency independent, suggesting a local seasonal distortion effect. Once corrected for distortion, nearly all of the variability in the apparent resistivity was removed. In all cases, high levels of sensitivity to subtle electromagnetic effects were demonstrated, but no effects which can be described as precursors to the Parkfield earthquake were found.

Professor Frank Morrison, Co-chair

Professor James W. Rector, Co-chair

To Oma

Contents

List of Figures

List of Tables

Foreword

There have been many reports of electromagnetic phenomena associated with earthquakes. There is a 30+ year history here at Berkeley of researchers who have become absorbed by the question: Are there electromagnetic phenomena, or phenomena detectable by electrical, magnetic, or electromagnetic instruments which occur prior to seismic activity? The generally mysterious nature of electromagnetic fields to the layperson, has coupled with the destructive power associated with great earthquakes to generate a body of literature and anecdotal evidence in which it is almost impossible to distinguish scientific observation from lore. Thus, a Berkeley experiment was initiated to accurately measure natural fields at an array of stations over a long period of time in which it would be clear if significant electromagnetic signals occurred above the noise level of the instruments in association with seismic events.

Acknowledgements

I foremost would like to thank Frank Morrison for his guidance through the process of this data analysis. I also gratefully acknowledge the support from my entire committee, Professors Jamie Rector, Gary Egbert and David Brillinger. I also thank Professor Barbara Romanowicz and the staff at the Berkeley Seismological Laboratory for the use of their computing facilities. There have been many others who have provided guidance, discussions, experience or inspiration. In particular, I thank my friend and colleague Vedran Lekic for helping me with many aspects of this work, Sierra Boyd who supported and maintained the array for many years, and Nestor Cuevas at Schlumberger for helpful discussions in MT data processing. I also thank Bill Nickerson for instilling in me an interest in geophysics, and Bill Pfaffenberger and Dan Bergerud for foundations in analysis and linear algebra respectively.

Chapter 1

Introduction

1.1 Introduction

Many investigators have suggested that observations of unusual geoelectromagnetic behavior are associated with seismic activity, e.g. *Corwin and Morrison* (1977), *Fraser-Smith et al.* (1990), *Molchanov et al.* (1992), *Uyeda et al.* (2000), *Varotsos and Alexopoulos* (1984). For a more extensive list, see references in *Pulinets and Boyarchuk* (2004), *Park et al.* (1993), and *Johnston* (1989,1997). Geller (2006, 1996) draws attention to a lack of consensus in the geophysical community on the validity and scientific merit of these reports. The reported anomalies take on a variety of forms: variations in quasistatic electric fields, self-potential, ULF magnetic fields, alternating electric fields in the ULF, VLF or ELF bands, and variations in ground resistivity to name a few. Many, if not most, of these reports seem inconsistent in the reported time interval between the anomaly and the earthquake, the “time-frequency character” of the field perturbation (sharp boxcar-like signals in electric field vs. general increase in spectral level of magnetic field in a particular frequency band), and/or the duration of the claimed anomalies. In general, the entire effort aimed at observing these sorts of phenomena suffers from a lack of multiple earthquake observations to demonstrate a consistent relationship between the observed phenomena and the earthquake. Past efforts have typically been of short duration making it difficult to tell whether an anomalous field was truly a unique precursor or whether such anomalies happen at many times when there is no significant seismic activity.

In 1991, an NSF-sponsored conference aimed at defining some guidelines for the design of experiments that could obtain credible results [*Park et al.*, 1993] was held at Lake Arrowhead, California. A conclusion of the workshop was that there did seem to be some credible observations of precursory effects in ULF magnetic fields, quasi-static DC electric fields, and in ground resistivity.

Three caveats regarding the observations were acknowledged:

a) The cited phenomena are not clear outliers against an objective statistical criterion;

b) The instrument calibrations are not made with respect to some absolute standard, and/or instrument observations are published without citing system transfer functions and noise level analysis;

c) The observations cannot be shown to fit into accepted physical models, and are without plausible physical explanation.

In 1995, Berkeley researchers installed two EM monitoring sites on the San Andreas Fault (SAF) near Parkfield and Hollister, California, shown in Figure 2.1. The Parkfield segment of the SAF was chosen for a focused prediction experiment on the basis of repeating M 6 earthquakes, with a recurrence interval of approximately 22 years [*Bakun and McEvilly*, 1984]. By 1995, the Parkfield earthquake was late, but the probability of recurrence was high, so it seemed a good place for the EM monitoring experiment. The site was also chosen for the availability of other geophysical data and experiments relating to the dynamics of the SAF in this location and the role of water in the seismogenic zone [e.g. *Nadeau and McEvilly*, 1997]. In fact the expected Parkfield earthquake did not occur until September 28, 2004, (M_w6), by which time the array had been in continuous operation for nine years It continues to function as of this writing. The Parkfield segment has also been monitored with a long dipole array for long period (>300sec) electric fields since 1988 [*Park et al.*, 1993, 2007].

The Berkeley EM monitoring array was designed to monitor general fluctuations in the spectra of natural fields in the 10^{-4} to 10Hz band, changes in ground resistivity through the magnetotelluric tensor impedance, and variations in the amplitude and phase of of intersite transfer functions. The array could also be used to search for anomalous variations in quasi-DC electric field as reported by *Varotsos* (e.g. 1984, 1991). However, we found so many anomalous transients in the electric field measurements associated with as yet not understood electrode phenomena that we have not attempted a thorough study of the electric fields. This dissertation reports on a number of different analyses conducted on measurements of electric and magnetic fields in the 10^{-4} to 10Hz band. The array had three key objectives:

a) To provide a continuous data source from which to conduct analysis of intersite transfer functions (TFs);

b) To provide continuous measurements of apparent resistivity using robust MT remote reference processing techniques [*Egbert and Booker*, 1986; *Egbert*, 1997]; and,

c) To archive the MT data stream in a publicly accessible storage medium for the use of other researchers.

The idea behind calculating intersite transfer functions was to use the TF obtained during a period known to be free of earthquake phenomena to predict the fields at one site from the fields at another. Ideally, the difference between the observed and predicted fields should be the effective noise level of the measurement system. Anomalous signals at a site associated with seismogenic events nearby should then stand out above system noise as opposed to being buried in the natural field signal, as they would be with single site data. Residual fields above the instrument noise levels represent signals which do not satisfy the incident plane wave assumption of the MT source-field – solar wind-magnetosphere interactions, or the earth-ionosphere resonance cavity [*Berdichevsky and Dmitriev*, 2002] – as these sorts of fields would be detected at both sites. To our knowledge, no long term monitoring of EM remote-reference residuals along a fault zone have been published to date. Long term monitoring of apparent resistivity has been done passively in the past by [*Eisel and Egbert*, 2002], and using current injection [*Zhao and Qian*, 1992]. We employ analysis similar to *Eisel and Egbert* (2002) . The archiving of the data allows the site to act as a remote reference for other researchers, and validate background fields they may observe.

The estimation of the transfer function between sites is biased by uncorrelated noise at either site. In principle, an unbiased estimate can be obtained using a third site which is presumed to see the same naturally occurring fields, but where noise is uncorrelated with the first two. Unfortunately, due to budget and land permit considerations we were only able to obtain data at three sites simultaneously during the one month period of February, 1999.

We were able to perform long term unbiased impedance (E/H) measurements at each site in the array and to thus monitor the changes in apparent resistivity and impedance phase as a function of time.

Besides these primary purposes, several other measurements and indices can be derived from the long period time series, including MA indices [*Bernardi et al.*, 1989], principal components [*Egbert et al., 2000*] and canonical coherences [*Brillinger*, 1975; *Lyubushin*, 1998]. Finally, the archiving allows observed field amplitudes to be easily cross validated against other auxiliary datasets, such as global geomagnetic activity indices, or local geophysical instruments. The public availability of the data enable any interested researchers to perform their own experiments and signal processing with the array data. The sections on SI-unit conversions and array fidelity measures may be helpful to such researchers.

Chapter 2

Sites and Instrumentation

The specific locations of the sites were chosen in anticipation of the next M_w 6.0 earthquake which seems to be a phenomenon repeating on a time scale of 22 years [*Bakun and McEvilly*, 1984]. In 1996, the array began collecting data. The locations of the sites, separated by 120 km, are shown in Figure 2.1.

Site	Latitude	Longitude	Elev(m)	Date	Location
PKD	35.945171	-120.541603	583	1999/02/05-	Parkfield
PKD1	35.8894	-120.426109	431.6	1995/06/06 - 1999/03/08	Parkfield
SAO	36.76403	-121.447722	317.2	1995/08/15	Hollister

Table 2.1: Specific site locations and dates of operation

At each observatory, three orthogonal induction coils (model EMI BF-4) measure the time varying magnetic field, and two 100m long electrodes measure the electric field in the surface plane. The horizontal coils are buried in trenches 0.5m deep, and the vertical coils

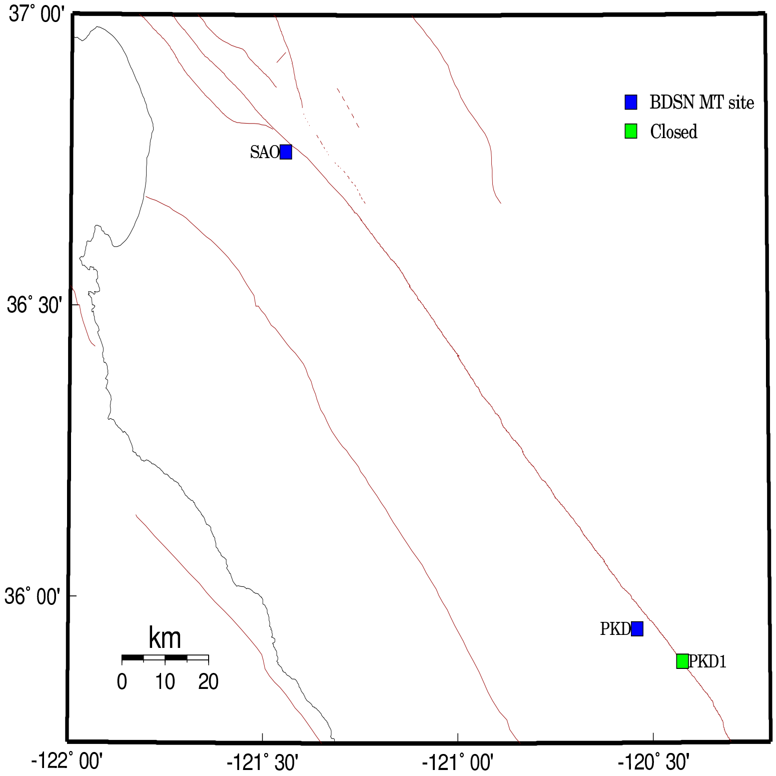


Figure 2.1: Map illustrating the location of operational (filled squares) and closed (grey squares) MT sites in central California.

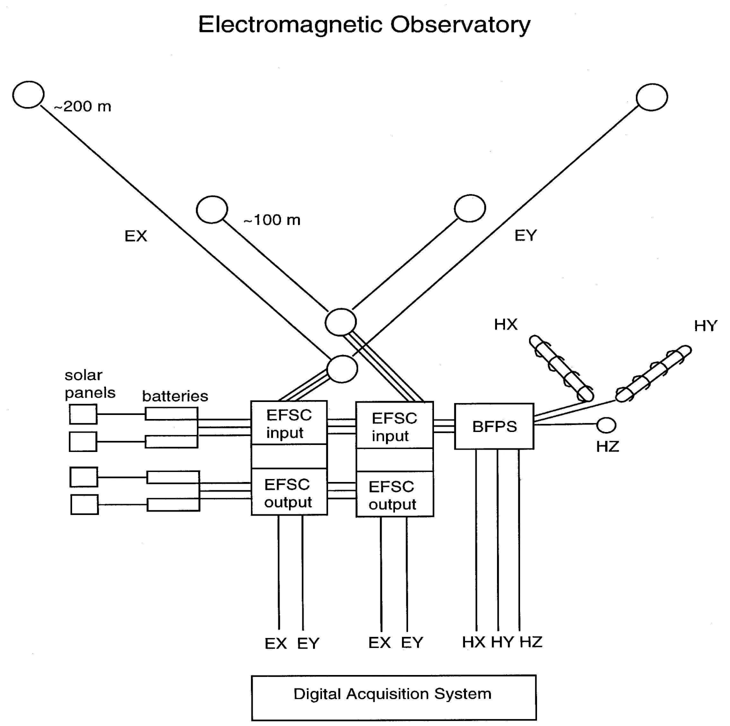


Figure 2.2: Schematic diagram of an EM observatory with two sets of electric dipoles.

are in drilled holes approximately 2m deep. The electrodes are Pb-PbCl non-polarizing type, and are placed in 3m holes with moist bentonite packing to keep contact resistance to a minimum. The entire system is powered by 12V batteries with constant trickle charge provided by on site solar cells. The measurements are filtered onboard the coils for a stable response over several decades of period. The dipole data are preprocessed with an Electric Fields Signal Conditioner (EFSC). The EFSC consists of a preamplifier, optical isolator, and main amplifier in series, with an optional high pass filter. The data are then digitized by 24-bit Quanterra digitizers (Q935 at PKD and Q4120 at SAO), at a sampling rate of 40Hz. Time synchronization is done via Global Positioning System (GPS). The data are then telemetered in packets to the Northern California Earthquake Data Center (NCEDC) where they are archived *(www.ncedc.org)*. A schematic diagram of the site instrumentation is provided in Figure 2.2. At PKD, an added independent pair of 200m dipoles collect data alongside the 100m pair. This is useful for recognizing when an electrode is creating voltage noise, for checking linearity of measurements, and is required should one want to run VAN method analysis [*Varotsos and Lazaridou*, 1991]. For the purposes of this paper we refer to the eight primary array channels as the horizontal magnetic coils at each site, together with the 100m electrodes at both sites. The electrode data preamplifier can be set to 10, 20, 30 or 40dB, where each 10dB of gain corresponds to $\sqrt{10}$ gain in voltage. Gain settings for the preamplifiers are recorded, and each electrical channel is divided through by 3.16, 10, or 31.6 where appropriate (the 40dB setting was never used). Figure 2.3 shows all electrodes and coils, where relative lengths of electrodes are accurate, as are cardinal orientations of all sensors.

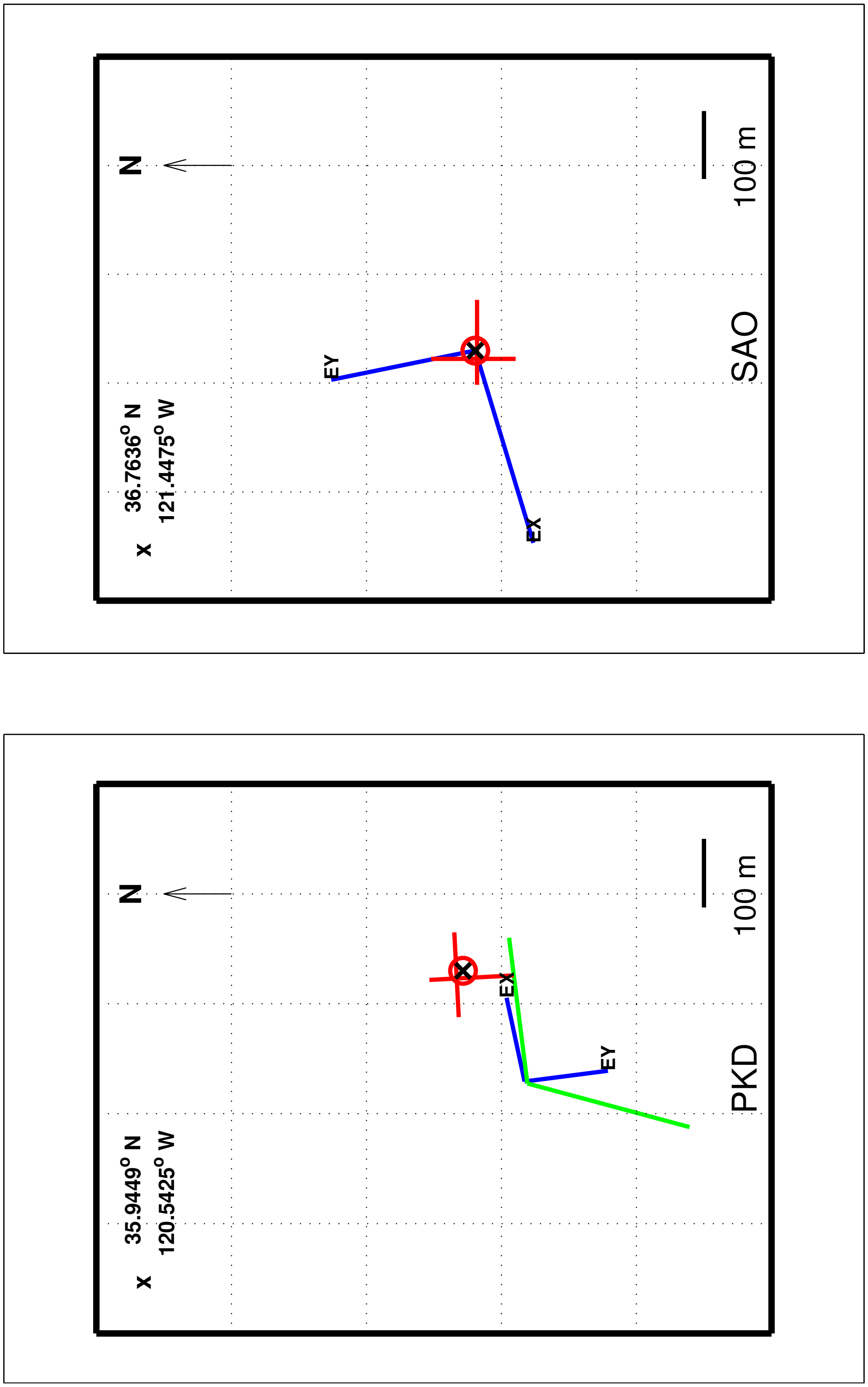


Figure 2.3: Schematic diagram of the EM observatory, showing the azimuths of the electrodes and coils. Electrode lengths are shown to scale, while the coils are scaled to be visible.

Chapter 3

Data Selection and Cleaning

The time window of analysis for this study spans four years, a total of 1461 days. The data examined, unless otherwise stated, were sampled at 1Hz. The raw data are stored as 8x86400 point (day-long) arrays. This is a natural window to use as ULF fields exhibit diurnal behavior. These raw data are stored in units of machine counts at the datalogger.

Before embarking upon the main data processing and interpretation, a data selection exercise is undertaken to ensure that the interpreted time series are relevant to this study. Most data processing will be reserved for days on which all eight channels in the array are “not severely corrupted”. There are many reasons a channel may fail to record physical data, most of which can be lumped under the umbrella term "lack of maintenance". Vandalism, extreme weather, and other unknown factors can also have adverse affects on data acquisition. A point in time is contaminated by missing data if there is no record for one or more channels at that sample. If more than ten percent of the time samples for a given day are contaminated by missing data, then that day is removed from the analysis. 61 days are flagged because of these large gaps in the data stream. An additional 16 days are flagged because EFSCs were not installed at SAO (72-87, 2002). The remaining 1376 days are all considered as candidates for analysis.

3.1 Simultaneity of ULF fields and criteria for omitting data

The science of magnetotellurics (MT) relies upon the recording of tiny variations in the earth's magnetic field. MT data interpretation is predicated on the assumption that the micropulsations of the earth's magnetic field are horizontally polarized, and hence spatially uniform over 100's of km (at least at mid-latitudes). One expects that, to first order, sensors at different sites ought to be strongly correlated. The magnetic fields should be nearly identical (neglecting local noise and instrument malfunctions), and once corrected for individual channel gains, the electric fields should also correlate well. The electric field amplitudes and direction, however, will differ by a scale factor which depends on the local conductivity structure.

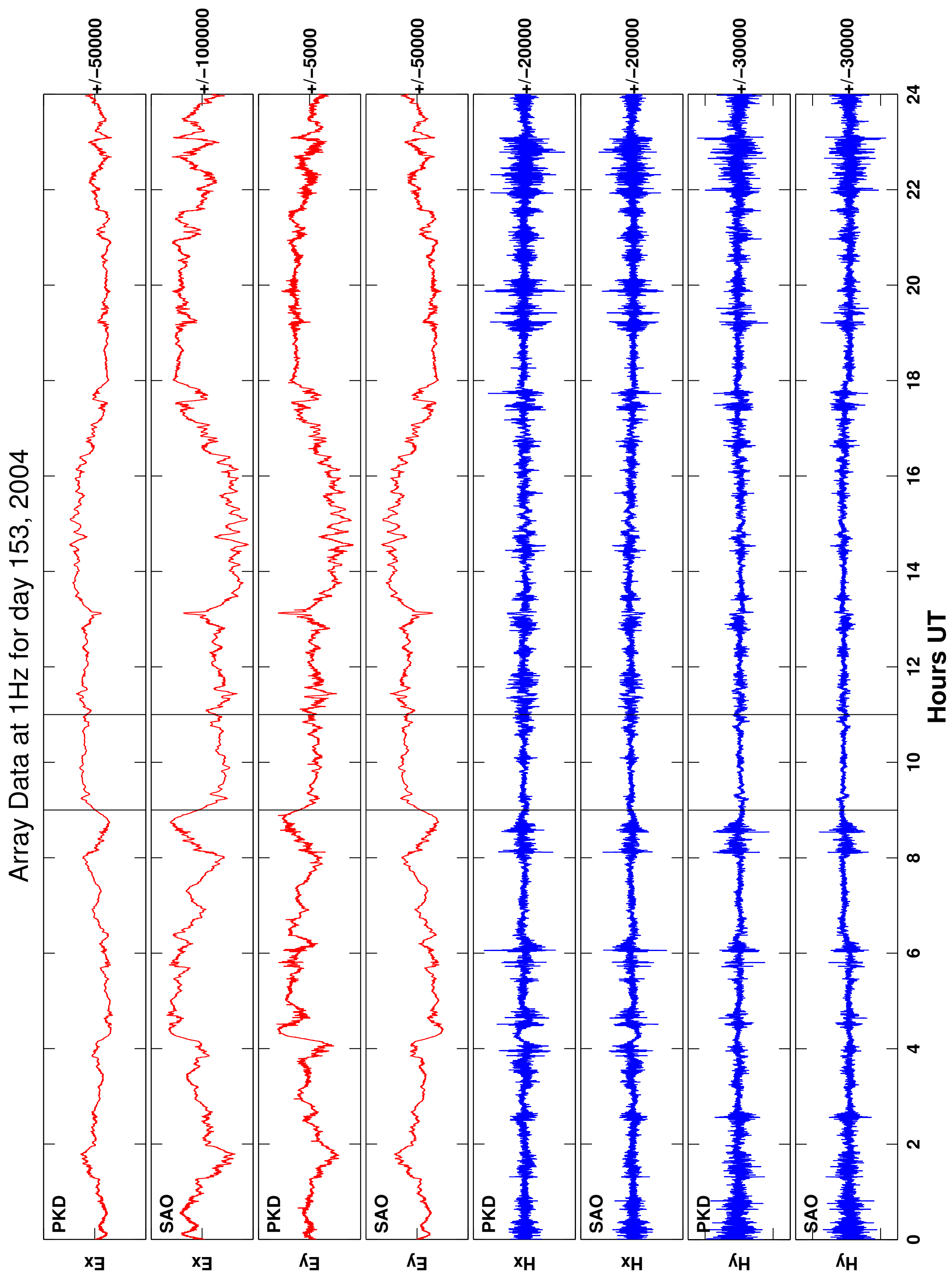


Figure 3.1: A plot of mean-subtracted array data for a full day in 2004. Electric fields are shown in red and magnetic fields in blue. Plots alternate between PKD and SAO at each field Polarity. Y values are in counts with axis limits shown to the left. The vertical lines mark the domain boundaries of Figure 3.2.

The strongly correlated behavior between magnetic channels is a property of the approximate uniformity of the source process over the earth, and the correlations of the electric channels is justified by the physics of Maxwell's equations, specifically Faraday's law. This uniformity of source fields extends over a broad range of frequencies, and as such, the correlations between channels should manifest over a variety of time scales. Figure 3.1 shows this correlation across all array channels. Note the similarity in the field variations, as well as the identical scaling of magnetic fields, compared to the scale factor difference in the electric fields. This similarity holds at higher frequencies, as shown by repeatedly zooming in on the data in Figures 3.2, 3.3, and 3.4.

The coherence of the fields can be clearly seen to extend from hours down to seconds in period. It is this inherent similarity in the fields which is exploited to identify windows in time when the array is not functioning correctly. Bad data is of several different time scales: whole days missing, parts of days or only some channels missing, and spikes or severe contamination of local origin. These are systematically omitted, filled in, or despiked, as described below. A useful statistic for evaluating the system fidelity is the variance of the data over small time windows. With the electric field data scaled as V/m and the magnetic field data still in instrument counts (effectively dB/dt), each day of raw data is partitioned into 450 windows. Each window has length 256, and there is a 64 point overlap from one window to the next. The 450th window is taken as the final 256 points in the day to avoid zero padding.

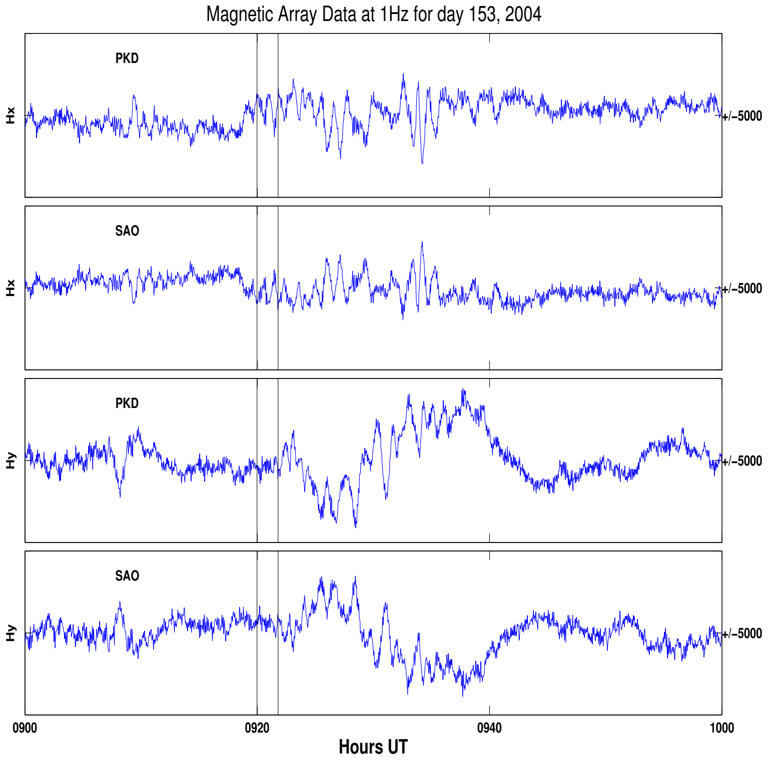


Figure 3.2: Mean-subtracted magnetic array data for one hour of the day shown in Figure 3.1. Plots alternate between PKD and SAO at each field polarity. The vertical lines mark the domain boundaries of Figure 3.3.

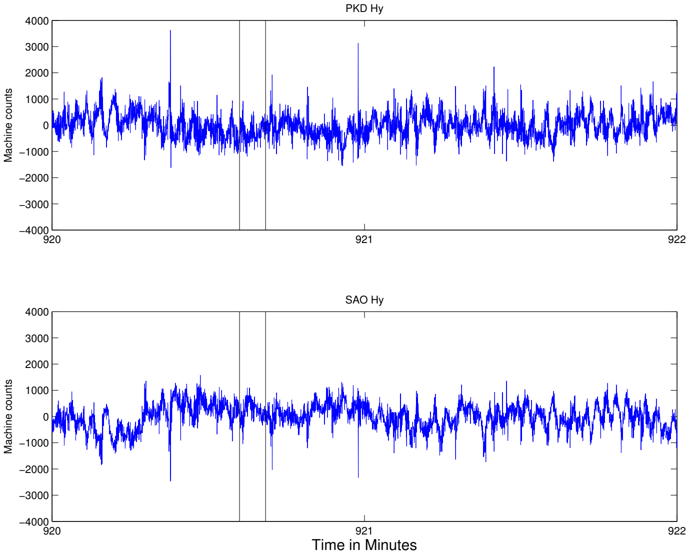


Figure 3.3: Mean-subtracted time series of y-polarity magnetic channels for two minutes within the hour shown in Figure 3.2. The vertical lines mark the domain boundaries of Figure 3.4

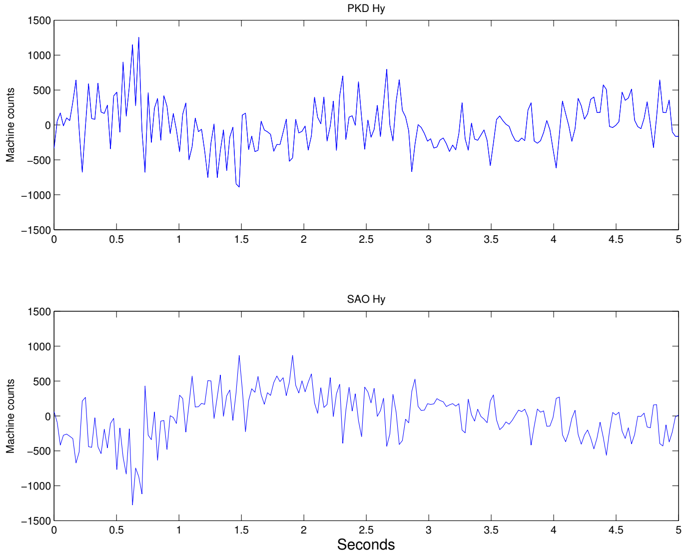


Figure 3.4: Mean-subtracted time series of y-polarity magnetic channels for 5 seconds within the window shown in Figure 3.3.

The ith sensor S$_i$ is then represented as a time series of window-variances $\mathbf{S}_i$(t), where the index t increments by one for each 256-second time window. By compressing the time series into this variance-window format, the ratios of the time series (taken bin by bin) can be examined for deviant field behaviour. Formally, for two sensors represented as $\mathbf{S}_i$(t), and $\mathbf{S}_j$(t), we calculate the log-variance-ratio time series $\mathbf{V}_{S1,S2}$(t), defined for t centered in a 256-point window as $\log_{10}$(var(S$_i$(t))/var(S$_j$(t))). Days where the 90th percentile of $\mathbf{V}_{S1,S2}$(t) deviated by more than 1 from its median value were flagged as unusable. This translates to removing days where more than ten percent of the data were off by more than a factor of ten from their well-behaved values. The median values cited above are calculated on blocks of days where instruments are unchanged. Combinations of S$_i$, S$_j$ are chosen to be sister channels at the remote site, i.e. (Hx at PKD together with Hx at SAO) or (Ey at PKD, with Ey at SAO) etc.

The flagged days (a total of 586 of them) were inspected by eye to confirm that one or more sensors were indeed malfunctioning. An example of the $\mathbf{V}_{S1,S2}$(t) is shown in Figure 3.5. The extreme data, with absolute y-values greater than 10^4, correspond to times when one site was not recording. Summary inspection of the raw data time series associated with these outliers shows a diurnal drift in one sensor on the order of around 200 datalogger machine counts (Boltzmann kT noise), with the digitization noise superimposed.

Previously, we had employed a code developed by Gary Egbert to obtain daily estimates of signal to noise ratio (SNR) for each instrument. Details of this code are laid out in the

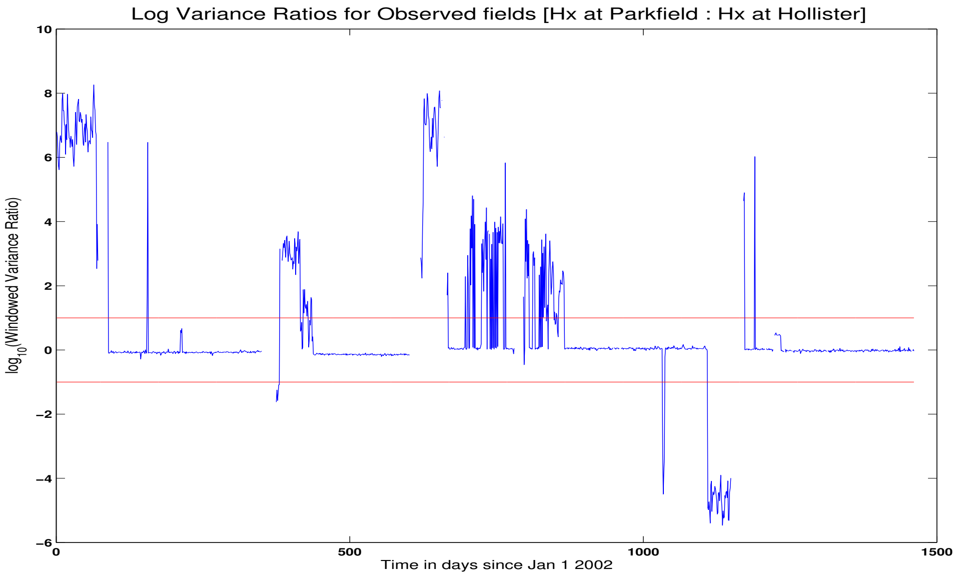


Figure 3.5: Variance ratios plotted for for each of the 1379 candidate days for analysis. This is the ratio of Hx at PKD to Hx at SAO.

next section. Sections of data where the code failed to converge, or returned SNR values less than two, had been classed as bad data and were removed from the analysis. This removed some severely corrupted data, but left many sections of poor data which had to be edited by hand. The 586 flagged days are a proper subset of the more than 800 days which were previously flagged by hand-selection, and a proper subset of the 678 days which were flagged by using the SNR criterion applied to Egbert's code. These days include days when the data logger was not working, or there was a problem with the power supply to the BFPS, or the amplifiers in the EFSC. Figures 3.6 and 3.7 depict a sample sensor, Hy at SAO, in terms of a histogram of its log variance ratio time series, before and after this initial data elimination phase. It is clear that there are two extra distributions in the data, which have been identified by inspection as corrupt data.

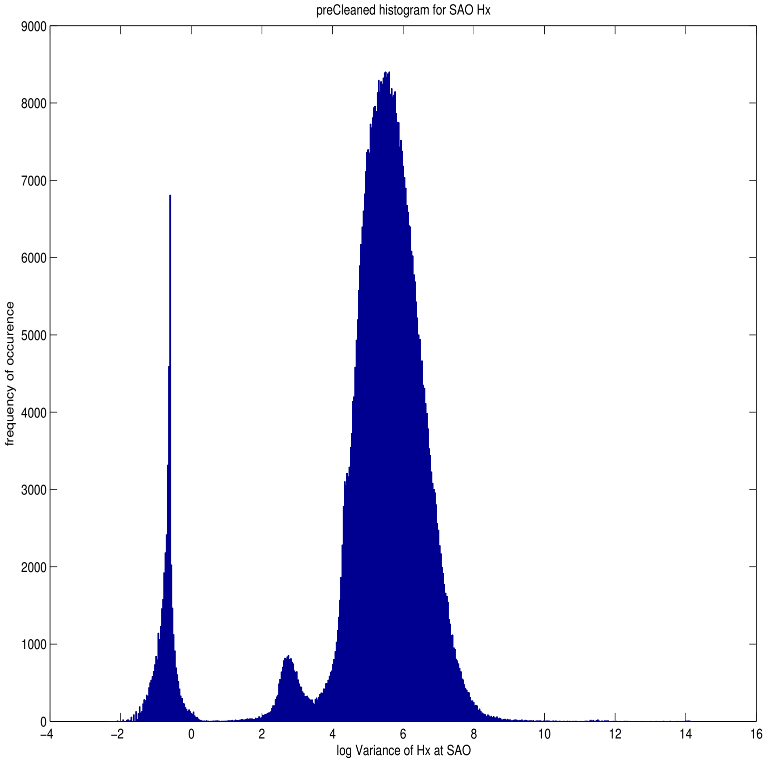


Figure 3.6: A histogram showing the distribution of the $\log_{10}$(window-variance time series) for Hx at SAO prior to the automated data rejection.

Though there are many flagged windows in the remaining 802 days, there are not enough flags to cross the threshold for eliminating the day's data. The remaining days are treated with a slightly more refined data rejection criterion. These remaining days are represented by $\mathbf{V}_{S1,S2}$(t), for four pairings of S_1, S_2 corresponding to each channel at PKD paired with its sister channel at SAO. For each day, the time series$\mathbf{V}_{S1,S2}$(t) is median subtracted, and scanned for outliers according to the following criteria: If a realization of the median subtracted $\mathbf{V}_{S1,S2}$ has absolute value greater than r one-sided α-trimmed standard deviations,

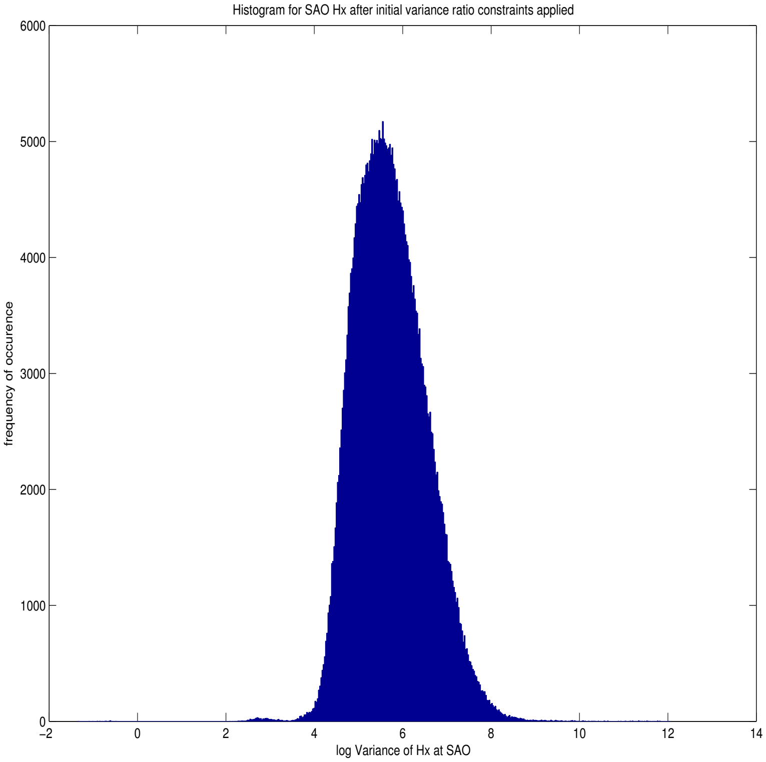


Figure 3.7: A histogram showing the distribution of the $\log_{10}$(window-variance time series) for Hx at SAO following the automated data rejection.

the window is flagged as needing replacement for the channel in which more energy was observed. The one-sided α-trimmed standard deviation is defined as the standard deviation of the set of all elements of $\mathbf{V}_{S1,S2}$(t) which are below the (1-α)th percentile. Using this measure prevents a few very large spikes from driving the standard deviation up so high that smaller spikes are not caught. Practically speaking, a value of r=5 works well for magnetic fields, and r=6 for electrics.

The described algorithm only identifies windows where the variance ratio between two channels is anomalous. It is another matter to decide which of the two channels in the ratio is the non-physical data. Because most nonphysical data tend to be spikes and sharp offsets, the channel with more energy is selected by default as the offending data. There is, however, a certain pathological case which occurred repeatedly during the period December 2003 through March 2004, marked by a power failure to the coil amplifiers at SAO. There are twin power supplies running to the coils, which take turns carrying the load, switching every few hours according to a program. When the faulty power supply was connected, the coils at SAO showed anomalously small variances. Consequently, the spike identification algorithm misidentified the large intersite variance ratios as "spikey" data at PKD, as opposed to malfunctioning amplifiers at SAO. Data suffering from this problem have two properties. First, the underpowered amplifiers affect the datastream for at least 20 minutes, and second, both coils at SAO are affected simultaneously. Because of these facts, the magnetic variance ratio time series from December 2003 to March 2004 were searched for contiguous segments of five or more flags, simultaneously present in both PKD coils.

It turns out that this simple search identified all the pathological cases, and only the pathological cases. The data replacement flags were then moved from PKD windows to SAO windows for these times. More sophisticated and general searches can be added for future data. For example, if the BFPS power problem needs to be more broadly searched for, one could use the approximate stationarity of Hx/(d(Ey)/dt) and look for times this quantity drops at SAO, but remains normal at PKD.

3.2 Time domain despiking and treatment of small discontinuities

The variance ratio statistic described above is by no means a complete noise-spike identification algorithm, nor are we guaranteed that the flagged windows are corrupted by spikes. Nevertheless, it does identify large spikes successfully. The despiking algorithm described below does not require that a true spike be present in the data, only that a window in channel i has been flagged as being non-physical. Flagged windows are removed from the data entirely, and as such become gaps in the data stream. Thus, a single algorithm can treat spikey data together with actual gaps in the data stream. The flagged window is then replaced by a window of physically plausible data, which is spliced into the observed time series in a manner which attenuates inherent offsets at the boundary between the observed and predicted data. A description of the calculation of plausible data and the method of splicing which have been applied to the dataset is as follows: Consider a 256-point time window for which one or more channels have been flagged. We focus on the i^{th} flagged data channel. The physically plausible data $\mathbf{d}_i$(t), hereafter denoted $\mathbf{d}$(t), is calculated by Wiener filtering the non-flagged channels to obtain an approximation of the flagged channels. That is $\mathbf{d}$=$\Psi_j \star \mathbf{m}_j$, where the double index denotes summation over all channels of input data $\mathbf{m}_j$, convolved with Wiener filters Ψ_j. The input channels $\mathbf{m}_j$ are mean subtracted prior to all calculations, and so $\mathbf{d}$(t) will be plausible in its variations, but offset by a static shift which will be accounted for below. Formally, to calculate plausible data for a given time window, channels are separated into two disjoint sets. One set $\mathscr{F}$ is a list of channels which have been flagged, and the other set $\mathscr{F}'$ is a list of all other channels in the array. For each data channel d$_i$ in $\mathscr{F}$, a set of K=$^{\#}\mathscr{F}$ N-point (N=13) Weiner filters are calculated which predict the data in channel i using the data in all channels of $\mathscr{F}'$. The training data used to calculate the Wiener filter coefficients is a two hour segment of unflagged data which has the following three properties:

1. The time window of the training data spans the time of day t of the flagged window; 2. The training data are from the same block as the flagged data (i.e. instruments are identical); and 3. The A$_p$ (geomagnetic activity) index of the day from which the training data are selected is as close to the A$_p$ index of the day of the flagged window as possible while respecting conditions one and two.

Let $\mathbf{d}_{trg}$(t) be the training data from channel i, stored as a column vector. Then let $\mathbf{m}_{j_{trg}}$(t) denote the training data from the jth channel in $\mathscr{F}'$, j=1..K, stored as a row vector. Then the prediction filters, Ψ_j are found by solving the system of equations

$$\mathbf{d}_{trg}(t) = \sum_{k=1}^{K}(\sum_{n=1}^{N}(\Psi_k(n)\mathbf{m}_{ktrg}(t - \frac{N-1}{2} + n))) \tag{3.1}$$

The inner summation corresponds to the inner product of the k^{th} Weiner filter with N points of the k^{th} predictor channel centered at t. Denoting the vector made of N-points of predictor channel j, centered at t as $\mathbf{m}_k^{t,N}$ we find

$$\mathbf{d}_{trg}(t) = \sum_{k=1}^{K} < \Psi_k, \mathbf{m}_k^{t,N} > \tag{3.2}$$

Translating the above expression into linear algebra with a convolution matrix we have:

$$\begin{bmatrix} d_{1+\frac{N-1}{2}} \\ d_{2+\frac{N-1}{2}} \\ \vdots \\ d_{T-\frac{N-1}{2}} \end{bmatrix} = \begin{bmatrix} m_{1,1} & \cdots & m_{1,N} & \cdots & \cdots & \cdots & m_{K,1} & \cdots & m_{K,N} \\ m_{1,2} & \cdots & m_{1,N+1} & \cdots & \cdots & \cdots & m_{K,2} & \cdots & m_{K,N+1} \\ & \vdots & & & \vdots & & & \vdots & \\ m_{1,2} & \cdots & m_{1,T-N} & \cdots & \cdots & \cdots & m_{K,T-N} & \cdots & m_{K,T-N} \end{bmatrix} \begin{bmatrix} \psi_{1,1} \\ \vdots \\ \psi_{1,N} \\ \vdots \\ \vdots \\ \vdots \\ \psi_{K,1} \\ \vdots \\ \psi_{K,N} \end{bmatrix} \tag{3.3}$$

We solve the above equation where **d** and **m** are made of training data for the KxN Wiener filter coefficients Ψ. Where the subscripts i,j, on the m denote the j^{th} element of the i^{th} channel of training data indexed by $\mathscr{F}'$. Because the training data are unflagged, ordinary least squares is sufficient to solve the problem, i.e we needn't worry about damped least squares or outliers.

Note that the input data matrix M has the form of a concatenation of K convolution matrices each having N columns. When M is constructed of input time series having length T we do not recover T points of **d**, losing (N-1)/2 points on either side of the data time series. This is because estimates of **d**(t) when the filter is 'entering or leaving' the time series **m** are rejected.

Once an appropriate Ψ has been obtained, **d**(t) is predicted using Equation 3.3 but this time with Ψ as a known input, and the **M** matrix replaced by the observed data in the time window needing repair on channel *i*. The resulting **d** is the output plausible data. Here, enough padding is applied to the $\mathbf{m}_j$ so that the resultant $\mathbf{d}_{plaus}$ has width 1.1*T, with 5% overlap of predicted data to either side of the gap. At this point it is time to replace the gap. There are five regions in time domain that we must consider separately: $\mathscr{R}_i$, i=1..5. These are shown, partitioned by vertical lines, in Figure 3.8.

In each region, we adjust the data to be some combination of the predicted data and the observed data.

$$\mathbf{r}_1(t) = \mathbf{d}_{obs}(t) \quad \forall t \in \mathscr{R}_1 \tag{3.4}$$

$$\mathbf{r}_2(t) = [(1-\mathbf{T}_{down})(\mathbf{d}_{obs} - \bar{\mathbf{d}}_{obs}) + \mathbf{T}_{up}\mathbf{d}_{plaus}] \forall t \in \mathscr{R}_2 \tag{3.5}$$

$$\mathbf{r}_3(t) = d_{plaus} \quad \forall t \in \mathscr{R}_3 \tag{3.6}$$

$$\mathbf{r}_4(t) = [\mathbf{T}_{up}(\mathbf{d}_{obs} - \bar{\mathbf{d}}_{obs}) + (1-\mathbf{T}_{down})\mathbf{d}_{plaus}] \quad \forall t \in \mathscr{R}_4 \tag{3.7}$$

$$\mathbf{r}_5(t) = d_{plaus}(t) \quad \forall t \in \mathscr{R}_5 \tag{3.8}$$

The diagonal matrices T_{down} and T_{up} correspond to tapering matrices, which upweight the observed data far form the window, and downweight the prediction, by inversely varying the weights on the observed data and plausible data as the window edge approaches. Formally,

$$\mathbf{T}_{up(i,i)} = cos(\frac{-\pi(i-N)}{(2N)}) \qquad \mathbf{T}_{down} = \mathbf{I} - \mathbf{T}_{up} \tag{3.9}$$

Finally the time series in the five regions r$_i$ are shifted so that they are free from sharp offsets at each boundary, and concatenated into a cleaned data vector $\mathbf{d}_{new}$. To each of regions 2,3,4 is added a shift which is the difference in the median values of r1 and r2, and to r5 is added a shift which is the difference in the median values of r4 and r5, where only the 5 points nearest the boundary are considered in the median calculation.

Thus, a whole day's data channel is finally replaced by $\mathbf{d}_{new}$

$$\mathbf{d}_{new}(t) = [\mathbf{r1} \mid \mathbf{r2} + s12 \mid \mathbf{r3} + s12 \mid \mathbf{r4} + s12 \mid \mathbf{r5} + s45] \tag{3.10}$$

where

$$s_{1,2} = median\{\mathbf{r}_1(t) \mid junction - t < 5)\} - median\{\mathbf{r}_2(t) \mid junction - t < 5)\} \tag{3.11}$$

$$s_{4,5} = median\{\mathbf{r}_4(t) \mid junction - t < 5)\} - median\{\mathbf{r}_5(t) \mid junction - t < 5)\} \tag{3.12}$$

A sample result of the despiking is shown in Figure 3.8.

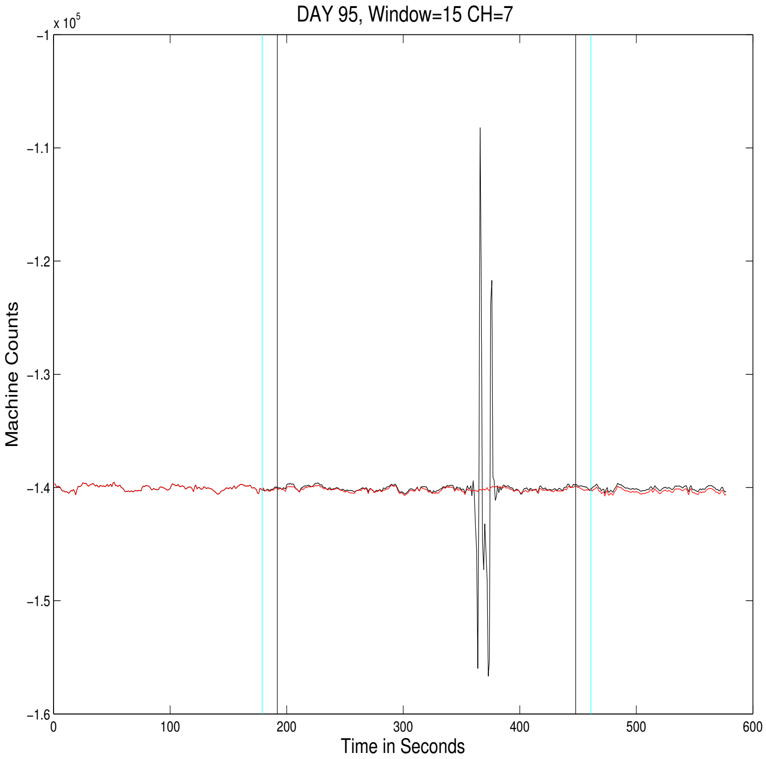


Figure 3.8: An example of the despiking routine results. The recorded data is in black, and the plausible data calculated by the method outlined in the text is shown in red. Black vertical lines bound the clipped section ($\mathbf{r}_3$) and the regions bounded between the cyan and black vertical lines are $\mathbf{r}_2$ and $\mathbf{r}_4$ (the splice regions).

The shift between r4 and r5 is especially important to incorporate when there are steps in the data, as for example in Figure 3.9.

A total of 3904 windows are flagged, and the distribution of these flags in time is shown as a sum over all channels. We note that there do not seem to be any periods of time which are much more contaminated than any other periods. In other words, there is no time which

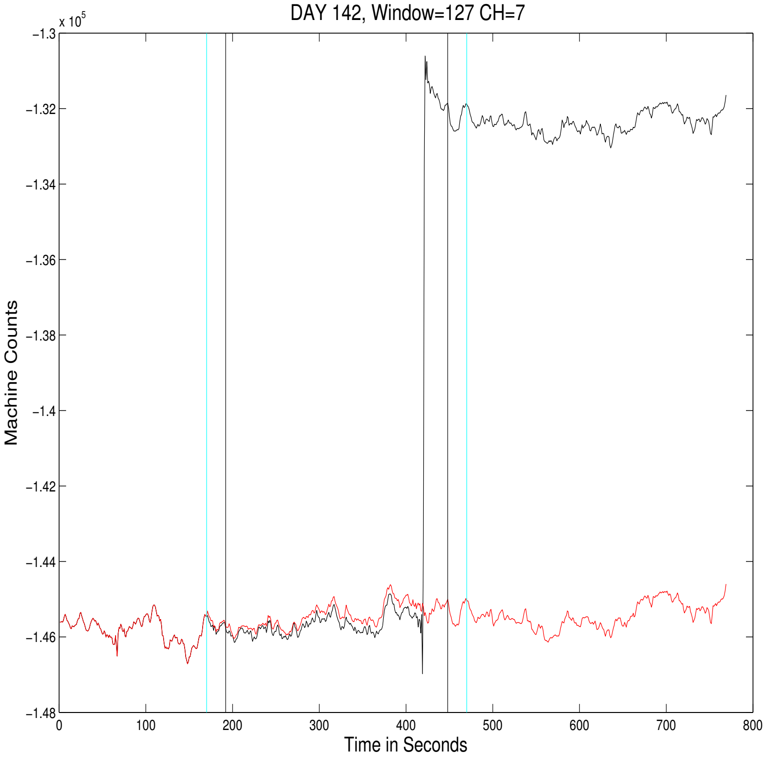


Figure 3.9: An example of the despiking routine results. The recorded data is in black, and plausible data calculated by the method outlined in the text is shown in red. Black vertical lines bound the clipped section ($\mathbf{r}_3$) and the regions bounded between the cyan and black vertical lines are $\mathbf{r}_2$ and $\mathbf{r}_4$ (the splice regions).

stands out as being especially prone to spikes.

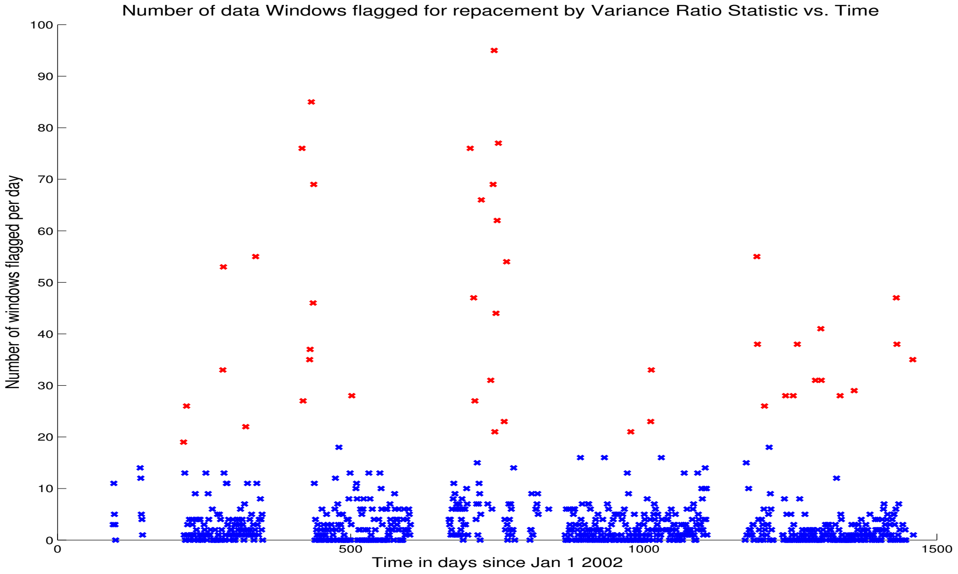


Figure 3.10: Distribution of flagged windows for all channels vs. time. The y axis shows the number of spikes on a given day, while the x axis is an index of the days. The 95th percentile of these points is shown in red.

Chapter 4

Signal and Noise

Now that large, clearly non-physical signals have been removed from the data, it is time to move to a more sophisticated measure of array fidelity in order to confirm times when the array can be truly considered to be functioning properly. Such a metric ought to be frequency dependent as both the character and strength of the ambient fields tend to vary with frequency. The data are transformed into the spectral domain, and a variant of Egbert's RMEV (Robust Multiple Channel errors in Variables) [*Egbert*, 1997; *Eisel and Egbert*, 2002] frequency domain cleaning algorithm is employed to generate estimates of signal strength and incoherent noise power for each channel in each of several frequency bands. By doing this, we can express the observed fields in terms of signal to noise ratio. Thus, should any anomalous fields be observed, they can be quantified in terms of their significance above the instrument noise level. This sort of noise level analysis is seldom done in the reports of signals associated with earthquakes. A byproduct of these calculations are daily averaged estimates of spectral matrices which will be exploited to explore the behaviour of ULF fields.

4.1 Conversion of data to SI unit Fourier Coefficients

For each instrument, and each day, we Fourier transform the data in order to obtain a time-series of Fourier coefficients (FCs). Hamming windows of 256 points, with 64 point overlap are applied in time domain. Although no band averaging is applied until after the RMEV routine is run, the bands are chosen at the beginning of the process so that extra FCs need not be stored or operated on. Long period Fourier coefficients are obtained by repeatedly decimating the data by a factor of two and reapplying the same windowing scheme. This results in six decimation levels, spanning periods from around 3 to 1500s. Windows are chosen to have a constant Q $=\mathrm{f}_{center}/\mathrm{f}_{binwidth}$. A total of 32 distinct bands are calculated. The resulting time series of the Fourier-transform and decimation routine have lengths 449, 224, 112, 55 27 and 13, for decimation levels one through six, respectively. During the FC calculation, manufacturer's instrument transfer functions are used to correct for amplifiers in EFSCs and the frequency dependent gains of the coils, so that FC time series data use the standard units of magnetotelluric data: mV/km/$\sqrt{Hz}$ for electrodes, and nT/$\sqrt{Hz}$ for coils.

Time domain data in units of counts are first scaled into units of Volts by using the conversion factors stored at the NCEDC. This accounts for the variation in the exact conversion factors from channel to channel of a single data logger, and from datalogger to datalogger. Nominally, the conversion is on the order of 400000 counts per Volt. The voltages observed across the electrodes are then corrected for the gain settings on the EFSCs, and normalized by the electrode lengths so they are expressed in units of V/m.

$$\mathbf{E}_{\mathbf{i},mV/km}(t) = 10^6 \frac{\mathbf{E}_{raw}(t)}{(cpv_i)(g_i)(l_i)} \tag{4.1}$$

where cpv_i is the digitization factor, g_i the EFSC gain setting, and l_i the length, all referring to the i^{th} electrode. The factor of 10^6 converts V/m to mV/km. For coils:

$$\tilde{\mathbf{H}}_{\mathbf{i},V}(t) = \frac{\mathbf{H}_{raw}(t)}{(cpv_i)} \tag{4.2}$$

The data are then windowed and FFT-ed which is equivalent to:

$$\tilde{\mathbf{E}}_{\mathbf{i}}(\omega) = \sum_{n=1}^{N} \mathbf{E}_{raw,i}(t) e^{-\frac{j2\pi(\omega-1)(n-1)}{N}} \tag{4.3}$$

and for coils

$$\tilde{\mathbf{H}}_{\mathbf{i}}(\omega) = \sum_{n=1}^{N} \mathbf{H}_{i,V}(t) e^{-\frac{j2\pi(\omega-1)(n-1)}{N}} \tag{4.4}$$

The time series of Fourier coefficients is then scaled as:

$$\tilde{\mathbf{H}}_{\mathbf{i,scl}}(\omega) = \sqrt{\frac{2dt}{1.36N}}\frac{\tilde{\mathbf{H}}_i(\omega)}{0.54} \tag{4.5}$$

$$\tilde{\mathbf{E}}_{\mathbf{i,SI}}(\omega) = \sqrt{\frac{2dt}{1.36N}}\frac{\tilde{\mathbf{E}}_i(\omega)}{0.54} \tag{4.6}$$

Finally, the magnetic data are scaled by the appropriate BF-4 transfer function from the manufacturer's calibration records. This corresponds to point by point dividing the Fourier coefficients at each frequency by the calibration value at that frequency from the plot below:

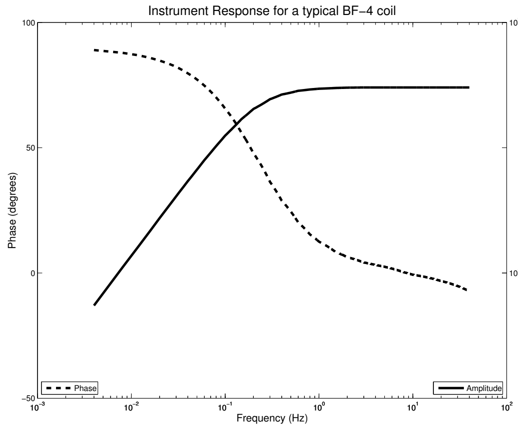


Figure 4.1: Manufacturer's calibration curves for coil BF4-9204

Referring to the instrument transfer function as ITF we can express the coil data in SI units by finally applying:

$$\tilde{\mathbf{H}}_{i,SI}(\omega) = \frac{\tilde{\mathbf{H}}(\omega)_{i,scl}}{ITF(\omega)} \tag{4.7}$$

Scaling the data like this results in the squared amplitude of the FCs corresponding to a true SI unit one-sided periodogram, with units of nT^2/Hz, and $(mV/km)^2/Hz$. The scale factors in equations 4.6 and 4.5 of 1.36 and 0.54 correspond to normalizing the data to account for spectral leakage in equivalent noise bandwidth and processing gain, respectively (*Harris* 1978). The $\sqrt{2}$ Parseval normalization allows the one-sided power spectrum to reflect the total energy in the time series.

A few amplitude spectra were estimated using these calibration factors and compared to calculations made by industry-standard codes used at Schlumberger [*Cuevas*, personal communication]. The comparison shown in Figure 4.2 reveals a reasonably good agreement. Note the two spectral peaks centered at approximately 8 and 14 Hz. These are the first and second Schumann resonances. These phenomena are attributed to standing waves set up in the spherically concentric capacitor whose conductive boundaries are the ionosphere and the solid earth. The source of the resonance are the bolts of intracloud and cloud-to-ground lightning continuously crackling at an average rate of between 40 and 50 bolts per second [*Christian et al.* 2003].

Because the instruments are swapped out from time to time, and because previous analysis of our data has shown orders-of-magnitude offsets coincident with these swaps (ex. Figure 4.3), an exercise is performed to determine how well the array observations compare to one another over the long term using the codes created for this study. By selecting a two-hour block of data at the same time each night (2-4am PST), 40Hz magnetic data are downloaded

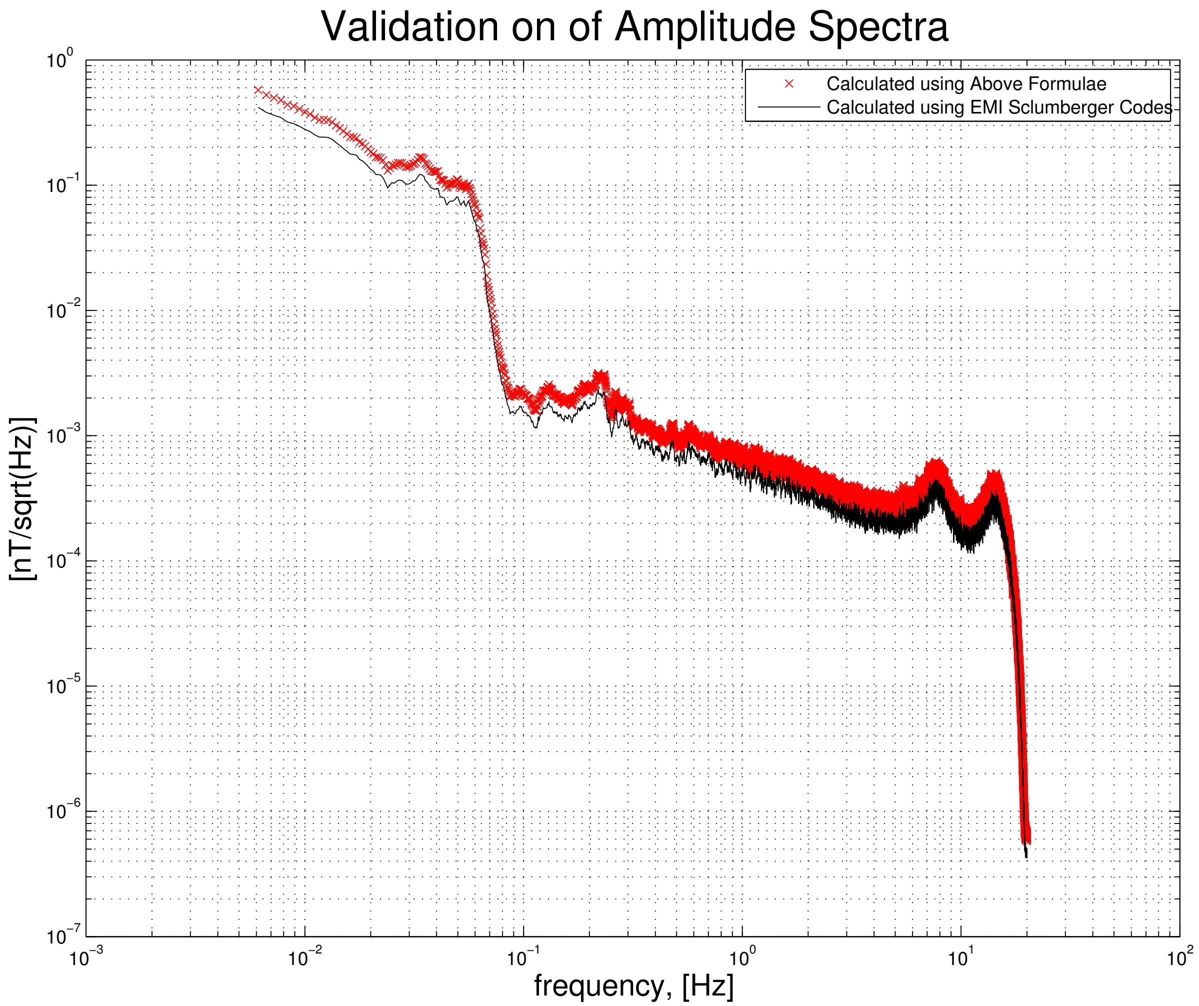


Figure 4.2: High resolution amplitude spectra for day 272, 2004. Two hours of data from 0000-0200hrs UT calculated using the above method (red) and using an industry standard code (black).

for the days when this two-hour window was not flagged by the despiker. The magnetic channels are then transformed according to Equation 4.5. The first Schumann amplitude is calculated by averaging the amplitudes of the Fourier coefficients between [7.413 and 8.2224] Hz. Long term plots of the stability of these indices are shown in Figure 4.4

The time series is fairly stable and the values are in agreement with the theoretical values predicted by *Heckman et al.* (around 0.2(pT)^2/Hz for the NS component (Hy) or $\approx$ 0.45 pT/$\sqrt{Hz}$). There is a clear seasonal variability with peaks in the summer months (JJ on the Figure axes) and lows in the winter months (DJ on Figure axes). This reflects the total amount of thunderstorm activity cycling through the year. Once the Fourier coefficients have been calculated and stored for each decimation level, the data are ready to be processed by a variant on the RMEV method proposed by Egbert (1997, 2002).

Figure 4.5 shows that the offsets in the field ratios are likely related to changes in site configuration. No maintenance records are available after March 2004. These plots show that changes in the absolute field values are reliable only to within around 10%. In order to see finer-scale variability, we express the data in terms of signal to noise ratio in the next section.

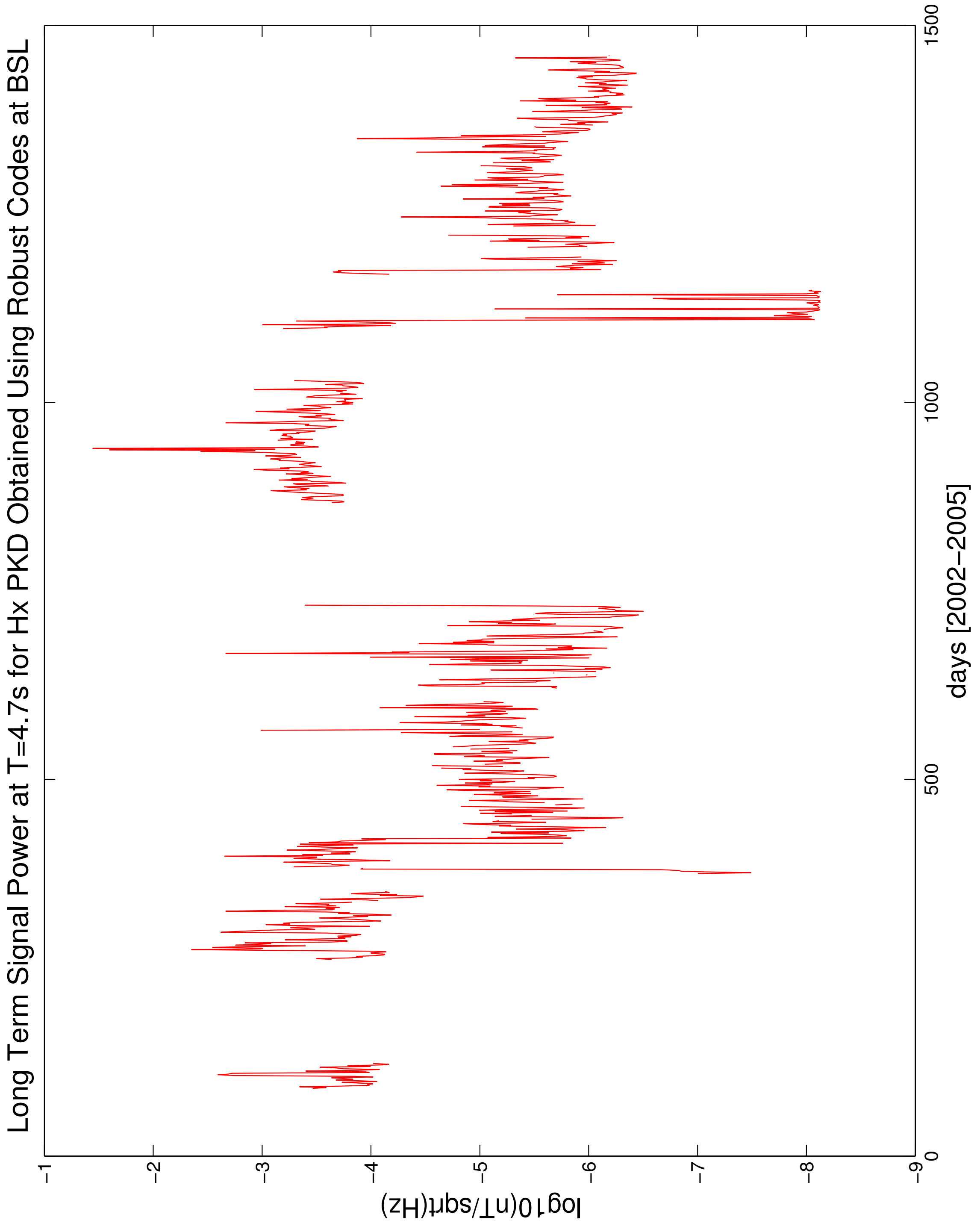


Figure 4.3: Long-term magnetic field amplitudes for 4.7s period data calculated in 2005 using standard software at the BSL. Logarithm of amplitude is shown because the large dynamic range renders certain tracts of data invisible on a linear scale.

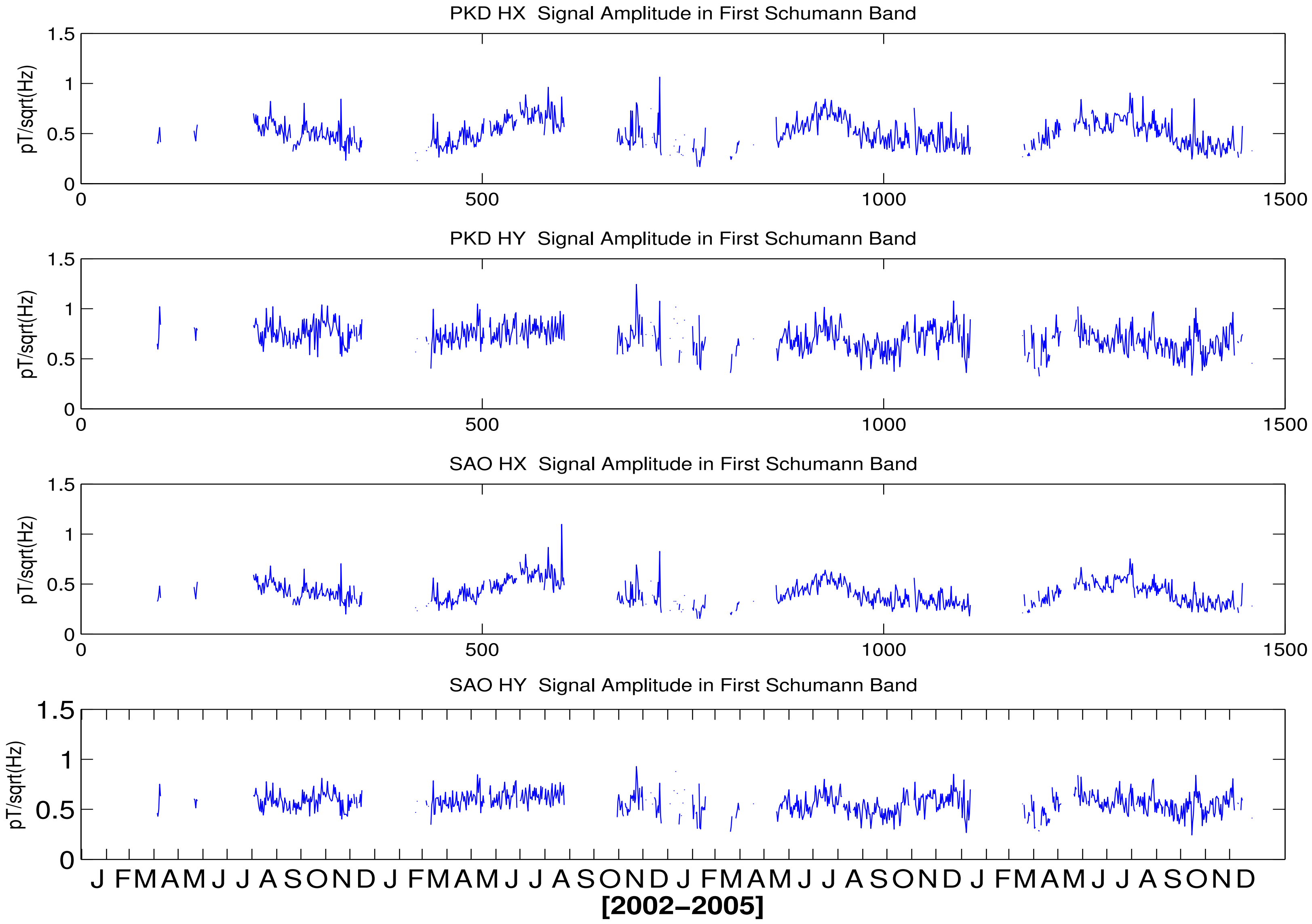
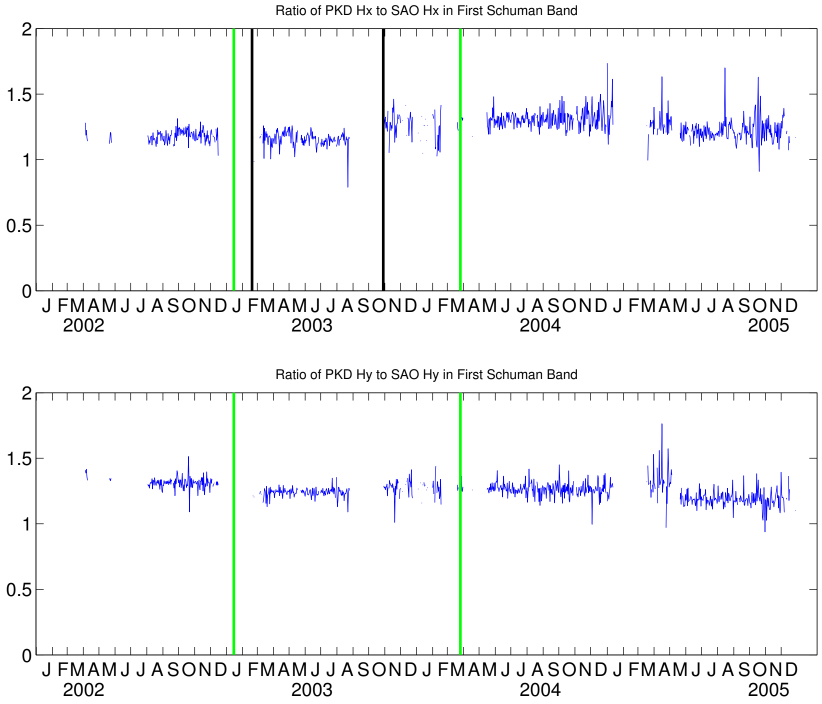


Figure 4.4: Seasonal variability of the Schumann resonances


Ratio of PKD Hx to SAO Hx in First Schuman Band

Ratio of PKD Hy to SAO Hy in First Schuman Band

Figure 4.5: Intersite horizontal magnetic field amplitude ratios. Green vertical lines represent times when BFPS were exchanged, and the black lines indicate when Hx Coils were replaced at SAO.

4.2 Robust estimation of the SDM

Following is a description of the RMEV code implemented to process the spectral domain data. The raw data for this method are the daily generated vector-valued time series of Fourier coefficients, with one row for each channel, and one time index for each partially overlapping 256 point window. Every frequency bin within each decimation level is processed independently for each day. Thus, we choose an arbitrary but fixed combination of these three parameters (day, decimation level, and frequency bin) and begin with a vector valued time series $\mathbf{X}(t)$ where $\mathbf{X} \in \mathbf{M}(\mathbb{C})_{\mathbf{8,T}}$. T is the number of windows in the decimation level. For a detailed description of robust estimation of the covariance matrix (hereafter referred to as the spectral density matrix or SDM), the reader is referred to *Egbert* (1997), and references therein. The SDM estimate is an iterative procedure of down-weighting outliers in the FC time series. Outliers are alternately downweighted in all channels, and then in individual channels. Outliers in all channels are treated by the affinely invariant approach Huber (1981), which essentially involves sphering the data, followed by downweighting observations (FC vectors corresponding to a single time window) which defy a Gaussian, spherically symmetric statistical model by more than some predefined number of standard deviations. After downweighting outlier in all channels equally, the method then re-estimates the data vectors by downweighting outliers in individual channels. This is done by approximating each channel's FC time series by a least squares fit to all other array channels, and then downweighting the FCs in time bins which have residuals greater than some number of standard deviations. These residual variances serve as estimates of channel-noise variances, although they are imperfect estimates, as they also incorporate noise from other channels. This contamination is reduced by an updating step which decouples the mixed incoherent channel noises from one another.

The SDM together with these estimates of incoherent noise serve as our basic dataset for a discussion of signal and incoherent noise. The structure and properties of the SDM will be discussed later. For now it is sufficient to note that with outlier-free data, the SDM is simply the expected value of the outer product of the data vectors, i.e.: The SDM **S** is calculated from **X** as

$$\mathbf{S} = \frac{XX^*}{T} \tag{4.8}$$

and as a trivial corollary, the ith diagonal element is simply the variance of the ith channel, or the autopower of the ith sensor.

STEP 1.

Downweight outliers equally in all directions. The data matrix **X** is preconditioned by the transformation

$$\mathbf{Y} = \mathbf{DUX} \tag{4.9}$$

where the rows of **Y** are uncorrelated, and of unit variance. Here **U** is defined by the eigenvalue decomposition of $\mathbf{S}=\mathbf{U}\Lambda\mathbf{U}^*$ and $\mathbf{D}$ is the diagonal matrix whose non-zero elements are the square roots of the non zero elements of Λ. i.e.:

$$D_{i,i} = \frac{1}{\sqrt{\Lambda_{i,i}}} \tag{4.10}$$

Converting $\mathbf{Y}$ into a single time series $\mathbf{R}$ whose elements are the norms of the columns of $\mathbf{Y}$, the elements of $\mathbf{R}$ which deviate by more than some fixed number of standard deviations from the median are downweighted by a factor of 0.5.

STEP 2.

An initial estimate of incoherent noise variance is determined by predicting each channel by all the others. To predict the ith row of $\mathbf{X}$ denoted by $X_{i,:}$ we employ the reduced matrix $X_{\sim i,:}$ where the $\sim$i denotes removal of the ith row. Thus by basic regression we obtain

$$\tilde{\mathbf{X}}_{\mathbf{i},:} = \mathbf{G}\mathbf{X}_{\sim\mathbf{i},:} \tag{4.11}$$

where

$$\mathbf{G} = (\mathbf{X}_{\mathbf{i},:})(\mathbf{X}_{\sim\mathbf{i},:})(\mathbf{X}_{\sim\mathbf{i},:}\mathbf{X}^*_{\sim\mathbf{i},:})^{-1} \tag{4.12}$$

and $\tilde{\mathbf{X}}$ is the predicted time series. The residual time series for the ith channel is simply then $\mathbf{X}_{i,:}$- $\tilde{\mathbf{X}}_{i,:}$. The RMS of this residual time series is taken as an initial estimate of the incoherent noise variance of the ith channel. When residual time series elements deviate by more than 3 form the 0.01 α-trimmed standard deviations, the data in channel i are downweighted by a factor 0.7

STEP 3.

The incoherent noise variances are decoupled. Following the work of Egbert (1997), we note that even if some channels were completely noise-free, the residual as defined above would contain some noise, because the predicted time series is a linear combination of other channels which do have some noise in them. Extending this argument to the case where a channel has some noise, it is clear that by estimating the noise by the residual RMS as in step two, one will consistently overestimate the noise. Because each channel's contribution to the residual noise is incoherent with all other channels, it will not cancel when summed, but rather, will grow as in a random walk.

Formally, consider a channel X$_i$(t), approximated as a linear combination of N-1 other channels, where the jth channel is represented as X$_j$+ε_j, where ε_j is the 'true incoherent noise' in the jth predictor channel. When SNR>2, we can approximate the residual variance $\mathbf{r}_i$ as a linear combination of the true channel noise variance plus the incoherent noise variances of the predictor channels. The coefficient controlling the amount of the jth channel's variance added to that of the ith channel is the square of the transfer function element linking the channels. More formally:

$$\mathbf{r_i} \approx \sigma_i^2 + \sum_{j=1}^{N} G_{i,j}\sigma_j^2 \tag{4.13}$$

Of course, by predicting each of the N channels with the N-1 others we obtain N equations of the form of 4.13 in N unknowns— a linearly solvable system. After Egbert (1997), if **r** is the residual variance vector, whose ith element is the RMS res of the ith channel, we solve:

$$\mathbf{r} = (\mathbf{I} + \mathbf{B})\sigma^2 \tag{4.14}$$

where B$_{k,l}$=G$_{k,l}$ (the coefficient in front of the lth predictor channel when fitting channel k). We apply a technique of regularization to solving this system to account for the possibility that negative, or very small variances could be returned by the linear system. The regularization is basically a constraint which prohibits the incoherent noise estimates from becoming smaller than 20% of their original estimate, and is enacted by repeatedly solving the modified version of 4.14:

$$\mathbf{r} = (\mathbf{I} + \mu\mathbf{B})\sigma^2 \tag{4.15}$$

trying at first to solve 4.15 with μ=1, but successively reducing μ if the ’twenty-percent’ condition is violated.

STEP 4. Repeat Step 1, then repeat Step 2, but estimate incoherent noise variances differently. Transform the data into units of signal to noise using the noise values from Step 3. Find the number of eigenvalues M of the scaled SDM significantly greater than 1 (here, to be conservative, we use strictly greater than 1). Approximate the ith channel by projecting $X_{\sim i,:}$ onto these M principal components. This is done by decomposing $\tilde{S}$=$\mathbf{U}\Lambda\mathbf{U}^*$, where $\tilde{S}$ represents the scaled SDM with the ith row and column deleted. Then, with the columns of $\mathbf{U}$ ordered with their descending singular values, select the submatrix $\tilde{U}$=$\mathbf{U}_{(:,1:M)}$ of the M dominant eigenvectors. Now project $X_{\sim i,:}$ onto these M principal components by multiplying $\tilde{U}X_{\sim i,:}$. Thus the time series has been compressed into an M-dimensional vector valued time series, where the vector elements are now the coefficients weighting each principle component. The regression is the same as in Step 1 except now $X_{\sim i,:}$ is replaced with $\tilde{U}$. Thus regression is performed on a channel's projection onto the M dominant principle components, rather than on the channel data itself.

STEP 5: Iterate 3 and 4 to convergence.

Once the RMEV algorithm has converged for each FC, the average SDM at the FC is calculated, and then the band averaged SDM is calculated over all SDMs for FCs in a band. Band averages are calculated directly with no taper.

4.3 A global reference for geomagnetic activity

Having described the robust SDM is described, we define the signal power in channel i as the i^{th} diagonal element of the robustly estimated SDM minus the estimate of incoherent noise variance. We define the average daily signal amplitude as the positive square root of the average daily signal power. Noise is defined as the effective instrument noise level, or incoherent noise. That is, noise which is uncorrelated between all sensors. We use the Daniell-window band average of the incoherent noise variances returned by the last iteration of the RMEV scheme.

We can examine the median values of signal and noise for each orientation of sensor and each type of sensor in Figure 4.6.

These exhibit what we expect: A 1/f noise level, and a signal level showing a significant drop as it enters the 'dead' band (0.1-1.0 Hz). Note that the amplitude of the electric fields at SAO is significantly larger than that at PKD. This suggests that the ground there is much more resistive than at PKD, which is consistent with the site geologies. Before tracking the dynamic behaviour of signal and noise, it is helpful to first consider the signal source process of the MT field. The micropulsations in the Earth's magnetic field have been monitored since the time of Gauss, and there is a worldwide network of observatories which contribute observations to a database at Potsdam which publishes summary indices known as geomagnetic activity indices. There are several scales on which geomagnetic

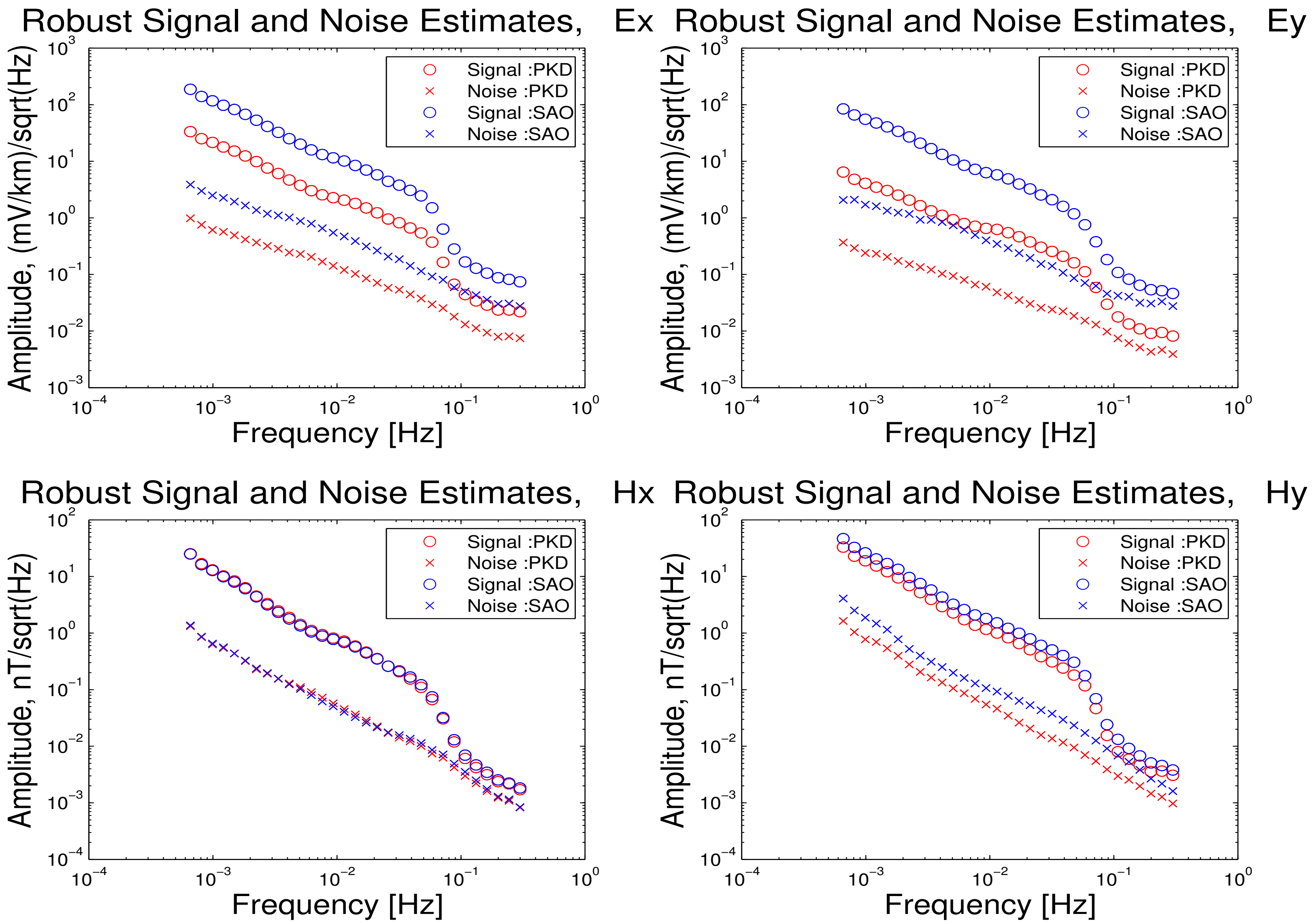


Figure 4.6: Signal and noise median amplitudes as a function of frequency calculated over the whole four year interval.

activity is measured. These indices are described by the NOAA website as follows: *“ The K-index is quasi-logarithmic local index of the 3-hourly range in magnetic activity relative to an assumed quiet-day curve for a single geomagnetic observatory site. First introduced by J. Bartels in 1938, it consists of a single-digit 0 thru 9 for each 3-hour interval of the universal time day (UT). The planetary 3-hour-range index Kp is the mean standardized K-index from 13 geomagnetic observatories between 44 degrees and 60 degrees northern or southern geomagnetic latitude. The scale is 0 to 9 expressed in thirds of a unit. This planetary index is designed to measure solar particle radiation by its magnetic effects. The 3-hourly Ap (equivalent range) index is derived from the Kp index as follows:”*

Kp=	0o	0+	1-	1o	1+	2-	2o	2+	3-	3o	3+	4-	4o	4+	5-	5o	5+	6-	6o	6+	7-	7o	7+	8-	8o	8+	9-	9o	9+
Ap=	0	2	3	4	5	6	7	9	12	15	18	22	27	32	39	48	56	67	80	94	111	132	154	179	207	236	300	400	

Table 4.1: Correspondence between Ap and Kp indices.

We expect variations in the intensity of global magnetospheric activity to relate to the signal power observed by the array. Plotted in Figure 4.7 are the A_p indices for the four-years 2002-2005.

4.4 Variation in signal and noise over four years

Dynamic behaviour of signal, noise and their ratios are shown in Figures 4.8- 4.13. There are several remarkable features about this collection of plots. Foremost, in all six plots, note the generally stable nature of the fields. Geomagnetic activity is typically stable, and

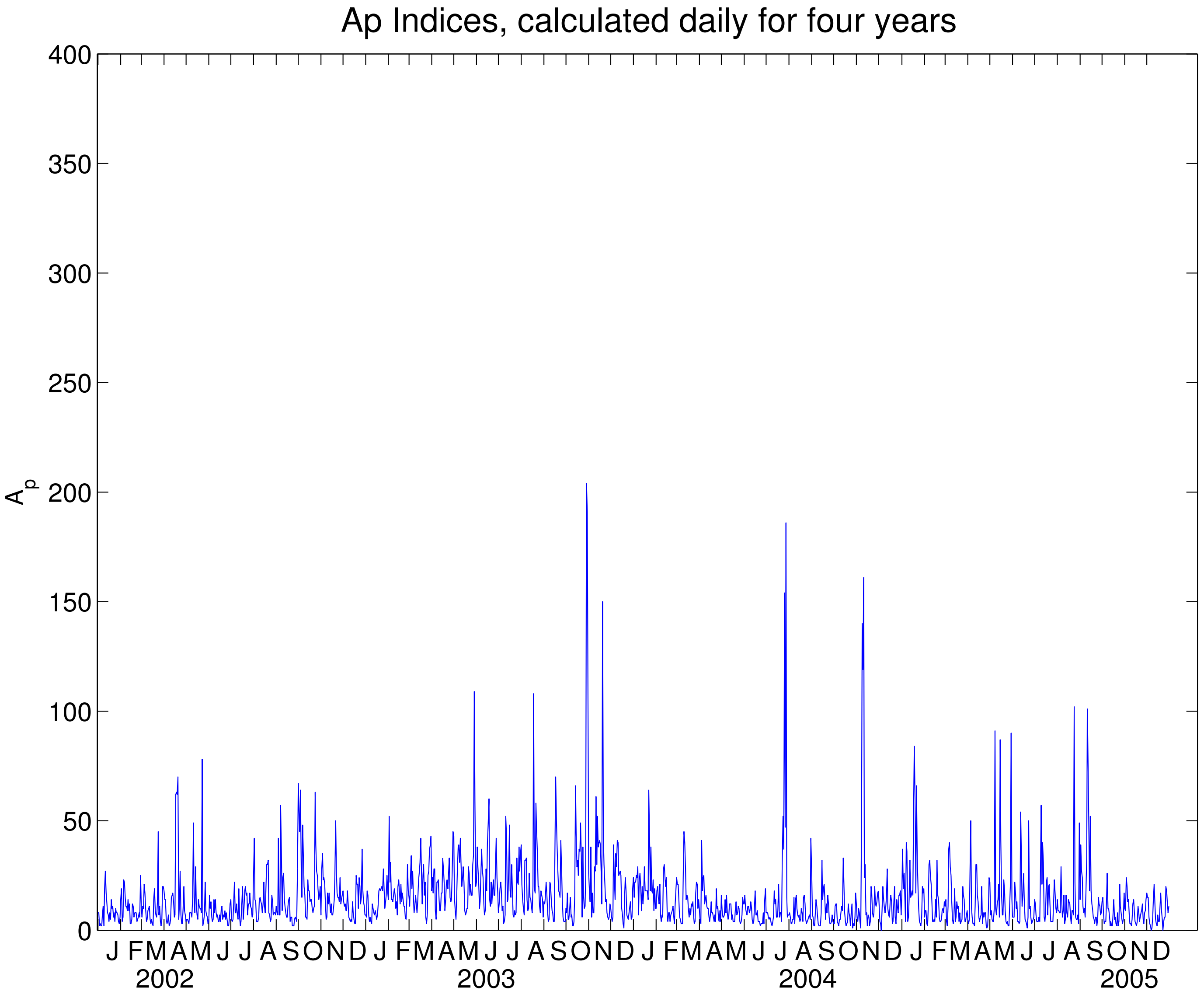


Figure 4.7: Ap indices over the 2002-2005 time interval.

the instrument noise levels are stationary from day to day. There is an offset in the instrument noise levels at September 20^{th} 2002, particularly noticeable in the electric channels (Figure 4.8). This date corresponds with a visit to both observatories during which routine maintenance was performed. Specifically, the electrodes were removed and a mixture of bentonite and salt water was added to the holes where they were re-embedded. This has the effect of improving local coupling to the earth.

Signal plots, especially the magnetics, show several broadband pulses, narrow in time, where signal levels were suddenly, unusually large. For example, these can be seen in July and November of 2004. The pulses are present at both sites and hence are not local phenomena. The pulses correspond to days of anomalously high geomagnetic activity. NOAA classifies days where the A_p index is greater than 49 as major geomagnetic storms, and Ap $>$29 as minor geomagnetic storms. Figure 4.14 shows magnetic signal power with storm days removed. Note that the broadband pulses are not present in these plots.

4.5 Variation of signal and noise over one half year

The SNR during the block of data around the earthquake was not as good as in some other years, but still reflects a reasonable level of array fidelity. The window of primary interest is the 163-day time from day 137, 2004 to day 299, 2004. This window is significant because:

1) It extends for more than four months prior to the 2004 Parkfield M 6.0 earthquake, a

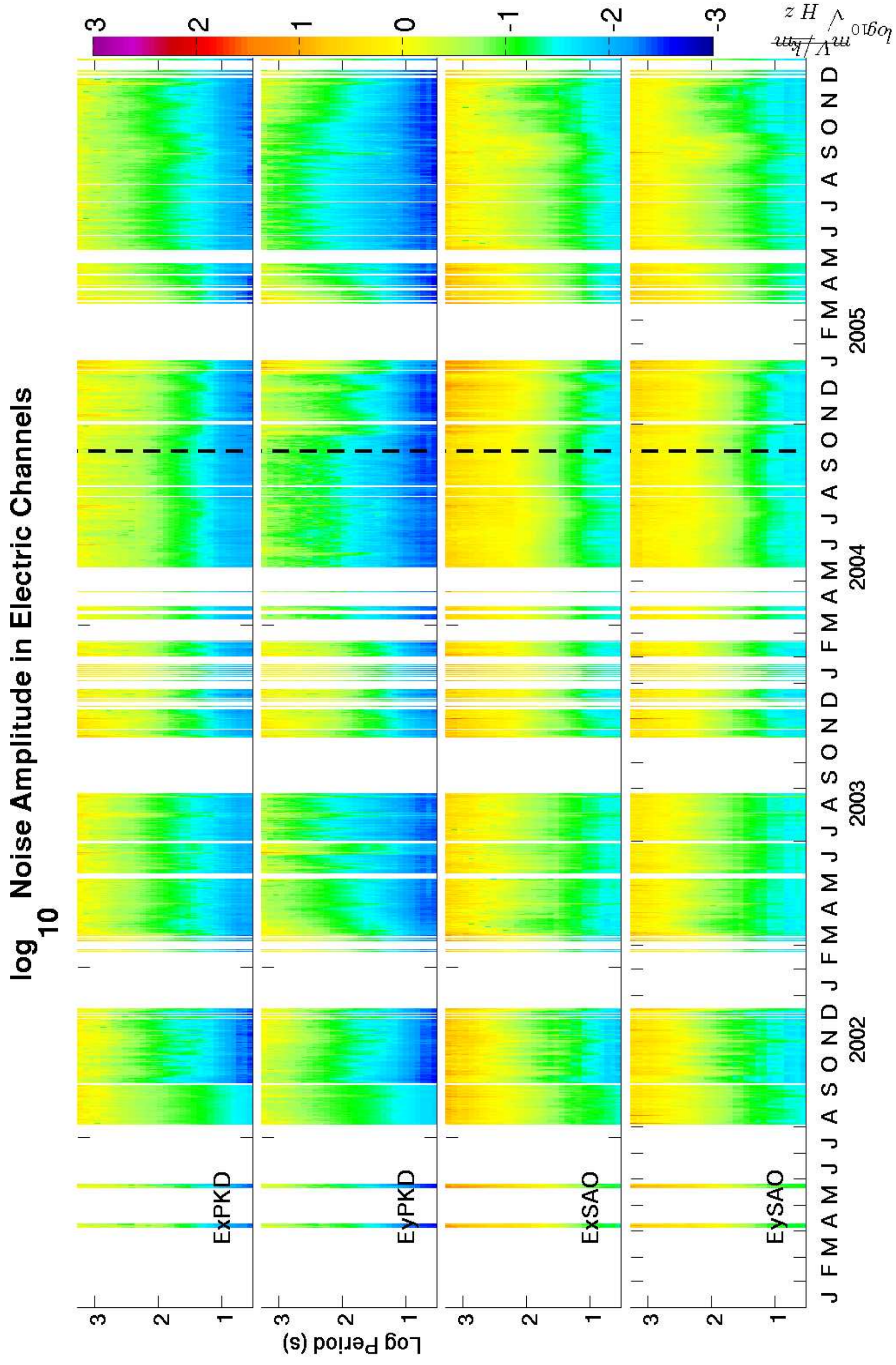


Figure 4.8: Noise amplitudes in $\log_{10}(\mathrm{mV/km}/\sqrt{Hz})$.

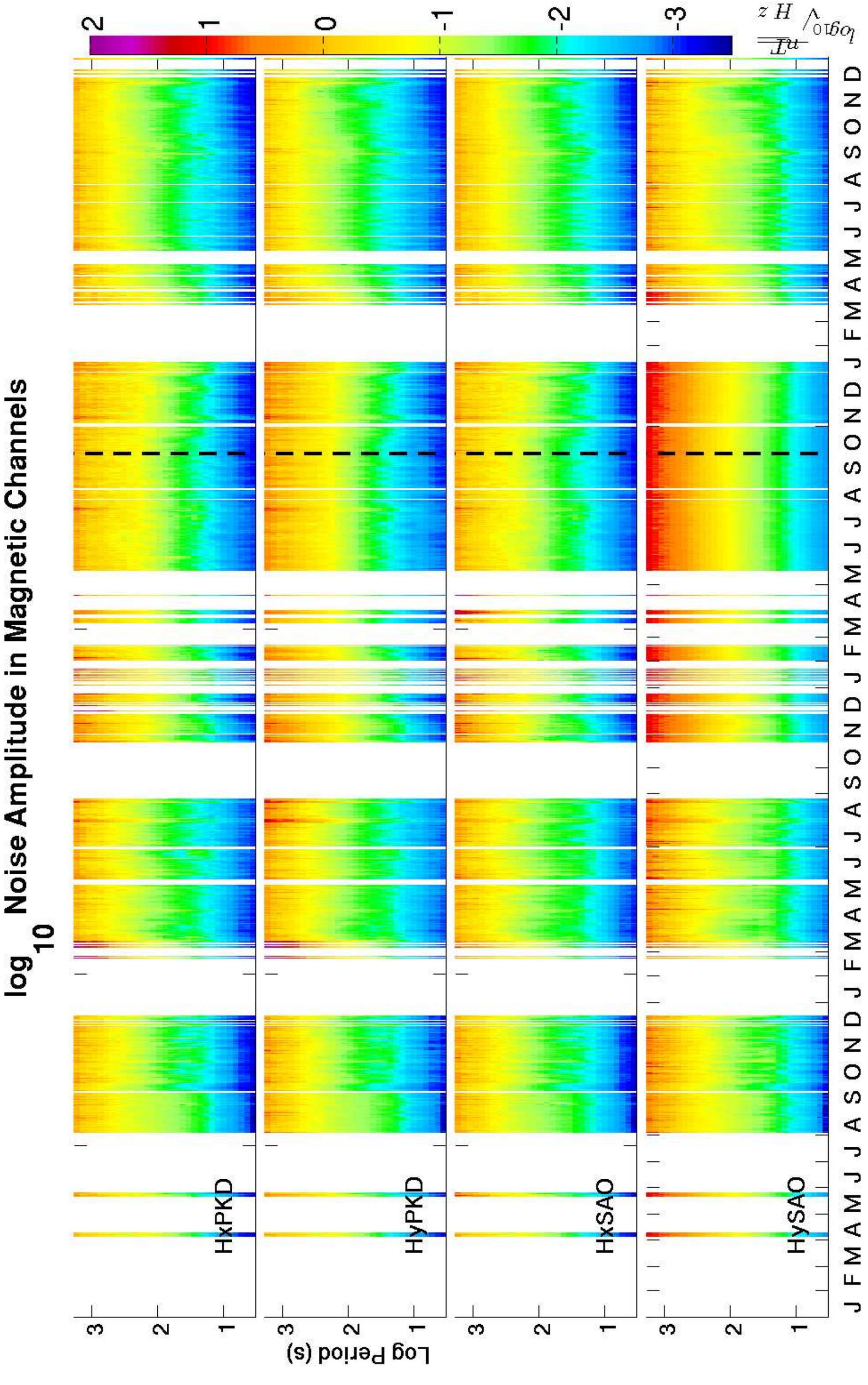


Figure 4.9: Noise amplitudes in $\log_{10}$(nT/$\sqrt{Hz}$)

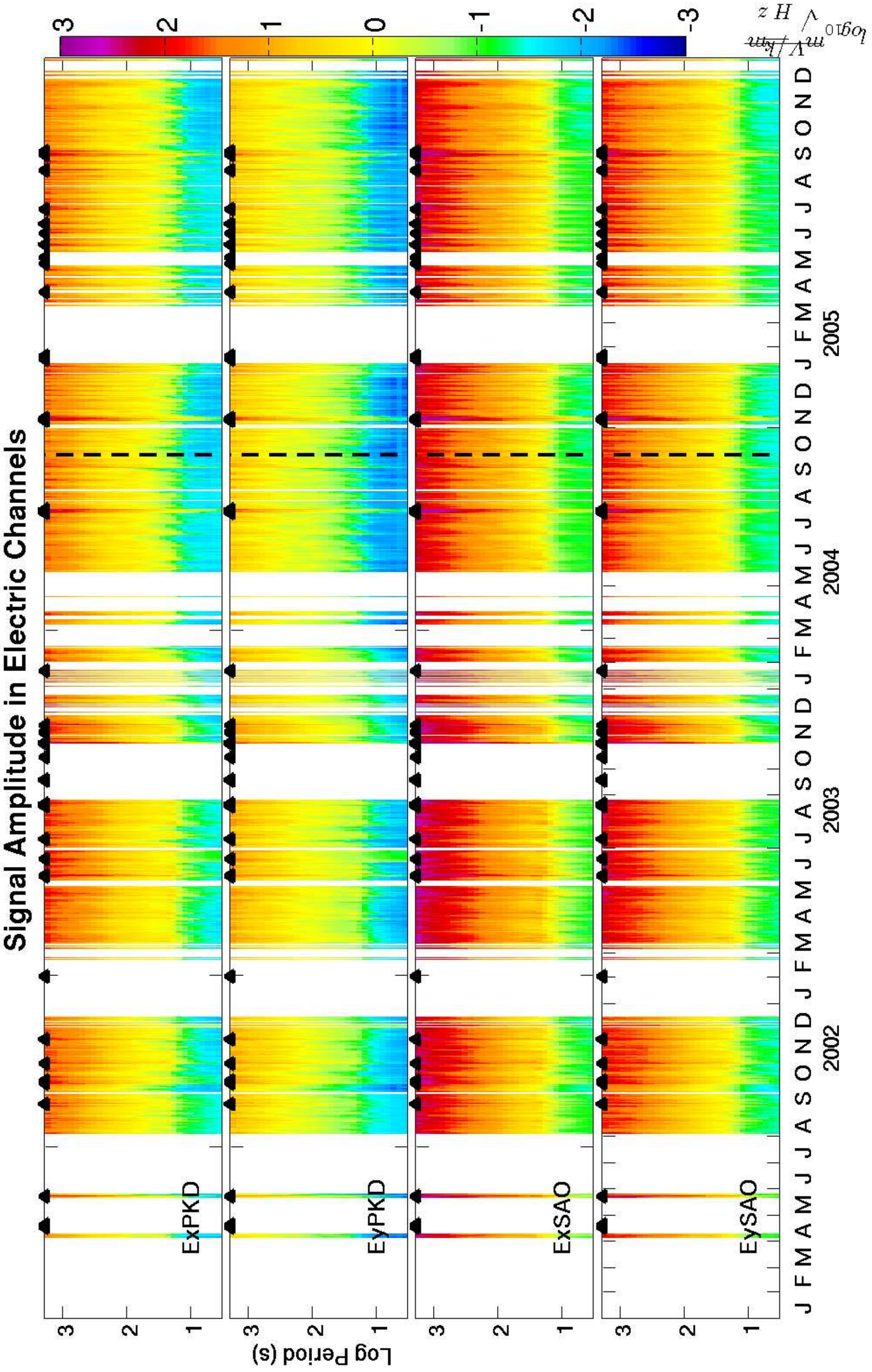


Figure 4.10: Signal Amplitudes in $\log_{10}(\mathrm{mV/km}/\sqrt{Hz})$. The colour scale is kept the same as in Figure 4.8. Black triangles mark days of major solar storms.

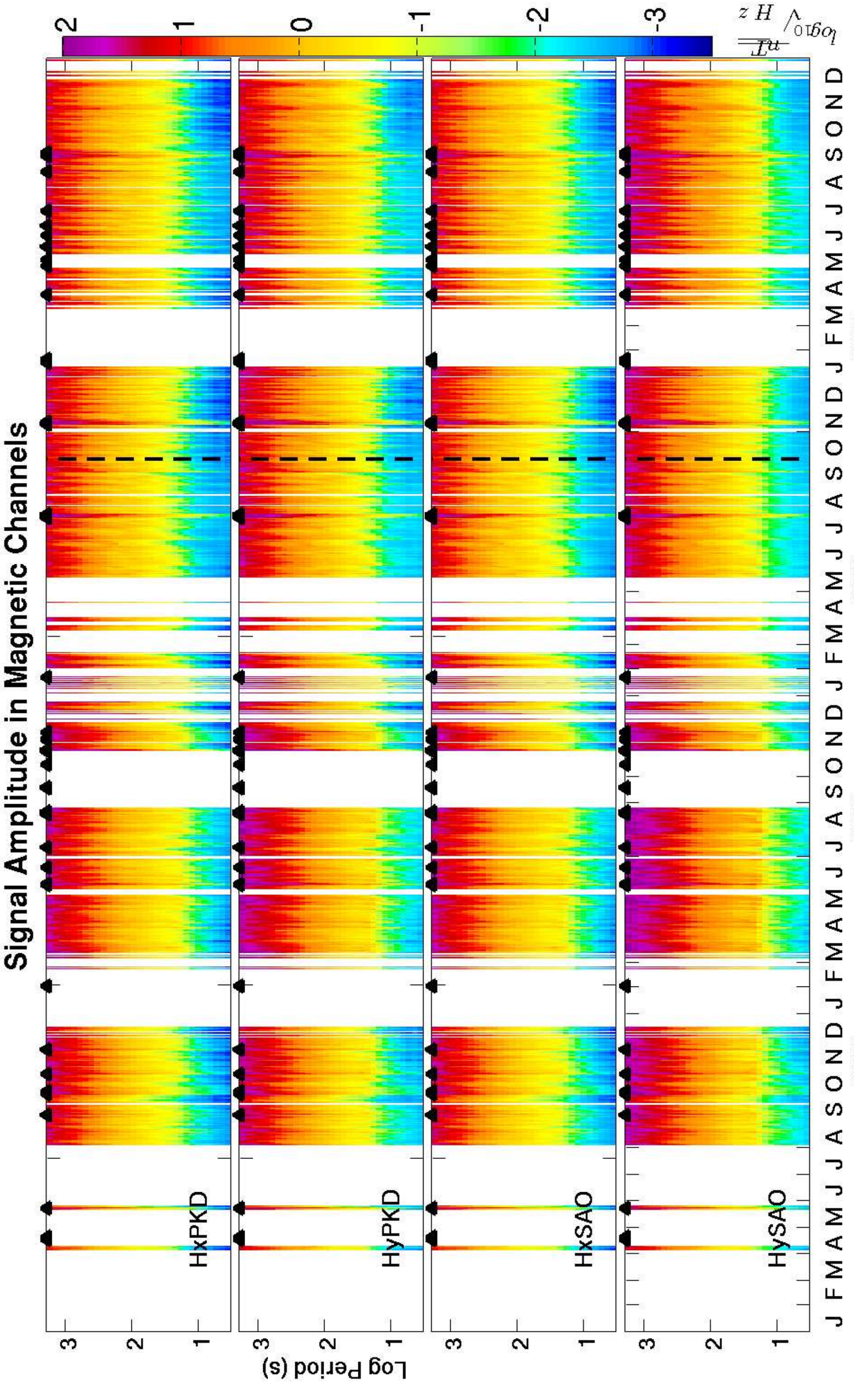


Figure 4.11: Signal Amplitudes in $\log_{10}(\mathrm{nT}/\sqrt{Hz})$. The colour scale is kept the same as in Figure 4.9. Black triangles mark days of major solar storms.

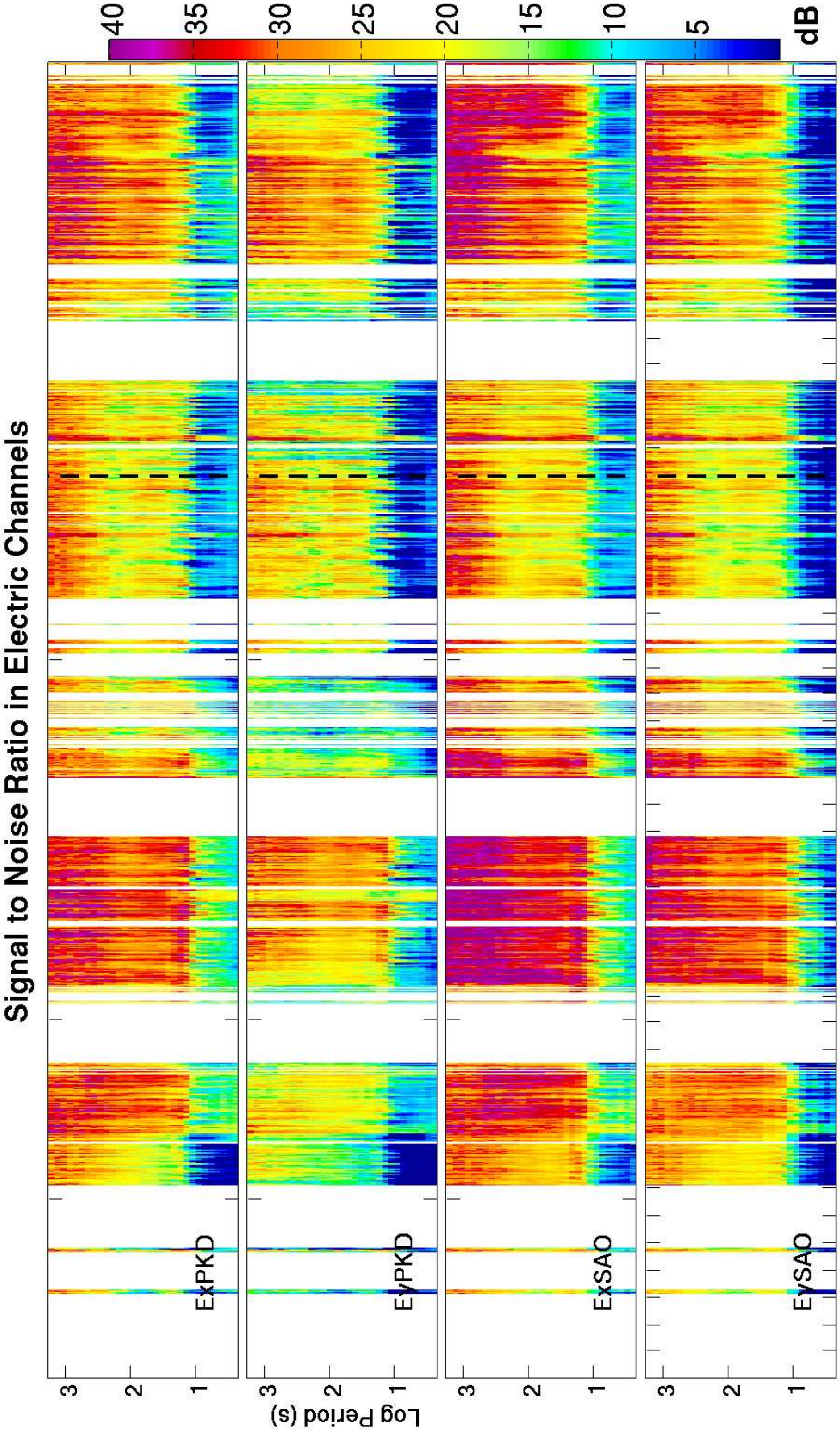


Figure 4.12: Signal to noise ratio in dB for electric channels.

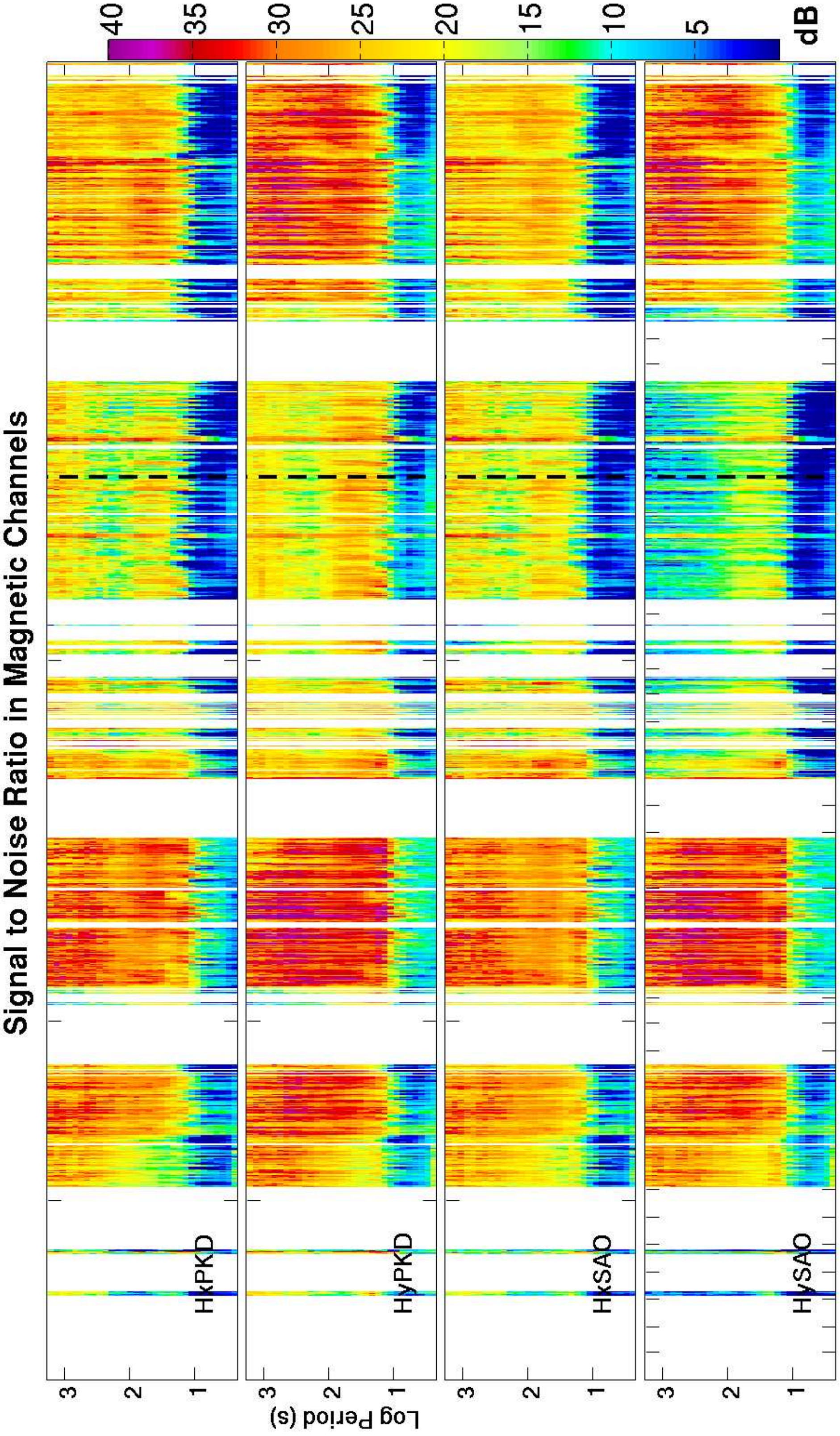


Figure 4.13: Signal to noise ratio in dB for magnetic channels.

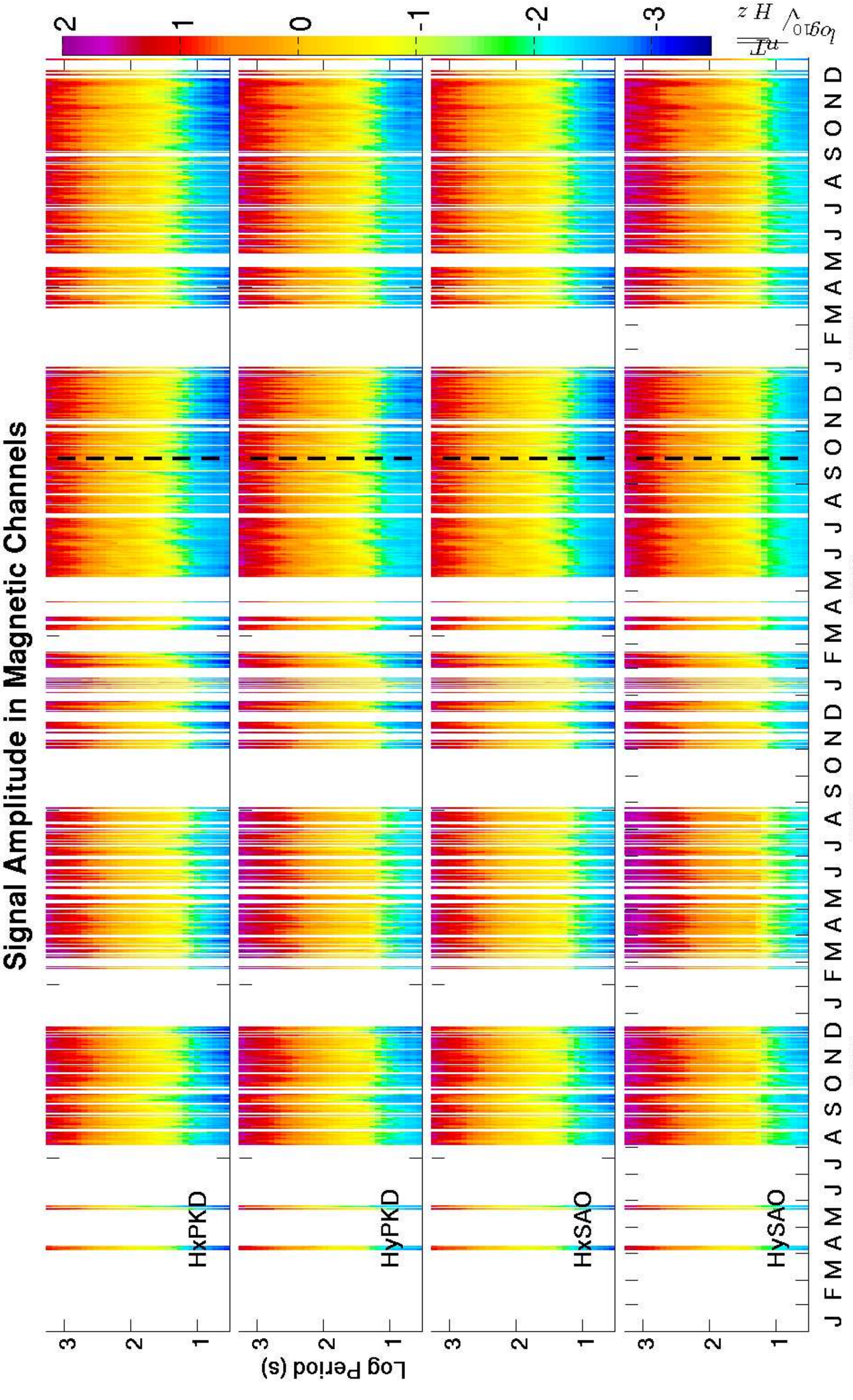


Figure 4.14: Magnetic signal power with storms removed.

longer period than has ever been treated by continuous ULF monitoring before an earthquake;

2) It is free of introduced artifacts in the data resulting from site maintenance, swapping of equipment, or equipment malfunction.

For this time window, we display noise (Figures 4.15 and 4.16) and signal (Figures 4.17 and 4.18) in terms of amplitude spectra and signal to noise ratio in units of dB (Figures 4.19-4.20).

Figures 4.15-4.20 show a few noteworthy features. The statements from the four-year plots about day to day stability apply. The signal power is clearly linked to solar storm activity. The noise power is mostly stationary. There is a noticeable outlier on day 162 where Ex at PKD suffers from increased noise levels (Figure 4.15), and a corresponding degradation in SNR (Figure 4.19). An inspection of Figure 4.17 shows no corresponding spike in signal power. Also, there is no corresponding spike in any of the magnetic field plots. This implies that the energy is solely in the electric field and is broadband—a highly unlikely scenario.

Inspection of the time series shows that this outlier is an artifact of spikes in the electric field measurements. Figure 4.21 shows the raw data and the despiked data. Although the despiked data are not as corrupt as the raw data, they still contain many smaller spikes. Since this behaviour is present in both electrodes, it could result from problems at the common electrode or within the EFSC itself. It is unlikely that these fields are natural,

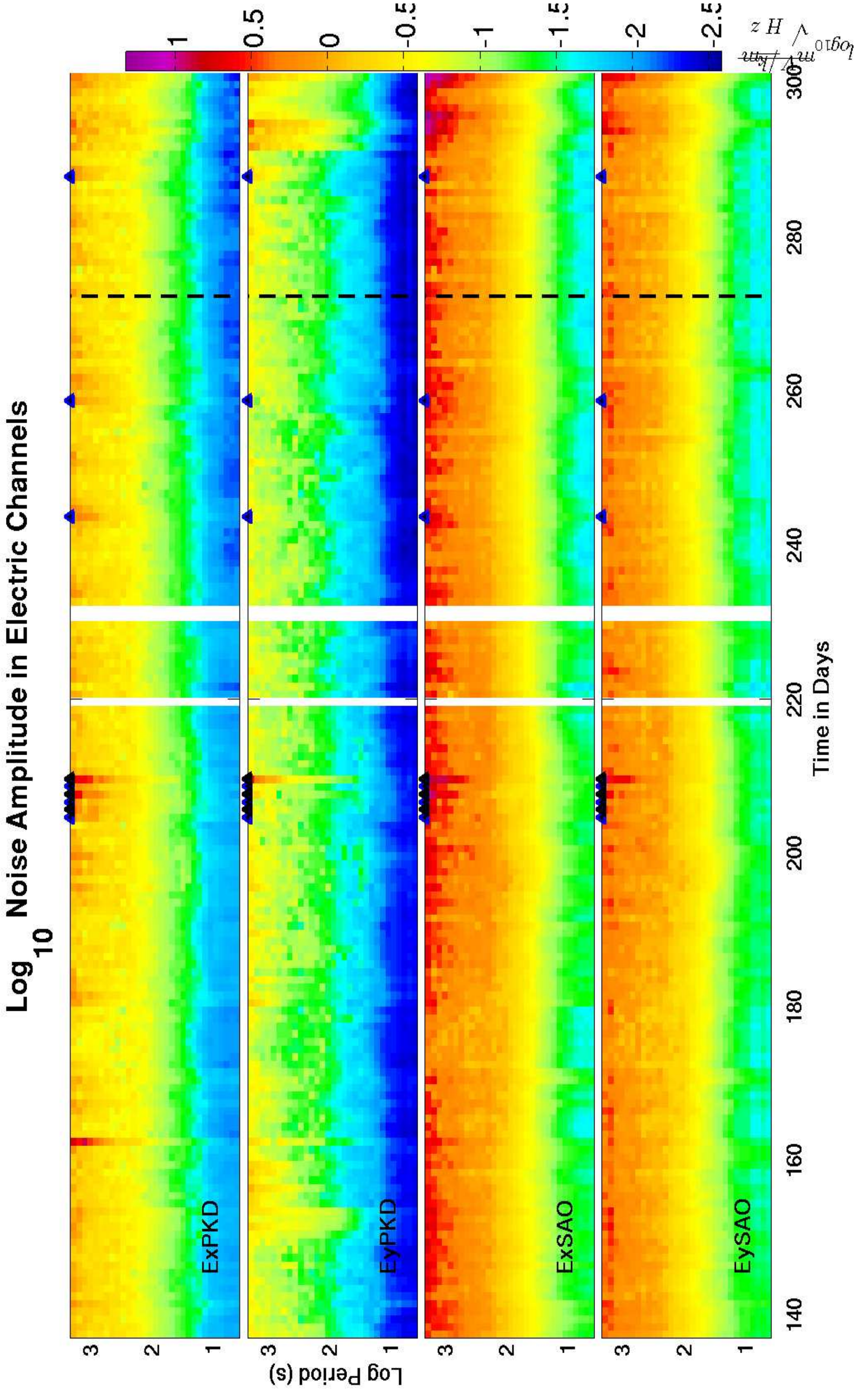


Figure 4.15: Noise amplitudes in $\log_{10}$(mV/km/$\sqrt{Hz}$). Black triangles mark days of major solar storms, and blue triangles mark minor storms.

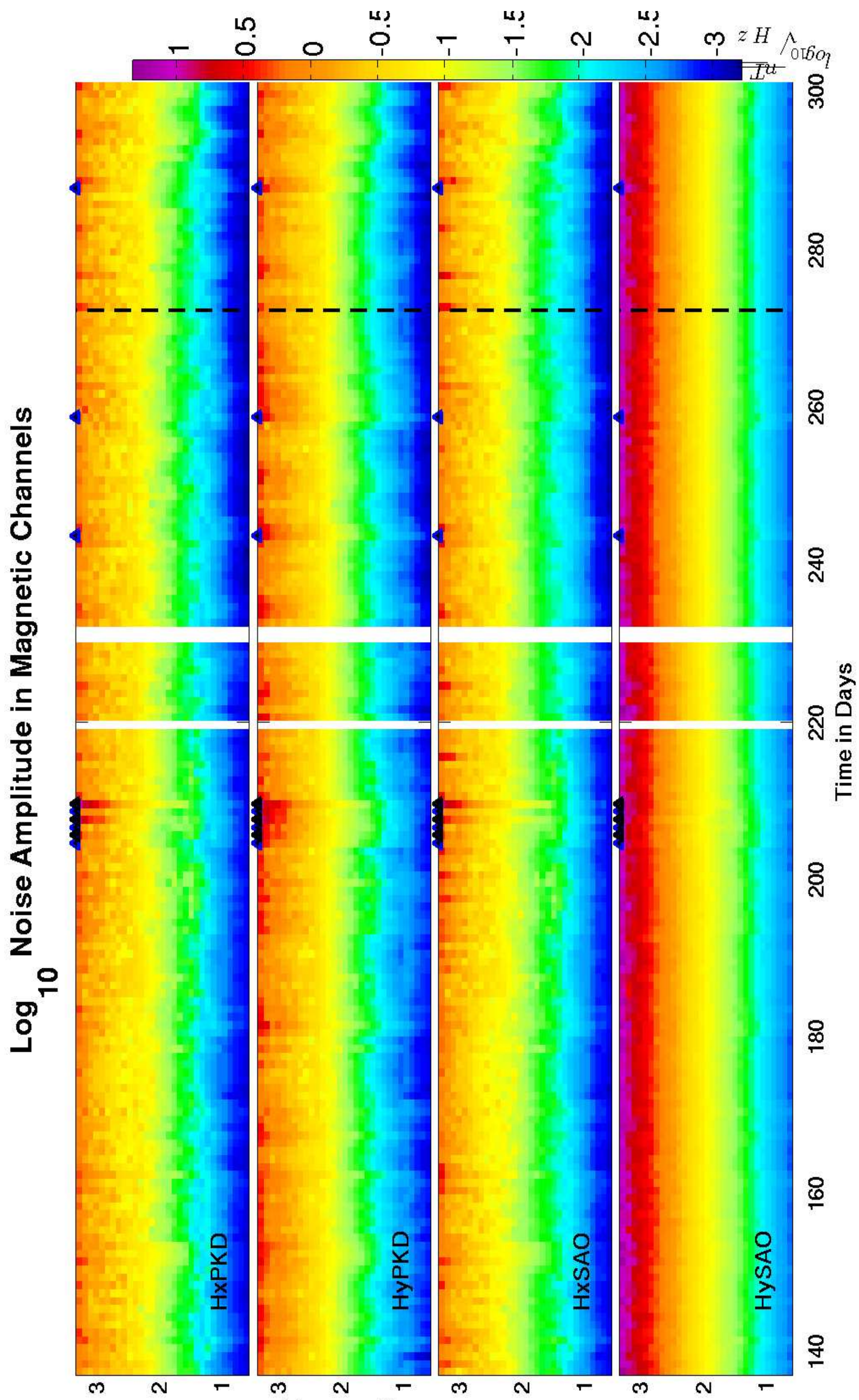


Figure 4.16: Noise amplitudes in $\log_{10}(\text{nT}/\sqrt{Hz}$. Black triangles mark days of major solar storms, and blue triangles mark minor storms.

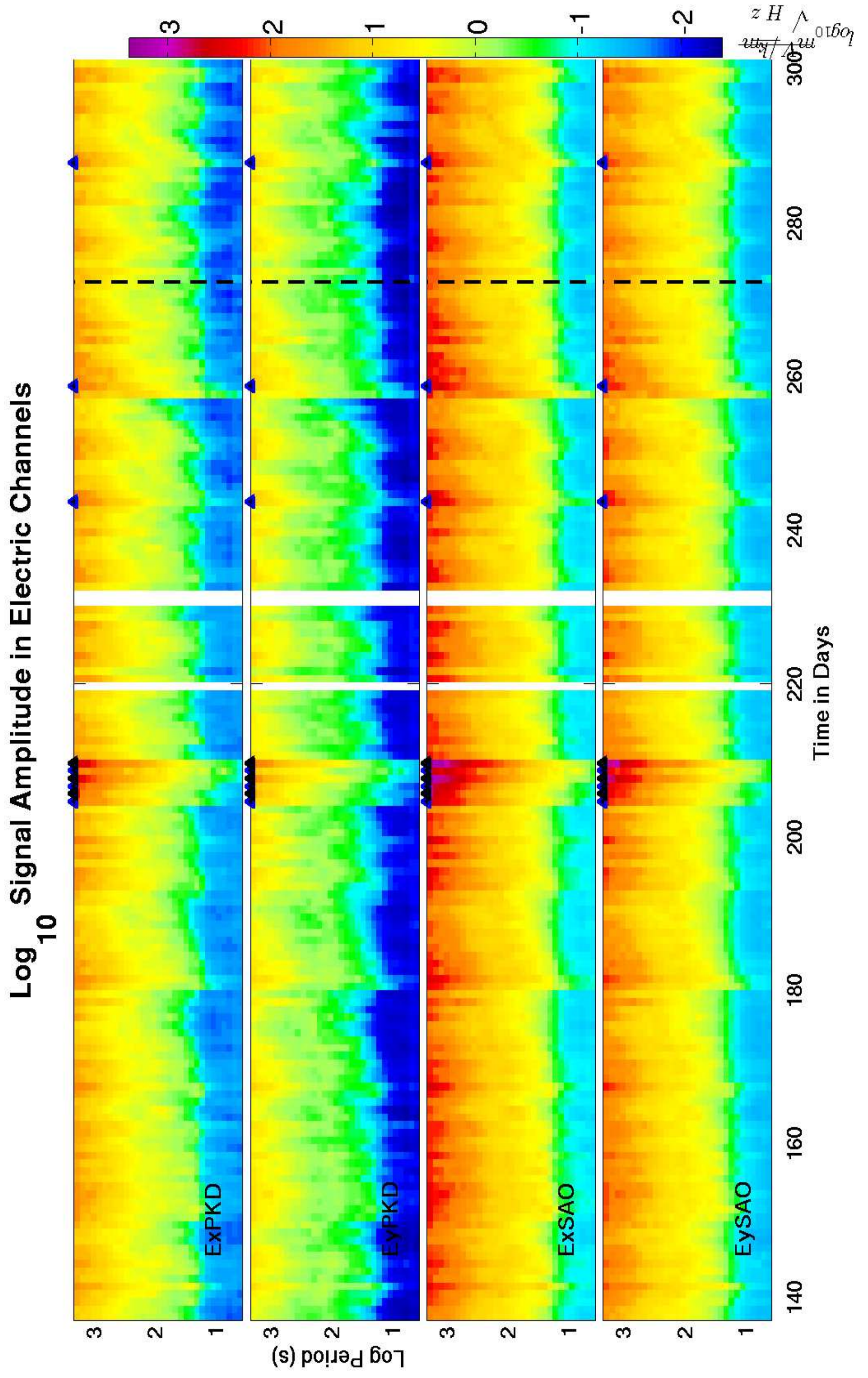


Figure 4.17: Signal amplitudes in $\log_{10}(\text{mV/km}/\sqrt{Hz})$. Black triangles mark days of major solar storms, and blue triangles mark minor storms.

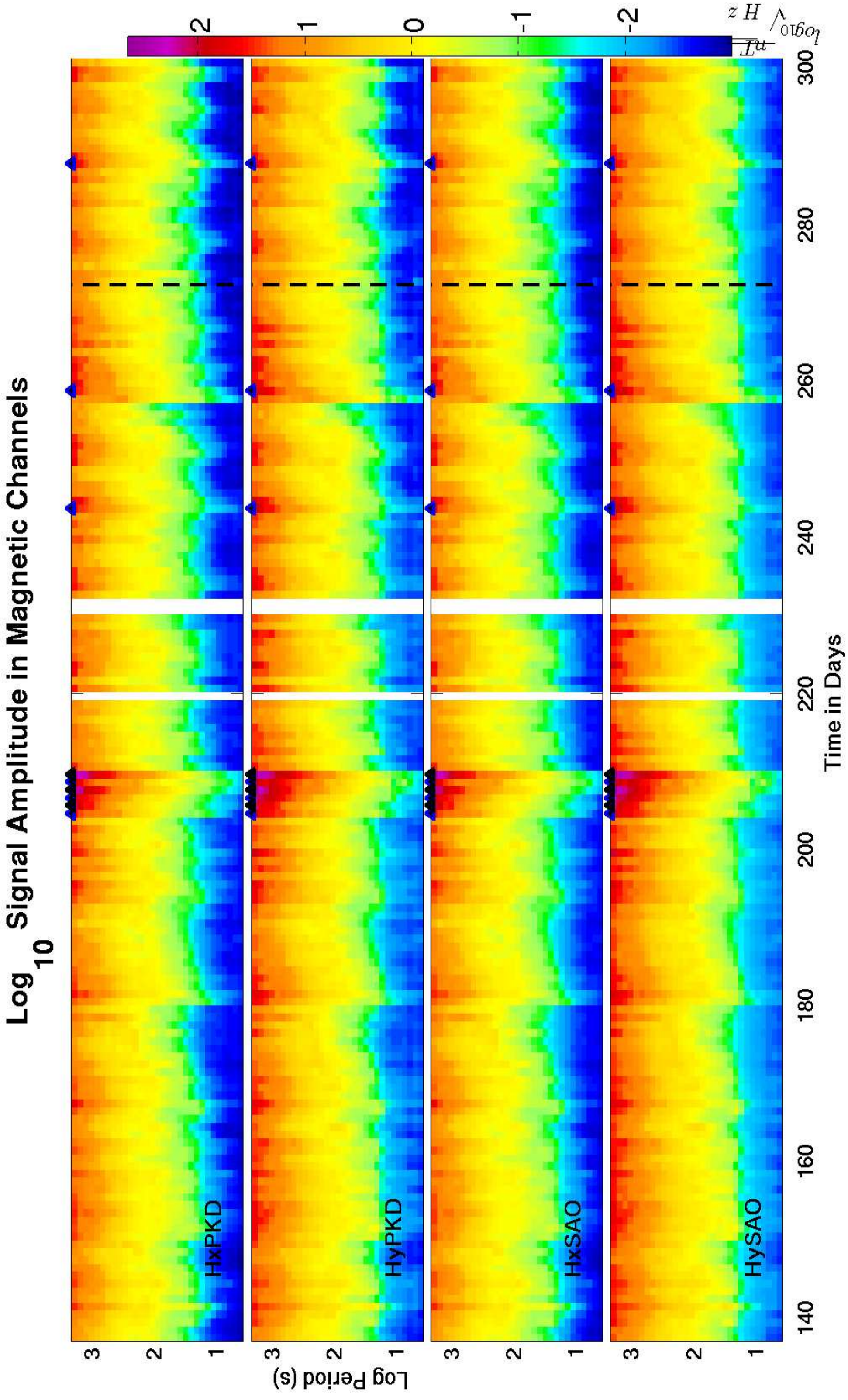


Figure 4.18: Signal amplitudes in $\log_{10}(\mathrm{nT}/\sqrt{Hz})$. Black triangles mark days of major solar storms, and blue triangles mark minor storms.

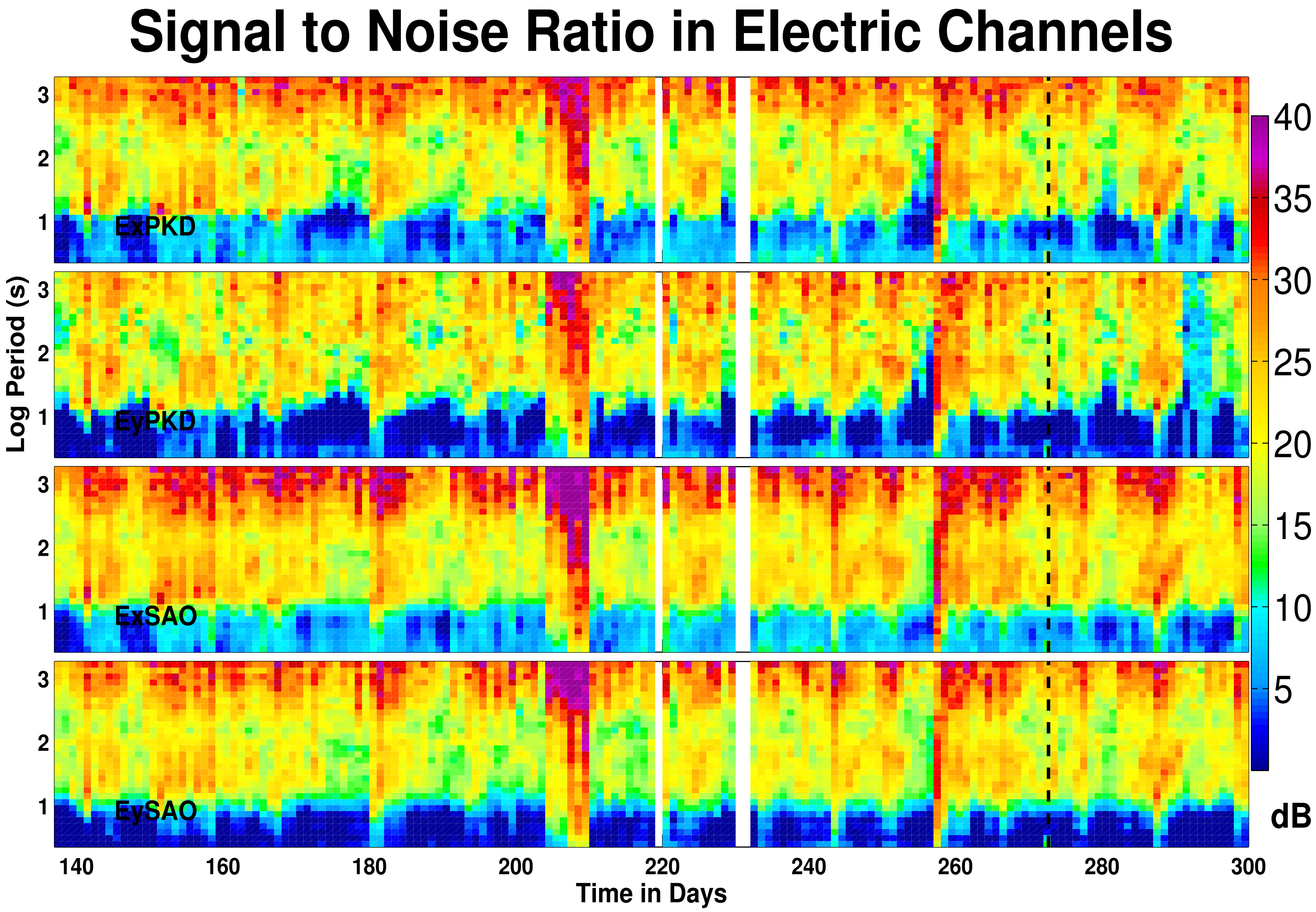


Figure 4.19: Signal to noise ratio in dB for electric channels.

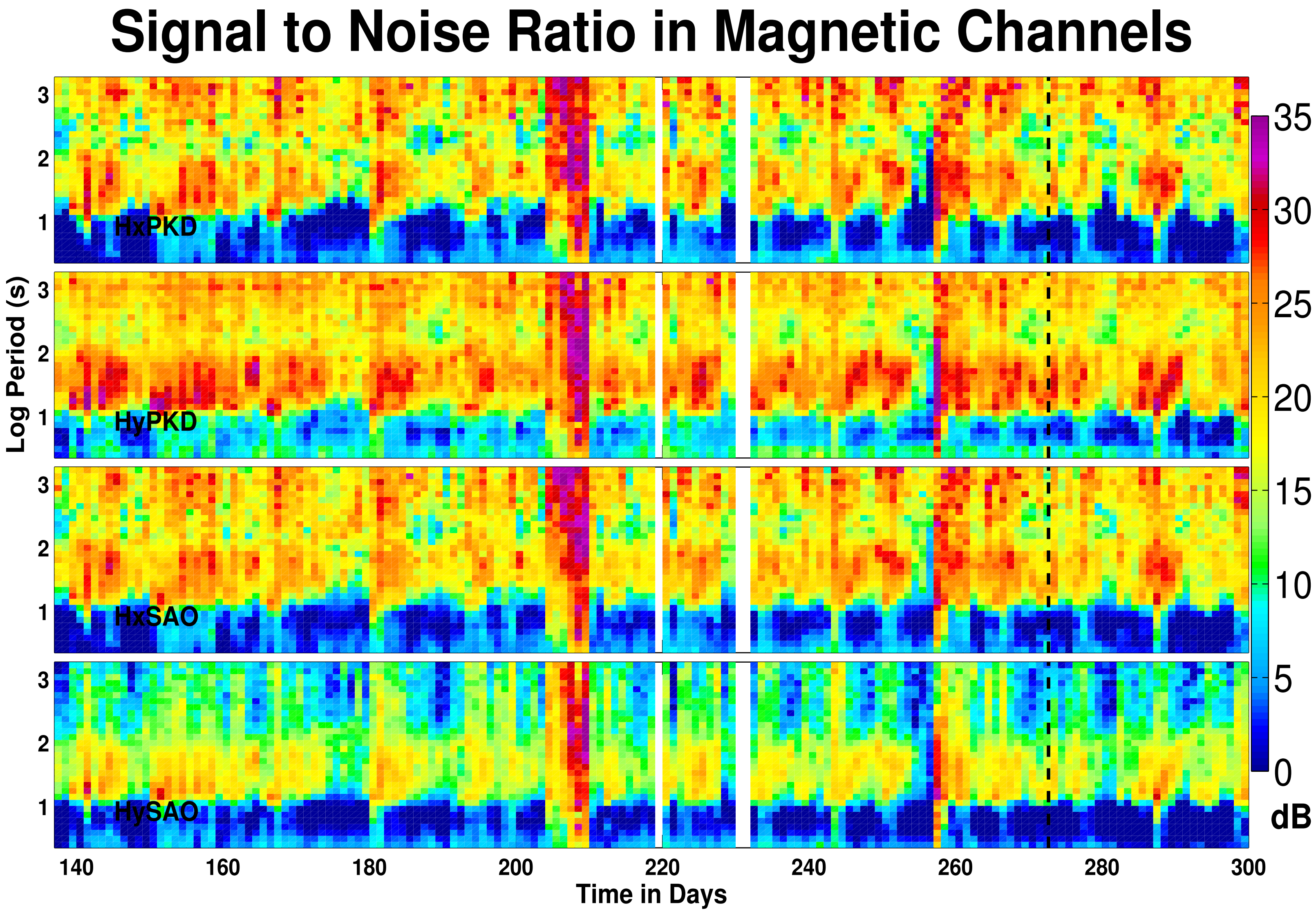


Figure 4.20: Signal to noise ratio in dB for magnetic channels.

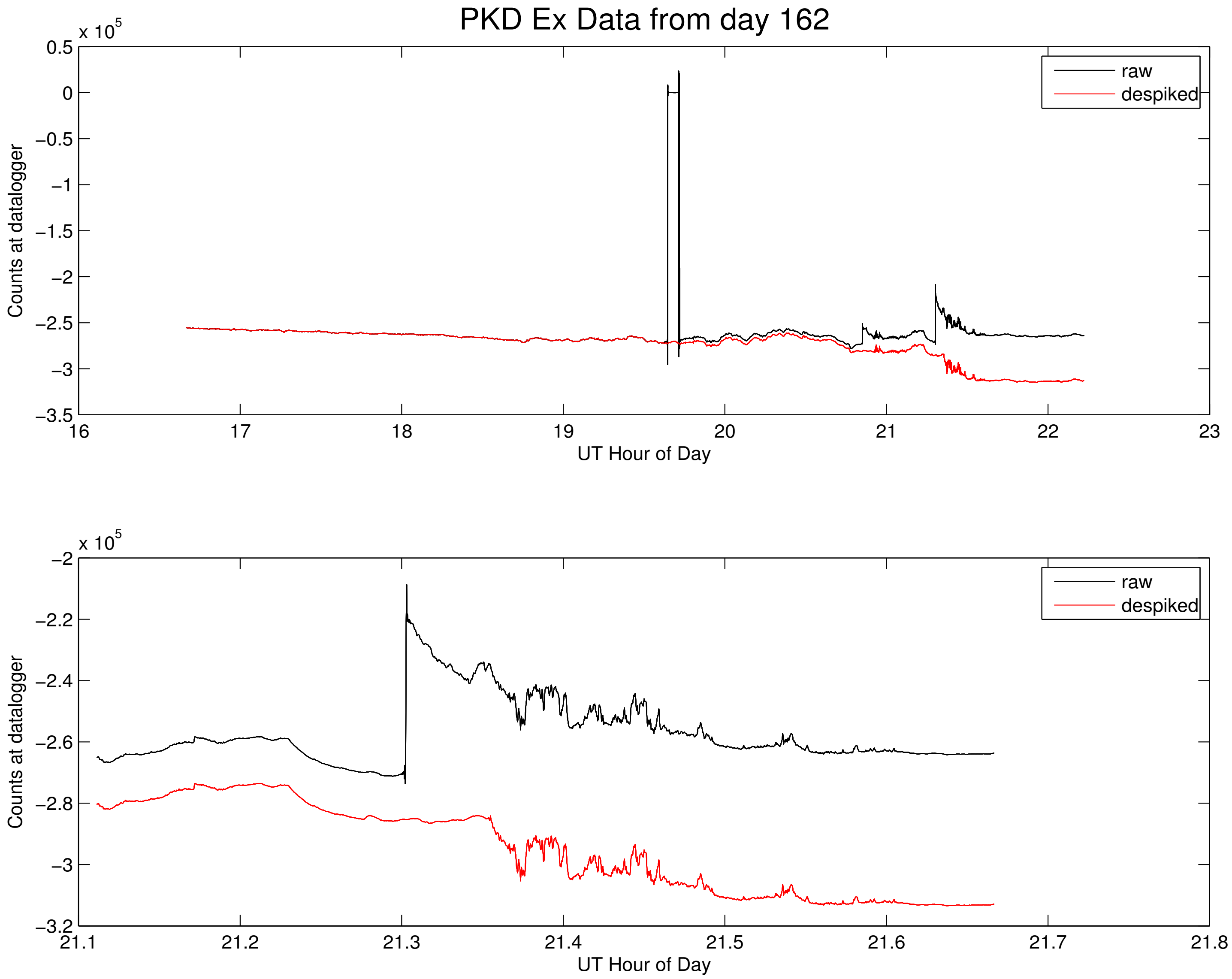


Figure 4.21: X-electrode data at PKD, day 162, 2004. The despiked data (red) is less corrupted than the raw data (black) but still shows unusual noise for a few hours.

otherwise one would expect to see accompanying magnetic activity, which we do not see.

Another interesting pair of features are the general increases in noise level in Ey at PKD between days 140 and 160, and again between days 280 and 300. These are seen as patches of yellow in the otherwise green bands around 100s period in Figure 4.15. An examination of the time series does not show a clear or simple explanation for these phenomena. In the first case, noise levels in the magnetic sensors show a corresponding increase, indicating the phenomena are more likely to be natural than those in the second time period, for which the effect seems constrained to electric fields only.

4.6 Relevance to previous observations

Undoubtedly one of the most famous electromagnetic anomalies which has been associated with seismic activity is that observed before the Loma Prieta earthquake by *Fraser-Smith et al.* (1990). The observation was made with a solenoid coil sensor whose cardinal orientation is not documented in the paper. Our plots of signal, and incoherent noise do not show, to first order, any strange behaviour which may be interpreted as being associated with the earthquake. Furthermore, there are several broadband pulses seen in the four-year plot. Two of these fall within the 163 day window of Figures 4.15-4.20, and span several days. These can be ruled out as being earthquake related under two principles. First, they correlate with days of high A_p geomagnetic activity index, and second, they can be observed at both sites. We point out, however, that these plots are not well-suited to be scrutinized for earthquake anomalies, as they are made from data which downweights outliers. The key point of these plots is that the broadband, natural, spatially correlated ULF micropulsations are to first order the main components of the observed signal. In order to look for signals originating in the subsurface, the components of the signal which are correlated between sites should be subtracted from the raw data. It is interesting to note, however, that the ULF magnetic field anomaly observed by Fraser-Smith et.al., which has been described as being a possible precursory signal to the Loma Prieta earthquake, was sufficiently large as to stand out clearly above the natural fields in several frequency bands, most notably, the MA3 band at 0.01Hz. We plot the field amplitudes in the “MA3 band” below for the 163 day period around the earthquake.

The MA index scheme is simply a method for chopping the broadband ULF fields into frequency bins. An explicit description of the signal processing used to generate MA indices is given in *Bernardi et al.* (1989). The method is a standard form of windowing time-series and band-averaging Fourier coefficients over logarithmically spaced bins in frequency domain. The raw data input to the MA index signal processor is sampled at 30Hz. In an effort to make a plot which is directly comparable to that which the MA index generator would make, we resample the 40Hz array data at three quarters the rate. This introduces error on the order of less than one half of one percent. The procedure is then as follows: Data, scaled in Volts, are Hamming windowed in 136 second (4096 sample) sections with no overlap. The data are then FFT-ed and converted into SI units via Equation 4.5. Band

MA Index	Frequency Band [Hz]	FCs averaged over 30min window at 40Hz	FFT Bins for 136s window at 30Hz
MA3	0.01-0.02	19-36	2
MA4	0.02-0.05	37-90	3-6
MA5	0.05-0.10	91-180	7-13
MA6	0.10-0.20	181-360	14-27
MA7	0.20-0.50	361-900	28-68
MA8	0.50-1.0	901-1800	69-136
MA9	1.0-2.0	1801-3600	137-273
MA10	2.0-5.0	3601-9000	274-682
MA11	5.0-10.0	9001-18000	683-1365

Table 4.2: Table of Fourier Coefficients used to relate MA indices after *Bernardi et. al.*

averages are then calculated (apparently using a Daniell window, as there is no reference to weighting in the averaging scheme) according the FC indices in Table 4.2.

Although Bernardi does not explicitly mention that the data are demeaned before FFT-ing, we assume as much, as otherwise DC spectral leakage would severely contaminate the MA3 band. To validate the MA indices thus calculated, we plot the spectra of a few 30 minute windows using MA averaging [*Bernardi et al.*, 1989], together with the data used to validate the SI unit transforms earlier this chapter.

The MA index calculations generate realistic values (Figure 4.22). Figure 4.23 shows MA3 indices converted to SI units for all horizontal coils in the array, shown over the 163 day window around the earthquake. Comparing this plot against the observations published in *Fraser-Smith et al.* (1990), shown in Figure 4.24, clearly shows an absence of variations of the sort reported by *Fraser-Smith et al.* (1990). The only variations anywhere near as large as those reported in the 1990 paper occur during times of enhanced solar activity.

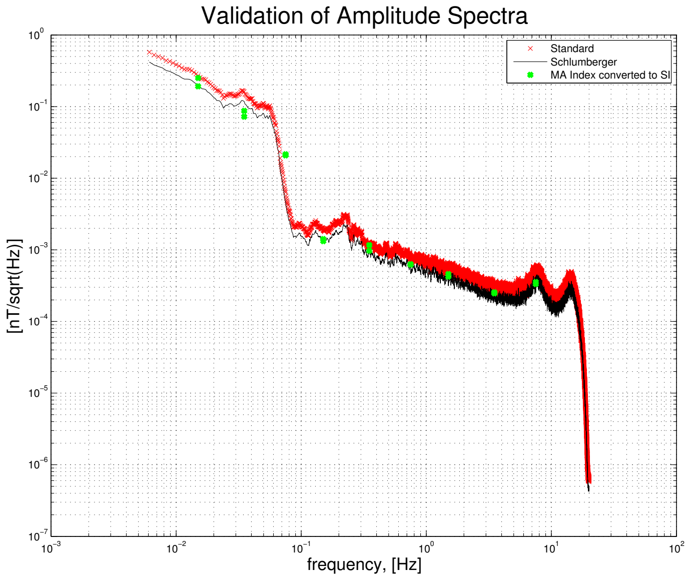


Figure 4.22: MA indices plotted over validation spectra.

Finally, we address the topic of boxcar-like variations in the electric field as cited by Varotsos and collaborators (1984,1991). We note that the days when data are flagged as bad by the variance ratio method are rife with these sorts of features. A survey paper on electrode instabilities by *Perrier et. al.* (1997) summarizes an experiment known as the Garchy electrode experiment. This experiment determined empirically that these sorts of SES-like signals appear in the electrode data for unknown reasons at random times. Pham et al. (1999) has suggested that some of these signals could be linked to coupling between radio wave transmissions and the solid earth. Figures 4.25 through 4.32 show so-called electrode instabilities in our data.

For example, Figure 4.26 depicts day 259, 2002, a time when no site visits or maintenance

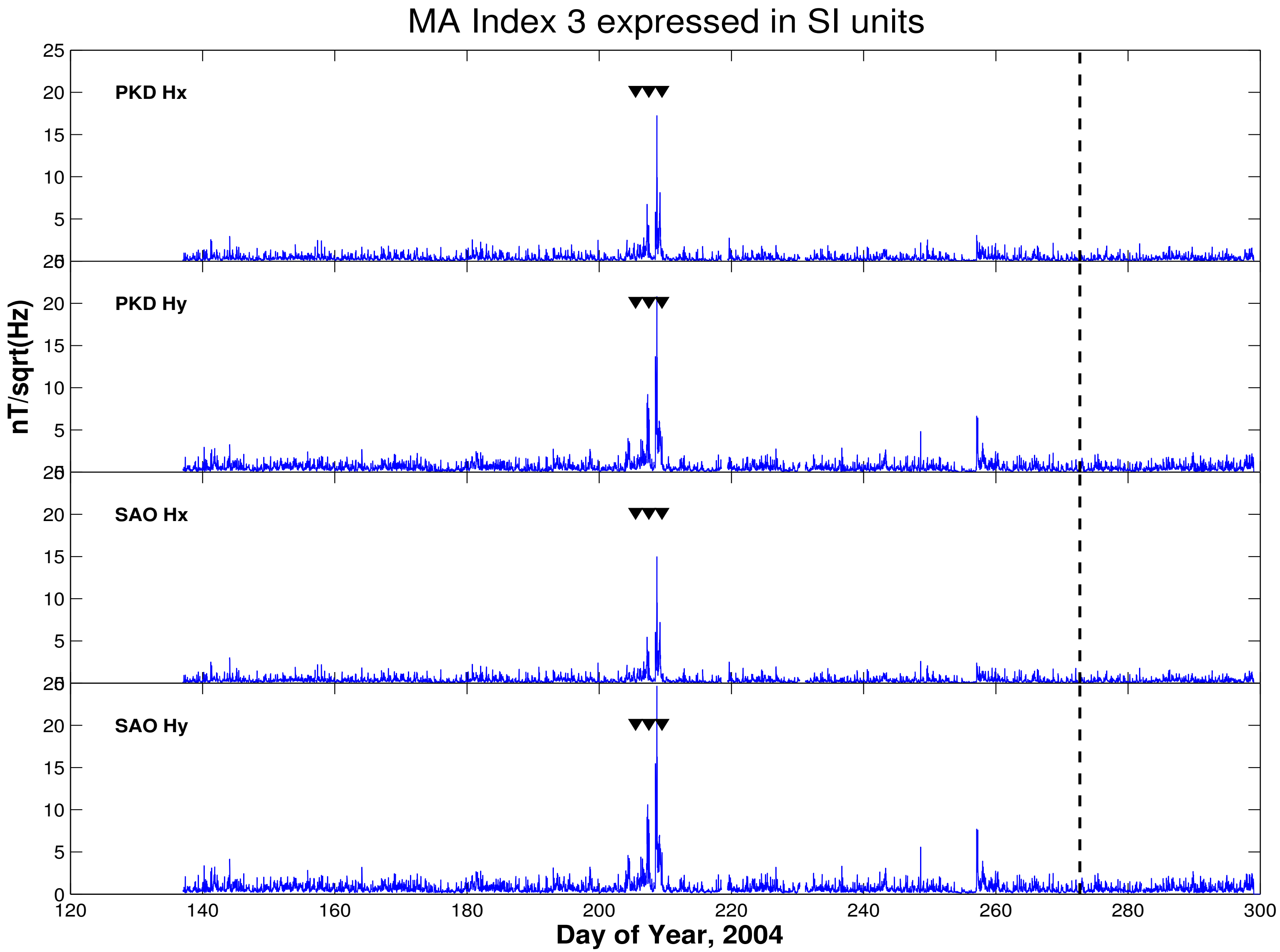


Figure 4.23: Time series of MA index 3 in SI units, to be directly compared with the figure from *FS1990*. Black triangles denote days of major solar storms.

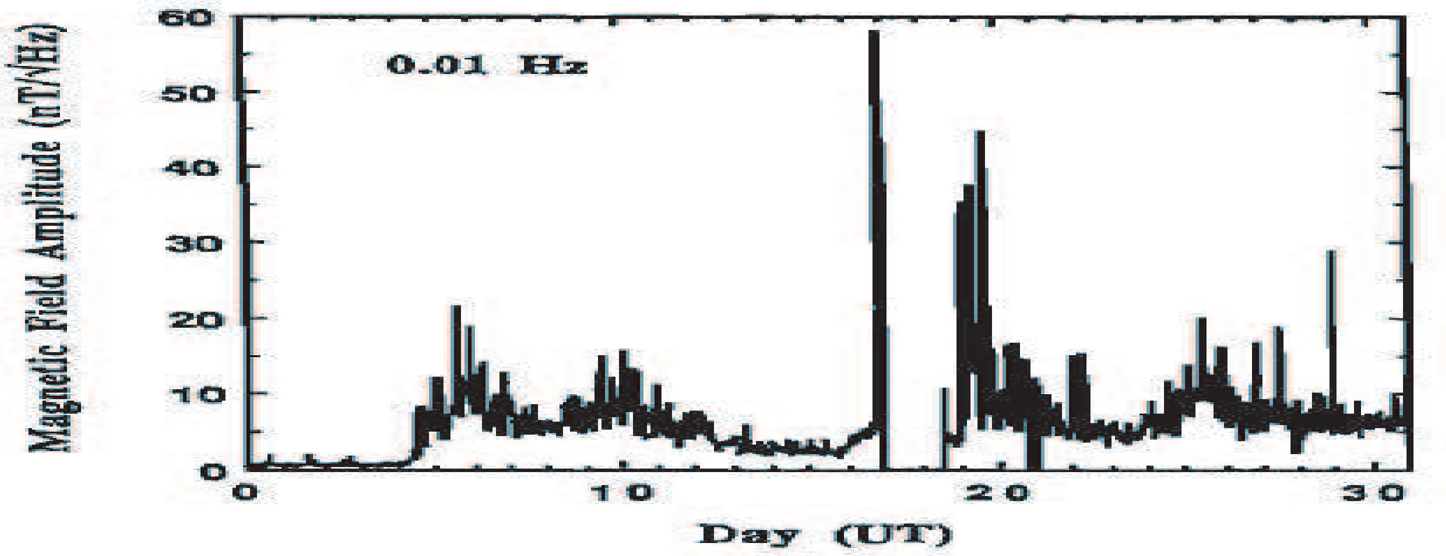


Figure 4.24: Time series of MA index 3 published in *Fraser-Smith et al.* (1990)

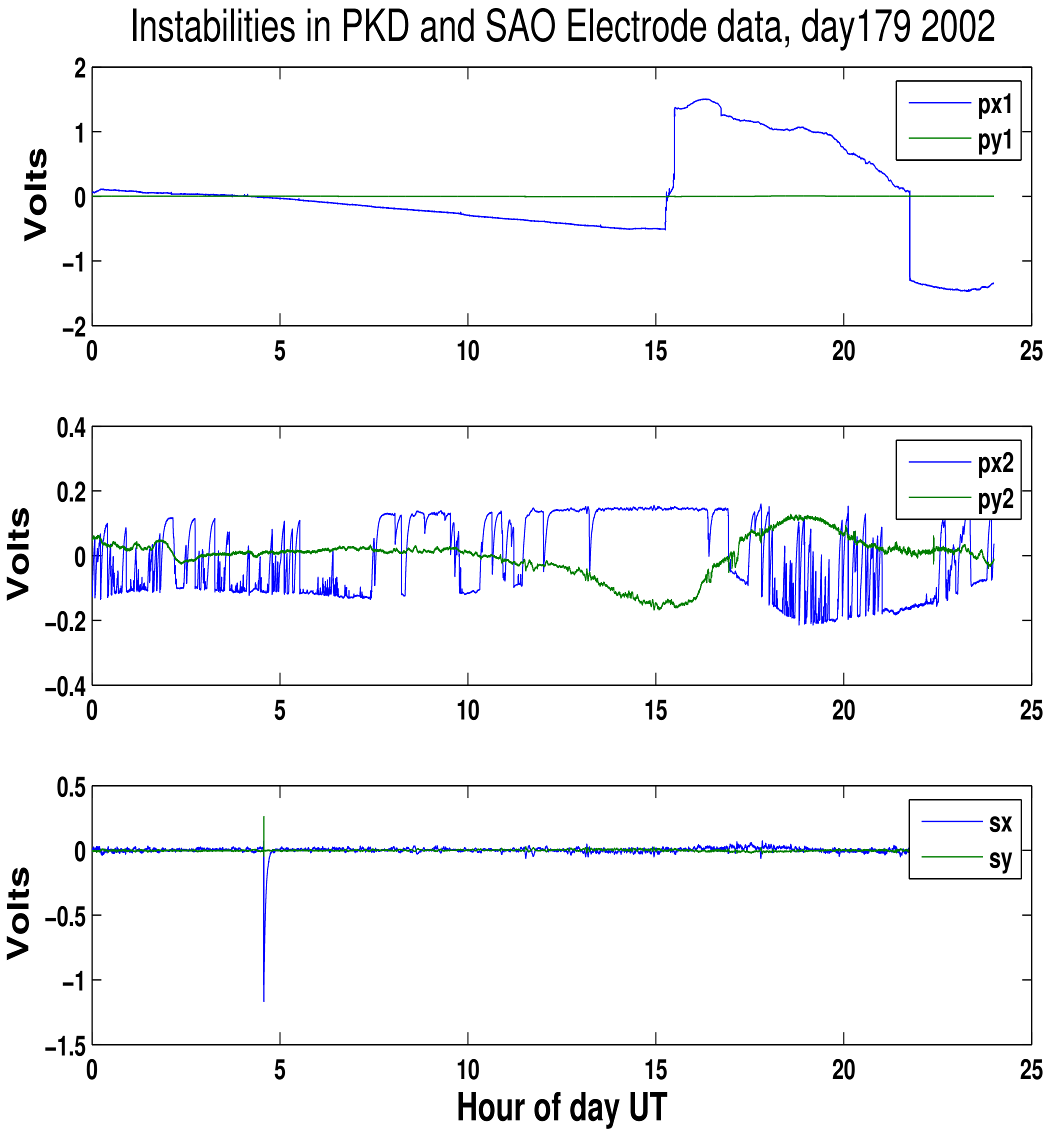


Figure 4.25: Parallel electrodes at PKD show completely different instabilities synchronously, while instability is present in only one channel of the two at SAO. Micropulsations on most channels agree however, i.e. these instabilities are superimposed upon natural fields. No site visits were documented at either site within over a week of this date. a) instability is so large that micropulsations are rendered invisible on scale of anomaly. b) completely different type of instability in 200m electrode. c) Hollister data also shows an instability.

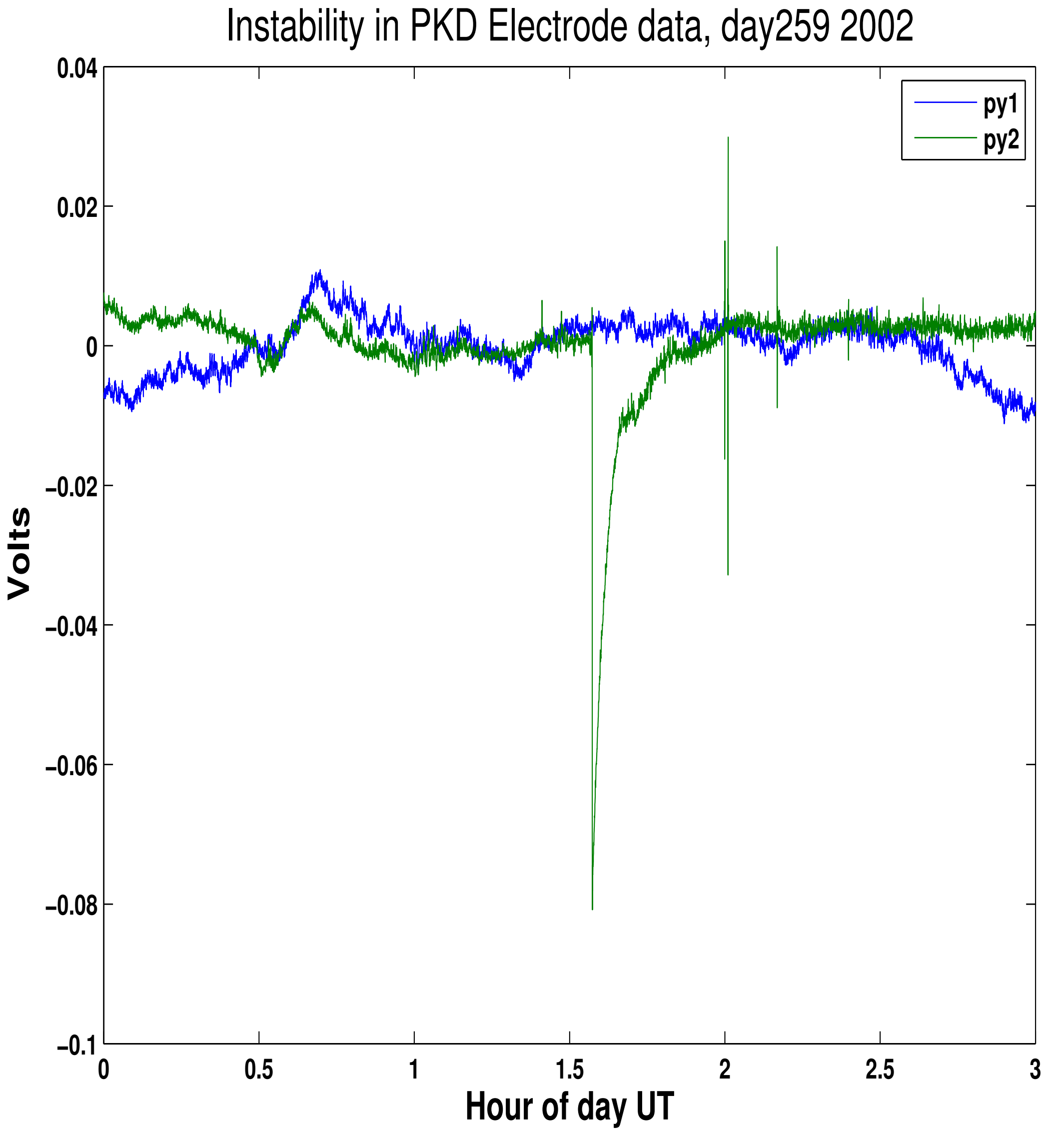


Figure 4.26: Electric field data from two parallel electrodes at PKD showing instability present in only one of the electrodes. No documented site visits occurred in the weeks prior to this time, but there was a visit four days later.

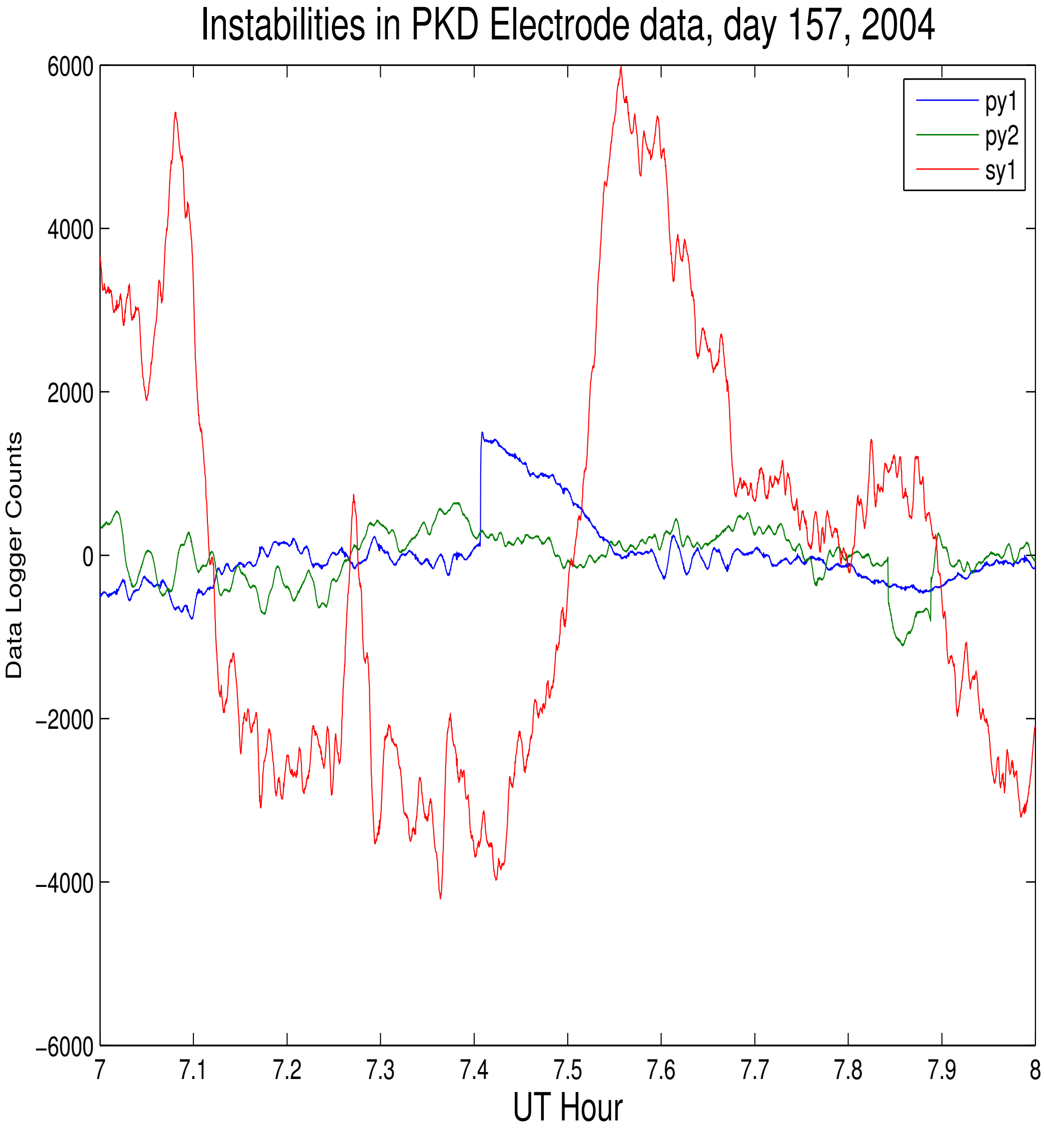


Figure 4.27: Electrode instabilities in both the PKD Ex (blue) and PKD Ey (green) data streams.

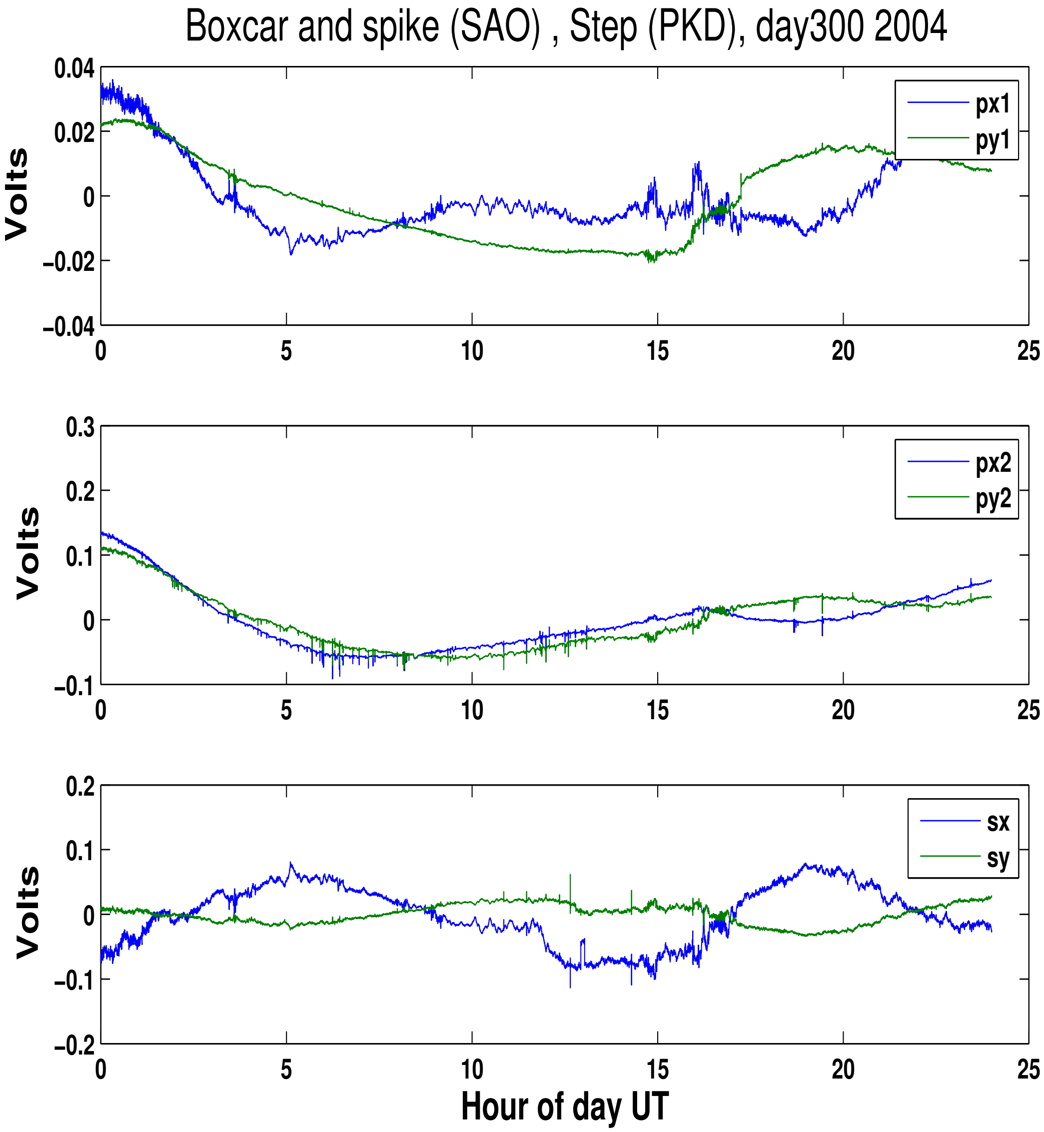


Figure 4.28: Electric field data showing step in PKD 100, spikes in PKD200, and a boxcar (SES-like) signal at SAO. No site visit occurred before, but there was one 2 days after the event, on day 1032.

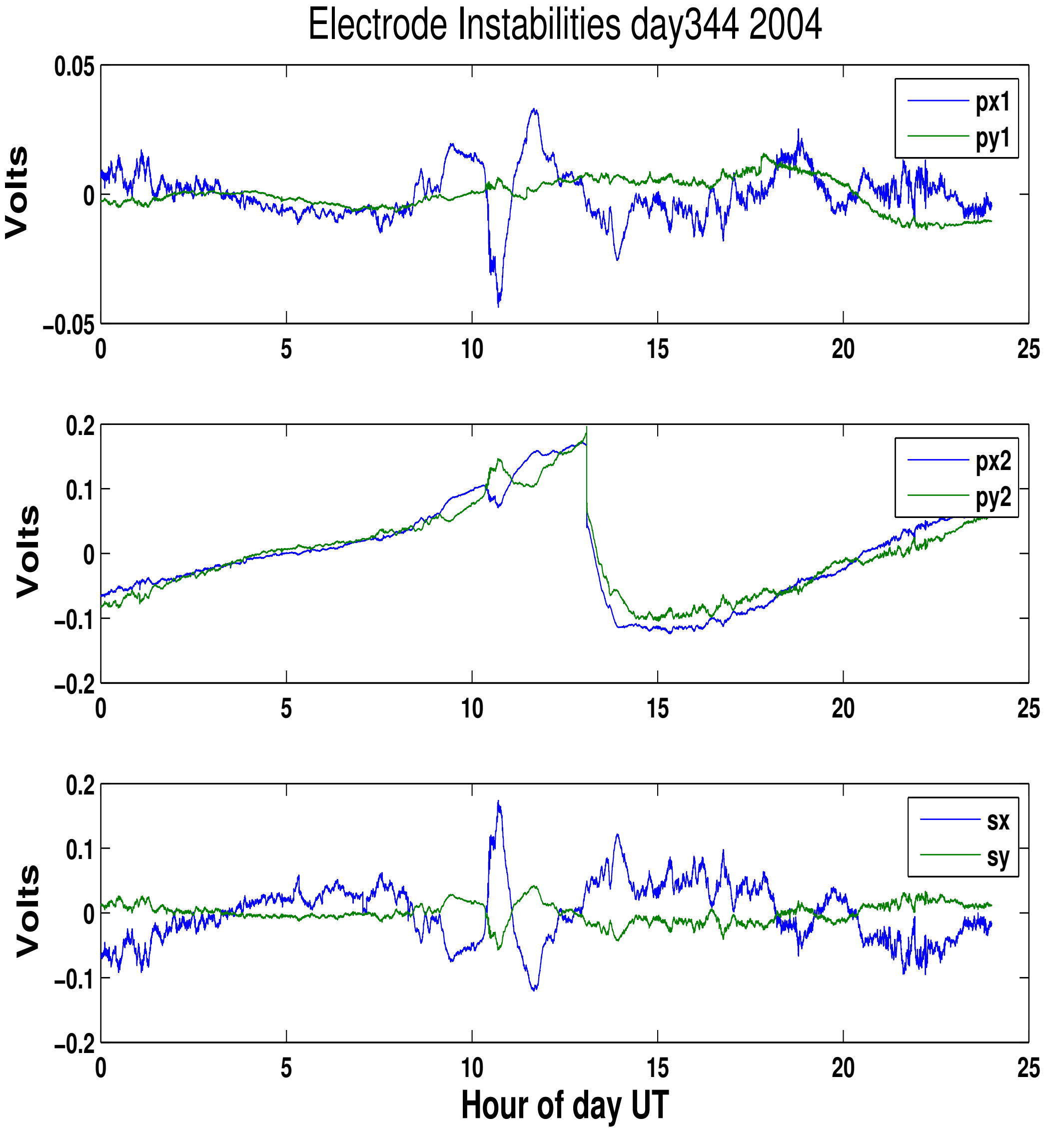


Figure 4.29: Strange instabilities in the 200m electrodes at PKD, which are not common to the 100m electrodes. SES-type signal is also present at SAO, but not visible at this scale.

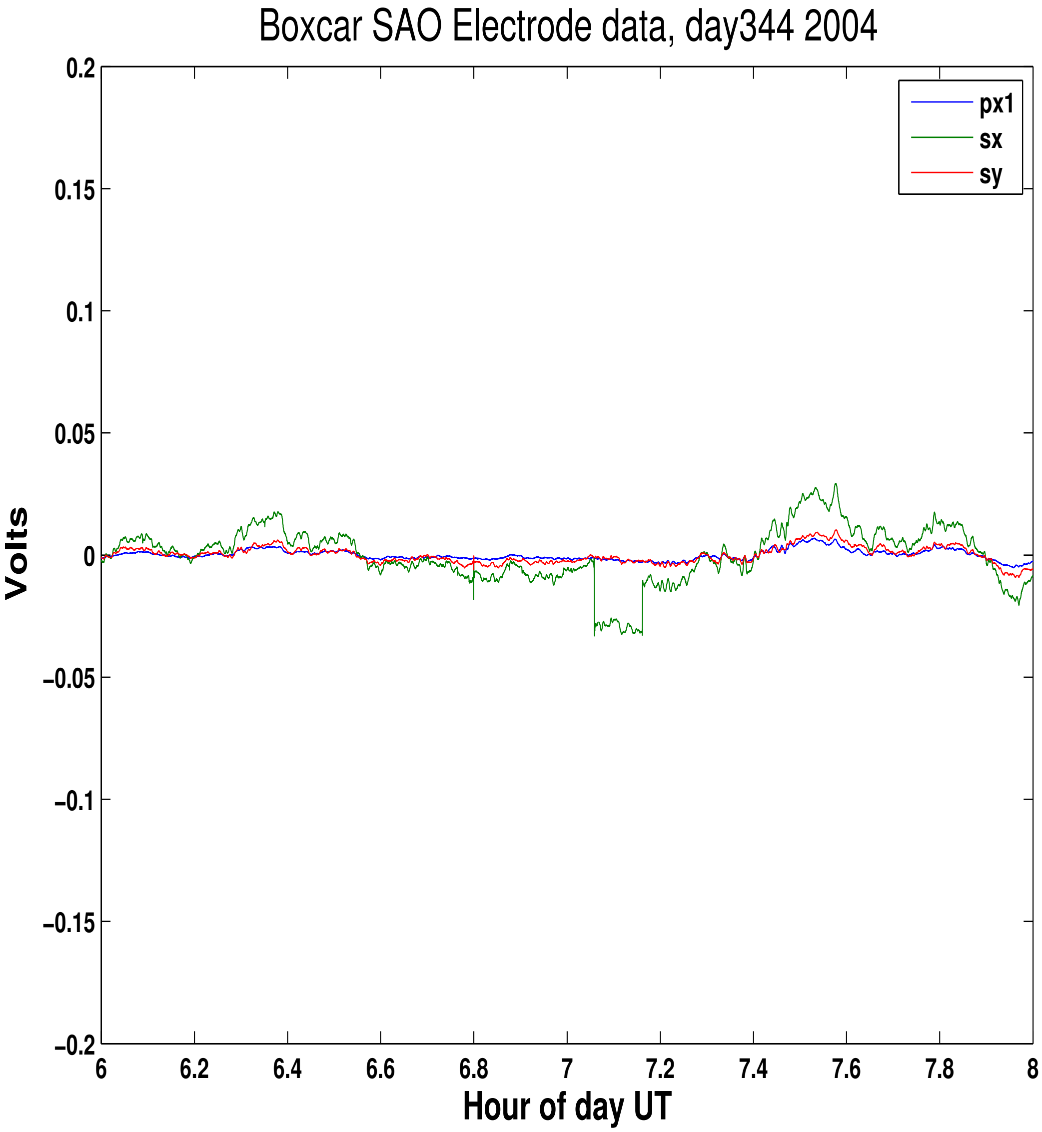


Figure 4.30: A zoom in on SES at SAO mentioned in Figure 4.28

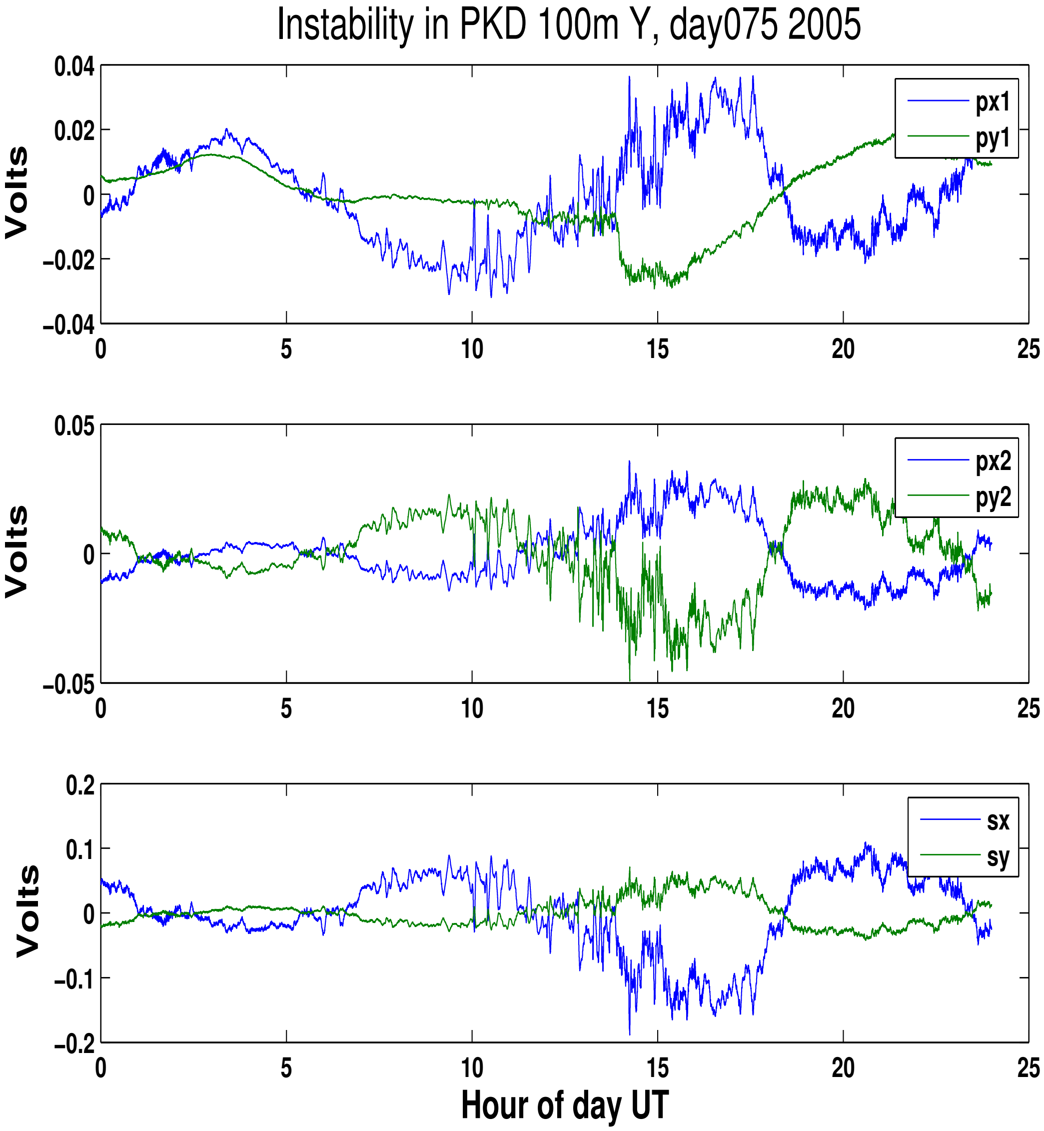


Figure 4.31: Instabilities in the 100m electrodes at PKD. These signals are possibly the result of watering of a dry electrode that occurred one day prior to the observation.

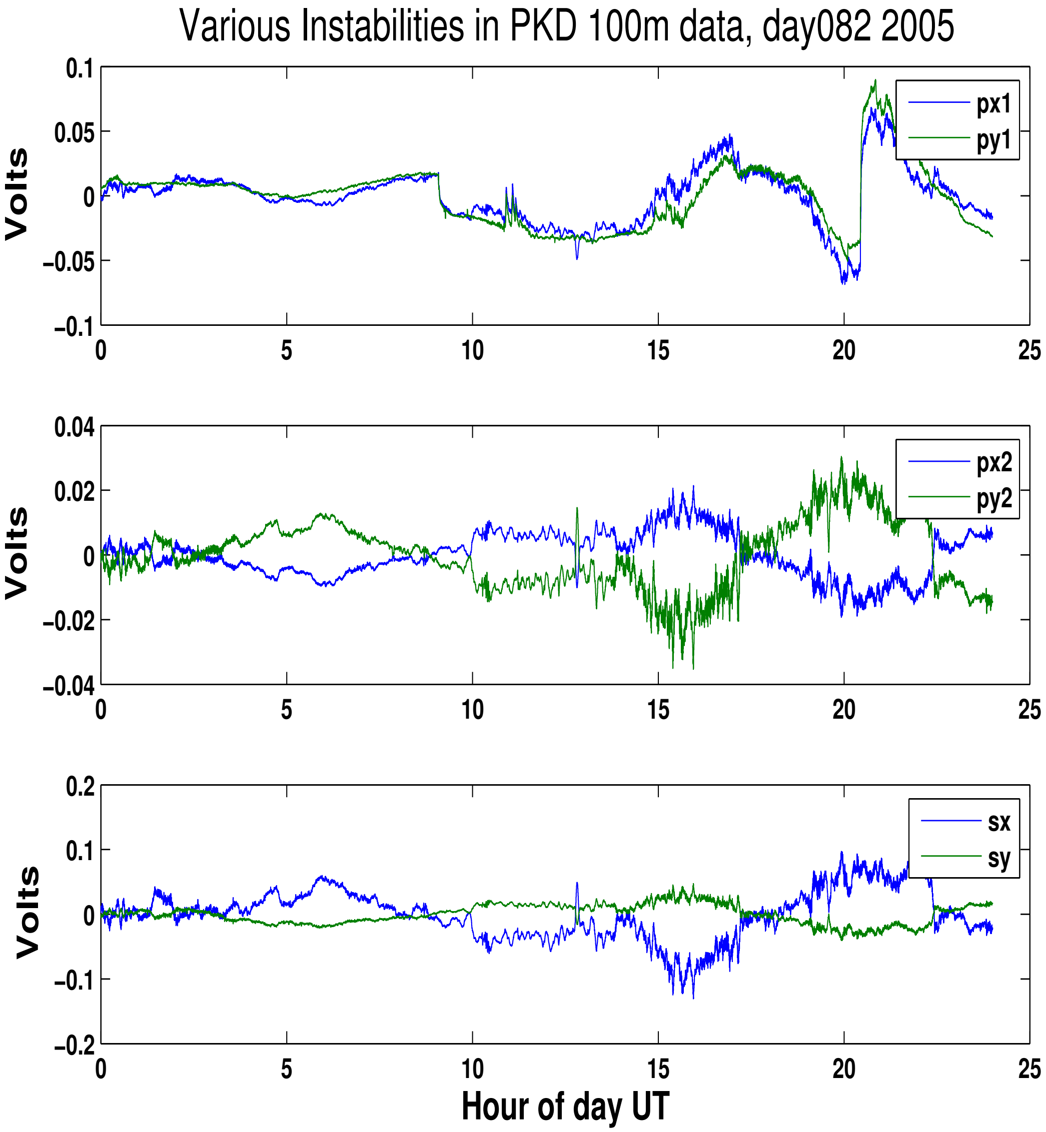


Figure 4.32: Sharp offsets in PKD electrodes which are not mirrored by SAO electrodes.

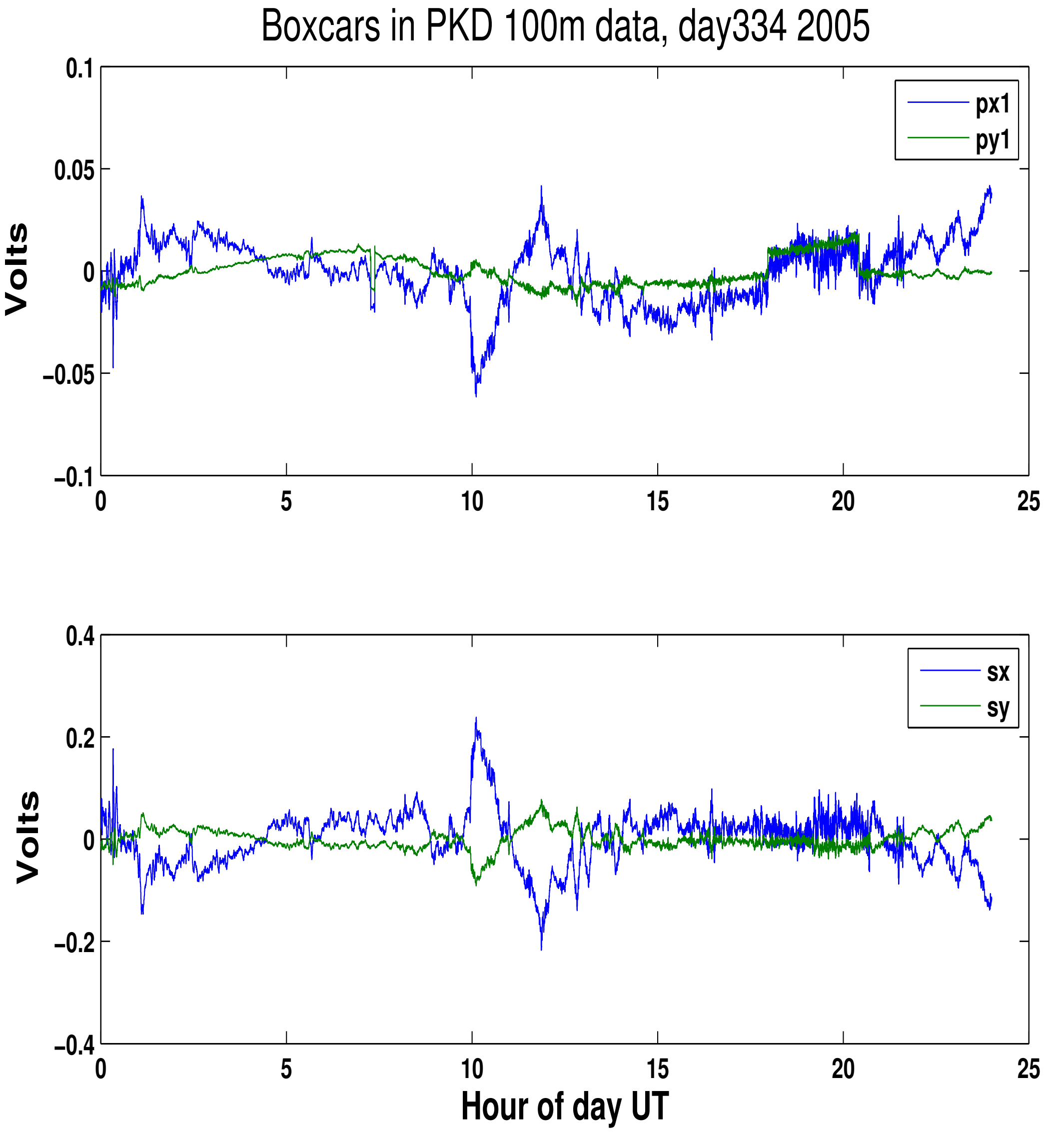


Figure 4.33: Boxcar-like signals of various lengths local to the PKD electrodes.

were documented. Two parallel electrodes track the magnetotelluric variations consistently, but some unexplained signals are superimposed on top of the data. It is hard to imagine a physical process which would result in these signals being present in one electrode and not in the other. Instead, chemical reactions occurring in the immediate vicinity of the electrodes might produce this effect. It is also possible that some of these instabilities are due to faulty electronics in the EFSC or data loggers.

Chapter 5

Frequency Domain Residuals

The monitored fields have been observed, at least to first order, to be free of anomalous behaviour associated with the Parkfield earthquake of 2004. This motivates an examination of lower order signals. There are a variety of techniques which can be used to do this. We begin with a simple and conventional technique, the classic remote reference residual calculation.

5.1 Remote-reference residual theory and calculation

From the previous chapter, it is clear that natural, everpresent, electromagnetic noise is typically observed on the order of a few pT to a few nT at all frequencies considered here. Moreover, the natural magnetic noise observed at the two sites is very nearly identical [*Kappler et al.*, 2005] and the electric fields are related by a time-invariant transfer function linking the fields observed at PKD to those observed at SAO. This transfer function has been shown empirically to be stable in [*Eisel and Egbert*] (2002). In order to see any local electromagnetic phenomenon on the order of this noise or smaller, these natural, 'magnetotelluric' fields must be separated from the observed data.

In an attempt to perform this separation, we employ a method of predicting the fields at Parkfield with the fields observed at Hollister using an intersite transfer function.

The motivation for this is the following assumption: Should there be a subsurface process, which were somehow generating a signal associated with the Parkfield earthquake, such a signal would likely be much attenuated, if detectable at all at SAO. Should such a signal exist at PKD, it could be present in more than one channel. Thus, by regressing a PKD channel on a collection of channels which includes other PKD channels, it is possible that a subsurface signal would be not be visible in the residual time series, as it may be canceled by the contributions of the local channels. If indeed a signal were present in only one channel, it should have shown up in the noise plots (Figures 4.8, 4.9, 4.16, 4.15). We thus regress each Parkfield channel on the set of SAO channels, with the intention of removing as much natural field noise at PKD as we can, leaving as a residual the "site noise". The principle is essentially the same as that underlying our calculations of incoherent noise in the previous chapter, but in this case we determine the channel residuals without using any local data. It is well known that the TF will vary as a function of frequency, and so, like incoherent noise, it is calculated separately for each band. As discussed in chapter four, residuals are only meaningfully calculated in the complex domain before band averaging is carried out so the prediction must be done separately for each time series of Fourier coefficients. Initially, we calculate the TF for each channel at PKD, for each FC in each band daily. Considering the 4xT time series of Fourier coefficients for all four channels at PKD as a matrix **D**, and the 4xT time series of Fourier coefficients for all four channels at SAO as a matrix **M**, we solve, daily:

$$\mathbf{D}_{PKD}(t) = \mathbf{G}\mathbf{M}_{SAO}(t) \tag{5.1}$$

for the complex transfer function **G** relating the data at the two sites. The residual time series are calculated and stored as data structures with the same configuration as the Fourier coefficient files. Examples of the application of this method three separate FFT bins are shown in Figures 5.1-5.3. These show the raw data observed at Parkfield, the predicted field by **G** acting on observations at Hollister, and the residual time series **r**(t):

$$\mathbf{R}(t) = \mathbf{D}_{PKD}(t) - \mathbf{G}\mathbf{M}_{SAO}(t) \tag{5.2}$$

Figures 5.1-5.3 show bins of higher frequencies, $\sim$4s, midrange frequencies $\sim$ 20s, and lower frequencies ∼500s, respectively. Note that the fit around 4s is not nearly as good as it is at lower frequencies. This is due to the “dead-band” effect, which is to say that there is very little natural signal at periods between 1 and 10 seconds.

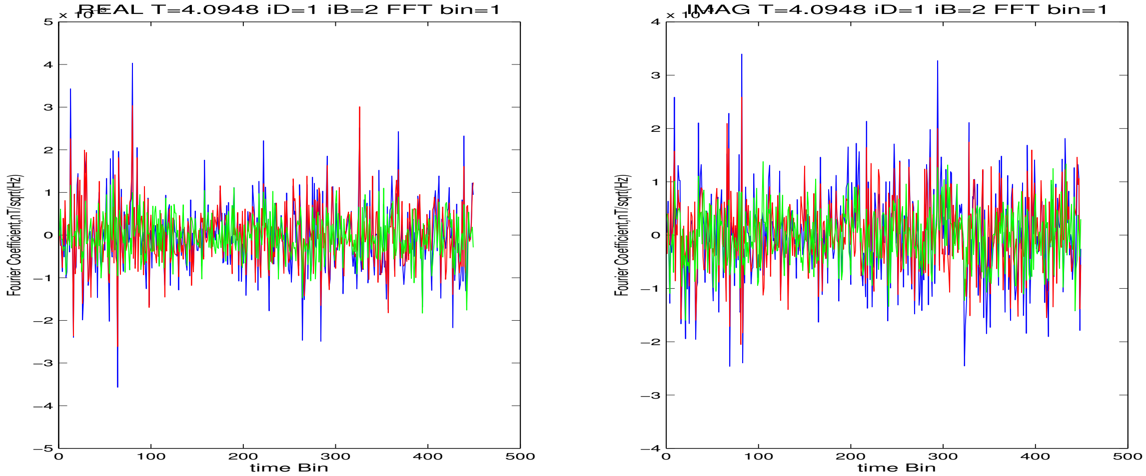


Figure 5.1: Least squares fit to the first Fourier Coefficient in the second band of the first decimation level. The band is centered at T seconds. Real part of the data is on the left, and the imaginary part on the right. Observed fields are shown in blue, remotely-predicted fields are in red, and residuals in green.

Spectral matrices can be calculated from these residual time series and the signal amplitudes in the residuals plotted over the long term. Four-year residual plots are shown below, where each day and frequency band are in terms of the RMS amplitude of the band averaged time series of residual Fourier coefficients. Note the relatively large residual in the second band from the bottom at around 4s, which is especially prominent in the Ey-component at PKD. This unknown signal component is local to the site. By inspecting this band more closely, we can see that this signal component is actually much more narrow in

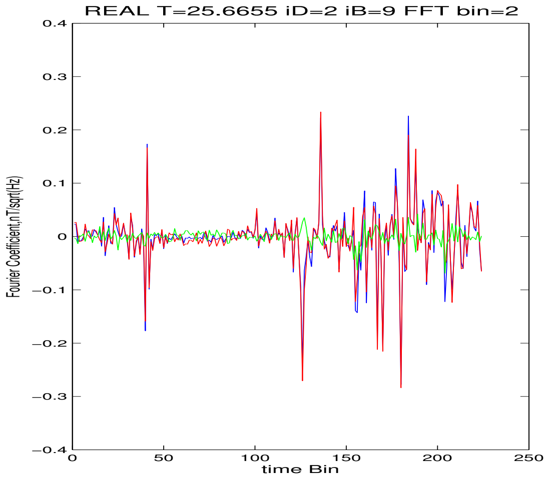


Figure 5.2: Least squares fit to the second Fourier Coefficient in the ninth band of the second decimation level. The band is centered at 25.6 seconds. Real part of the data is on the left, and the imaginary part on the right. Observed fields are shown in blue, remotely-predicted fields are in red, and residuals in green.

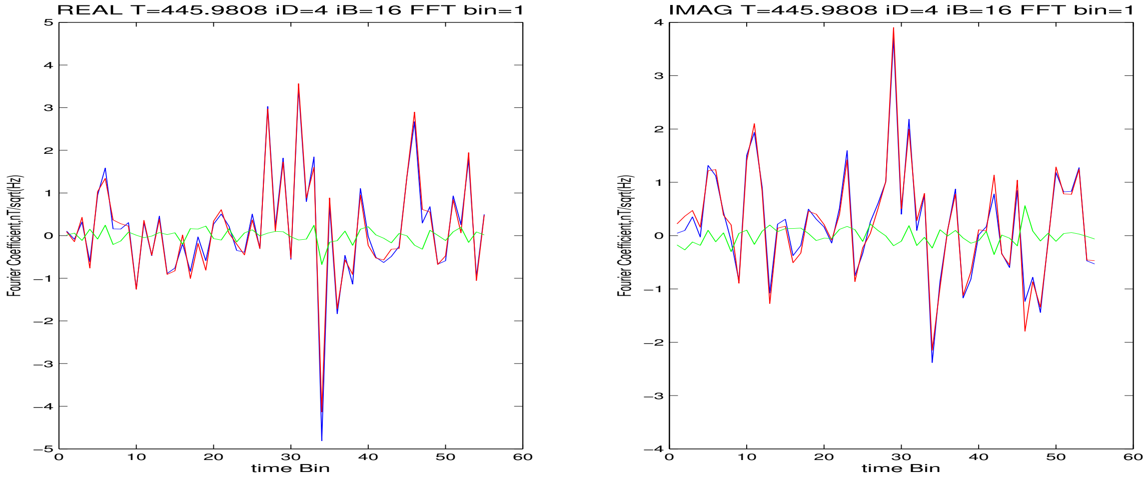


Figure 5.3: Least squares fit to the first Fourier Coefficient in the sixteenth band of the fourth decimation level. The band is centered at T seconds. Real part of the data is on the left, and the imaginary part on the right. Observed fields are shown in blue, remotely-predicted fields in red, and residuals in green.

bandwidth than shown at the spectral resolution in Figures 5.4 and 5.5. Only the fourth FC is significantly affected by this phenomenon.

A more condensed way to view this anomaly is to examine the residual of each FC in the band, as shown in Figure 5.8. Here, the residual and signal RMS amplitudes are typically near the same value for all FCs in the band with the exception of FC4 and its two nearest neighbours. This shows again that the signal is only present only at PKD, and is very narrow band (see Figure 5.8). It is more significant in Electric channels than in Magnetic channels.

Note also that though the magnetic storm is significantly attenuated in these plots as compared to the signal plots,it is still visible. This is likely due to a violation in the assumption of horizontally polarized uniform fields which can happen when large amounts of energy are injected into the global electromagnetic circuit. This is particularly evident when we zoom into the 163-day section around the earthquake time, as in Figures 5.9 and 5.10. Another unexplained phenomenon in these residual plots is the apparent broadband increase in noise power in Ey at PKD around days 150-153 (this corresponds to 29 May - 01 June 2004), which can be clearly seen in Figure 5.9.

In all the residual plots before and including Figure 5.10, regression was performed each day, and the transfer function allowed to vary in a way which optimized fits on a day to day basis. Of course, it is not obvious that a slowly appearing trend may not be subtracted out by this adaptive TF residual calculation. In order to ensure that no signals with this property

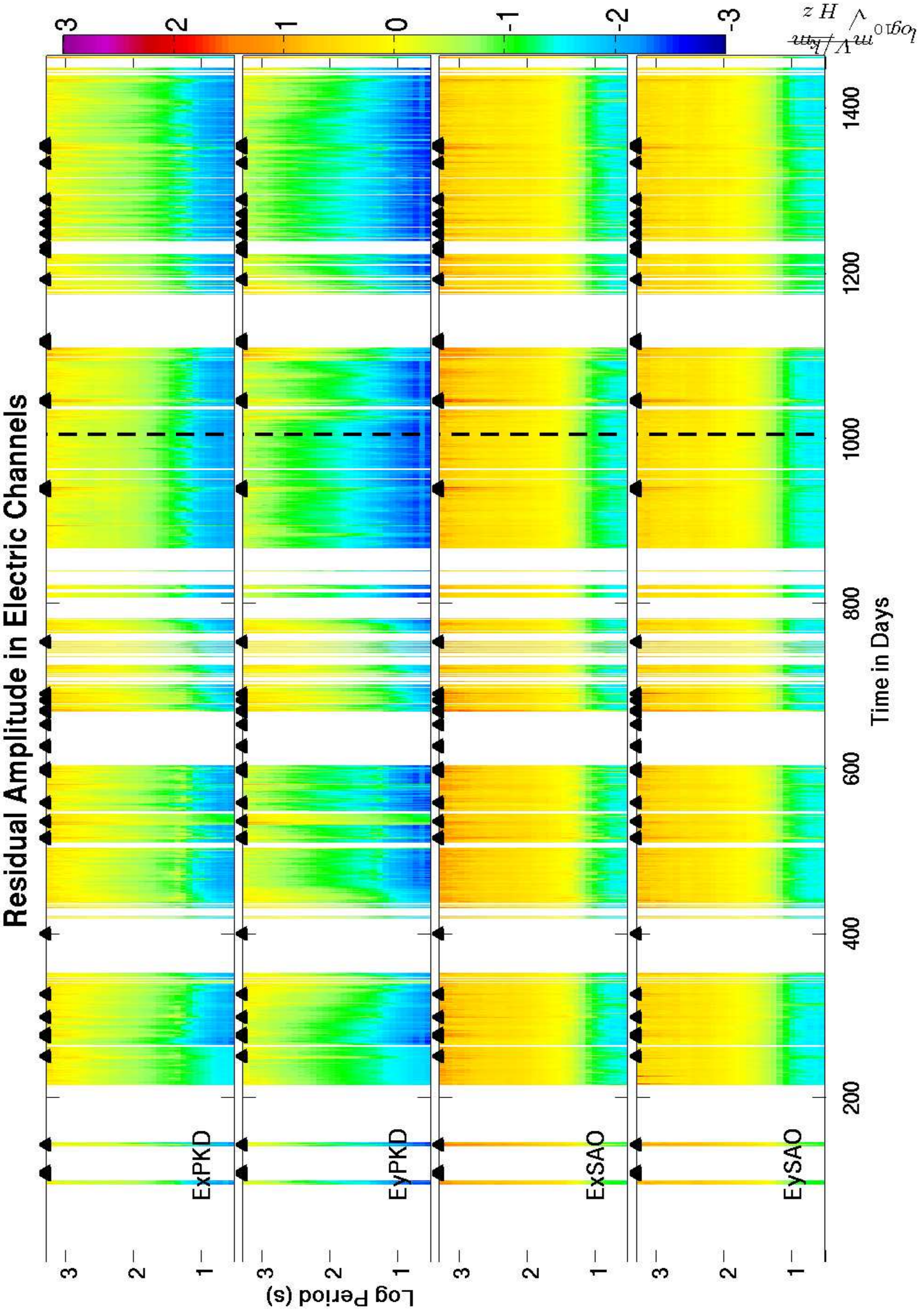


Figure 5.4: Residual amplitudes in $\log_{10}$(mV/kM/$\sqrt{Hz}$).

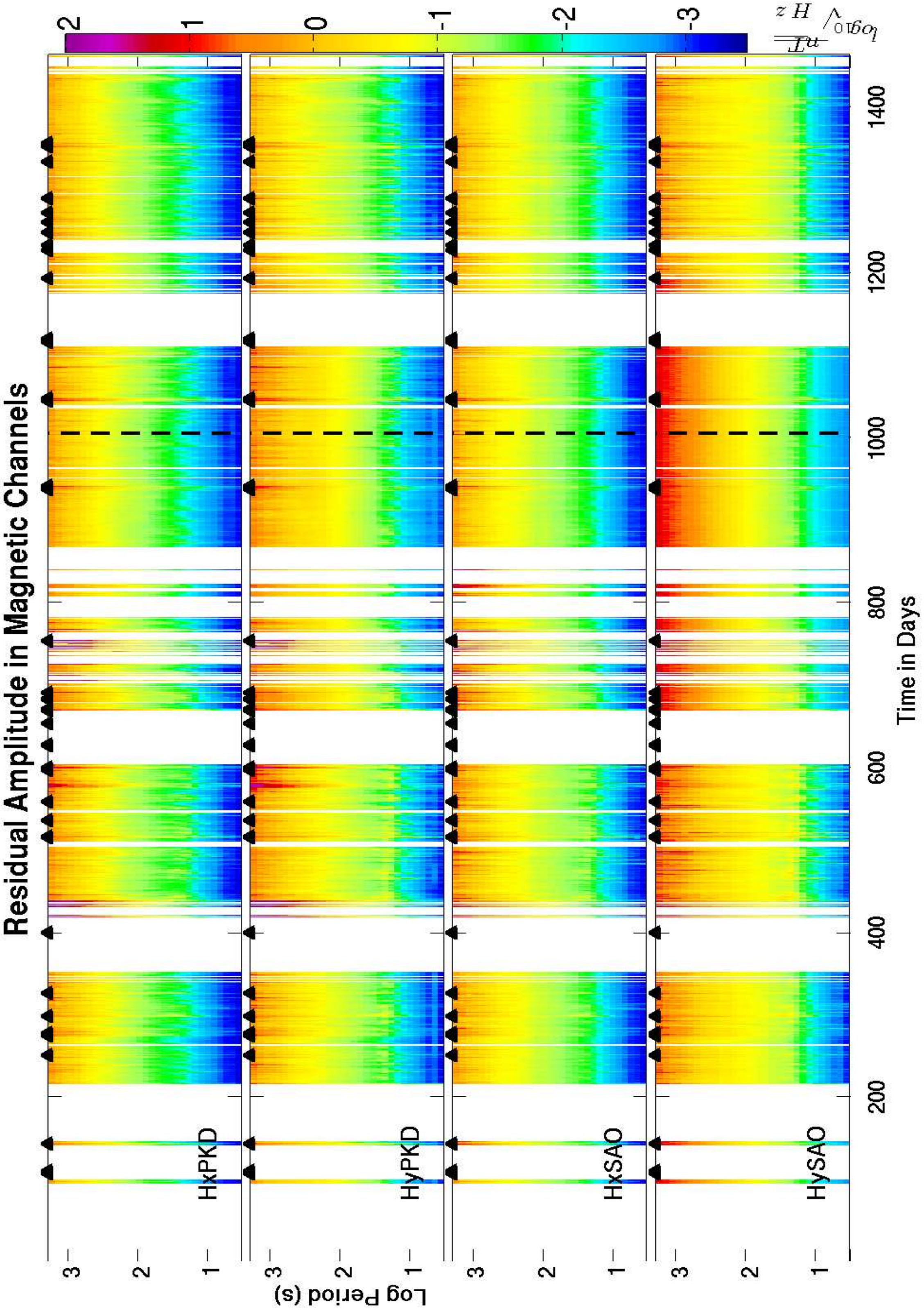


Figure 5.5: Residual amplitudes in $\log_{10}$(mV/kM/$\sqrt{Hz}$).

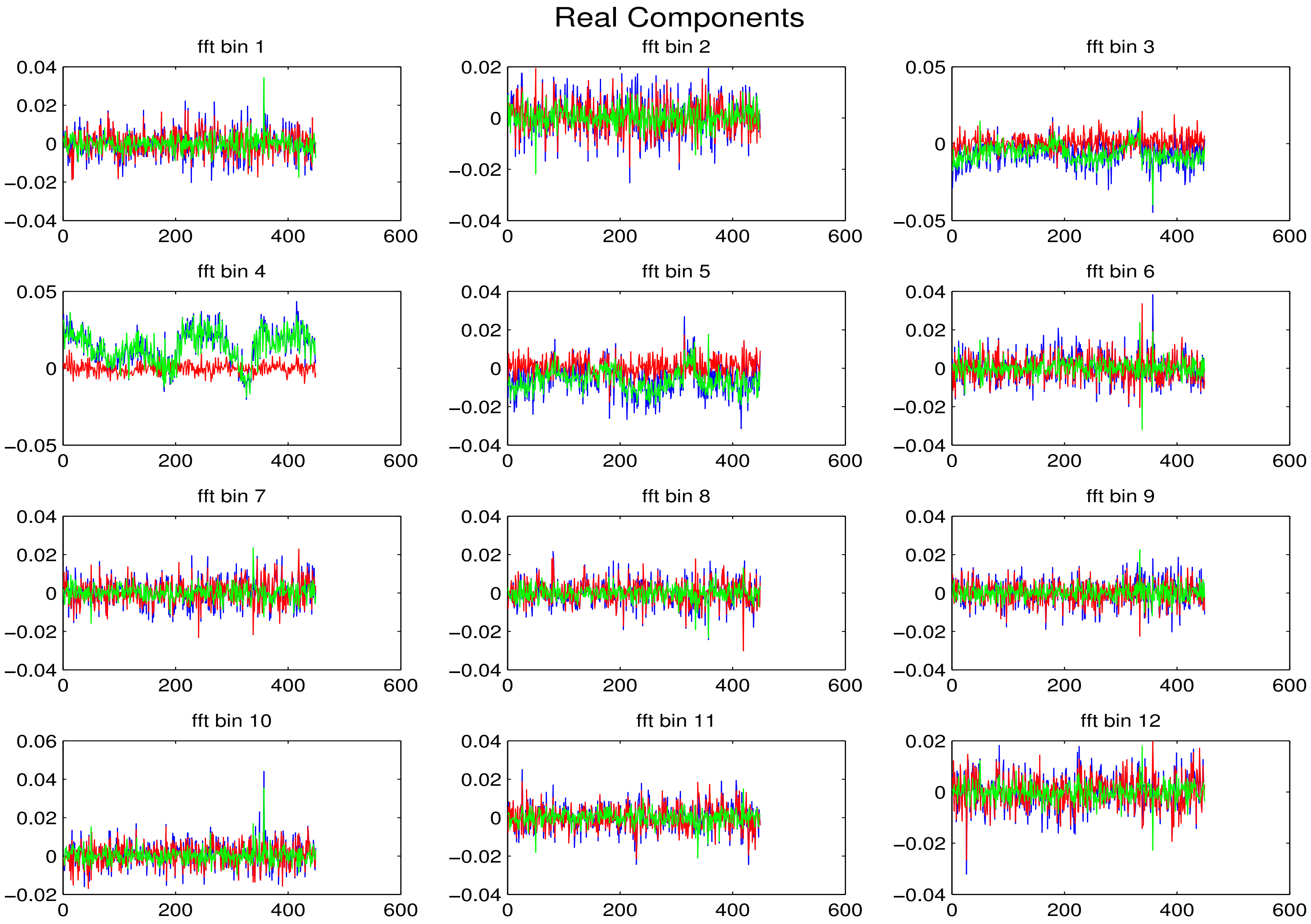


Figure 5.6: Observed fields in blue, remotely predicted fields in red, and residuals in green.

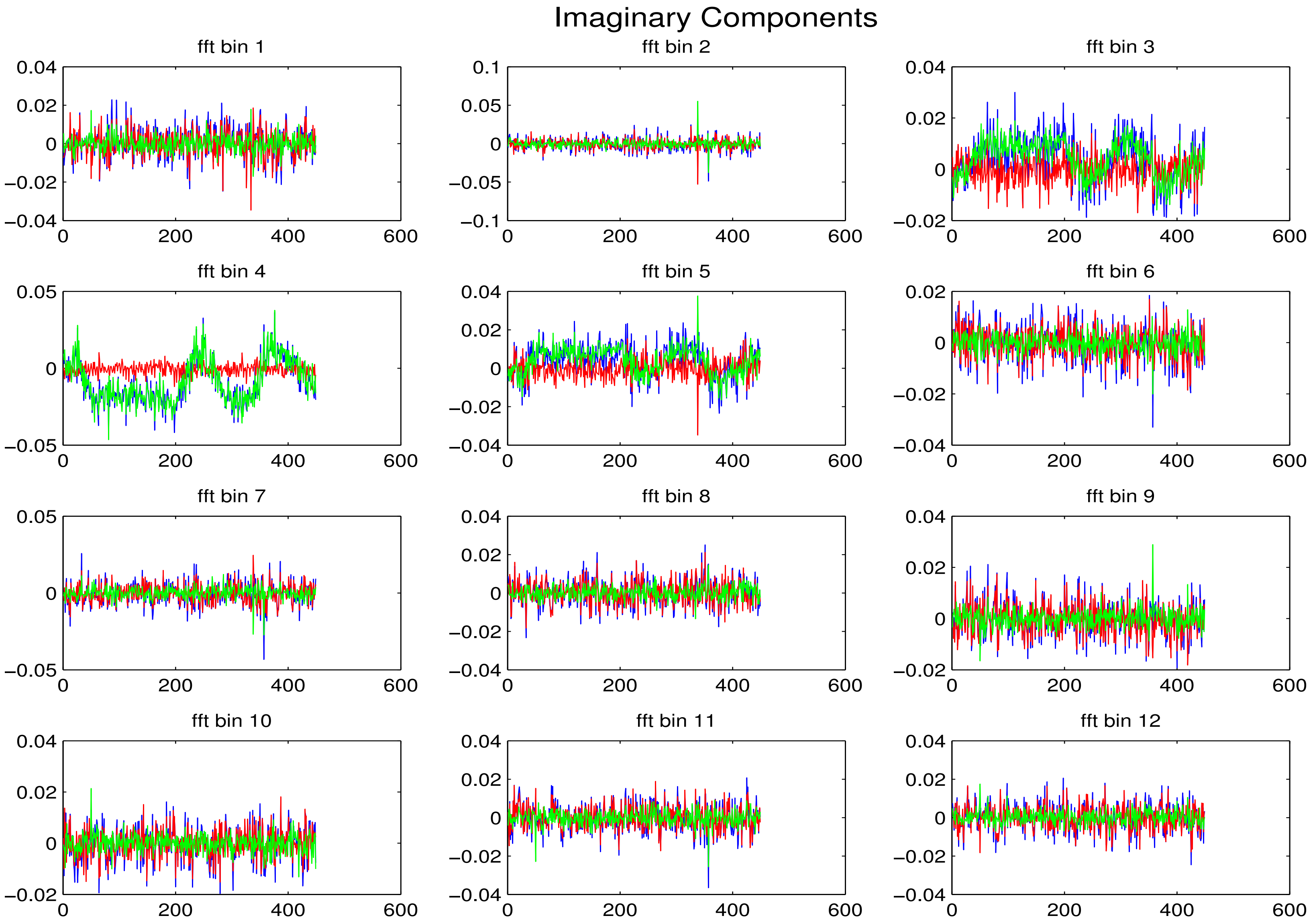


Figure 5.7: Observed fields in blue, remotely predicted fields in red, and residuals in green.

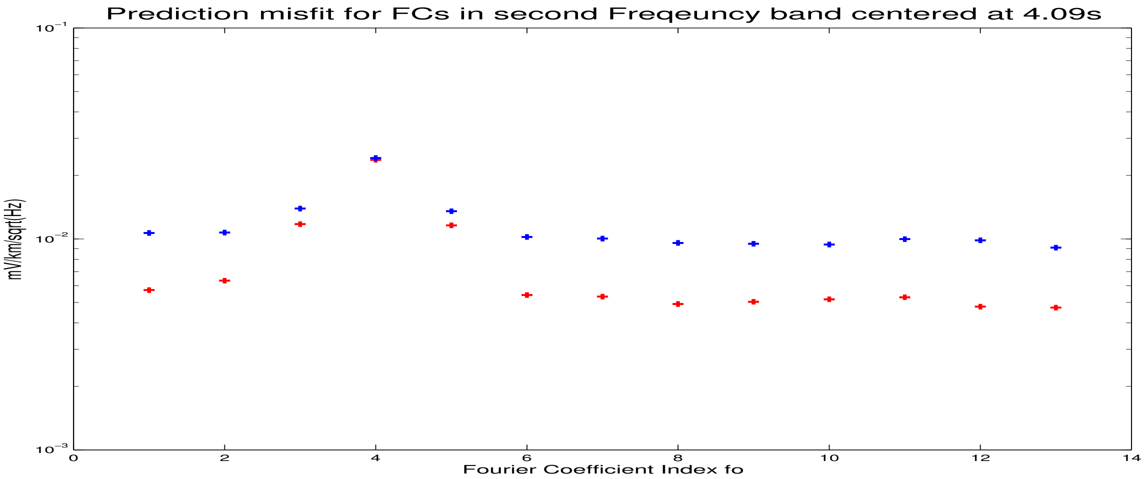


Figure 5.8: Misfit in a narrow frequency band. Observed FC amplitudes are shown in blue, and residual FC amplitudes are in red.

are present around the time of the earthquake, we generate an averaged transfer function over several weeks, and then apply this to the 163 day window around the earthquake in the next section.

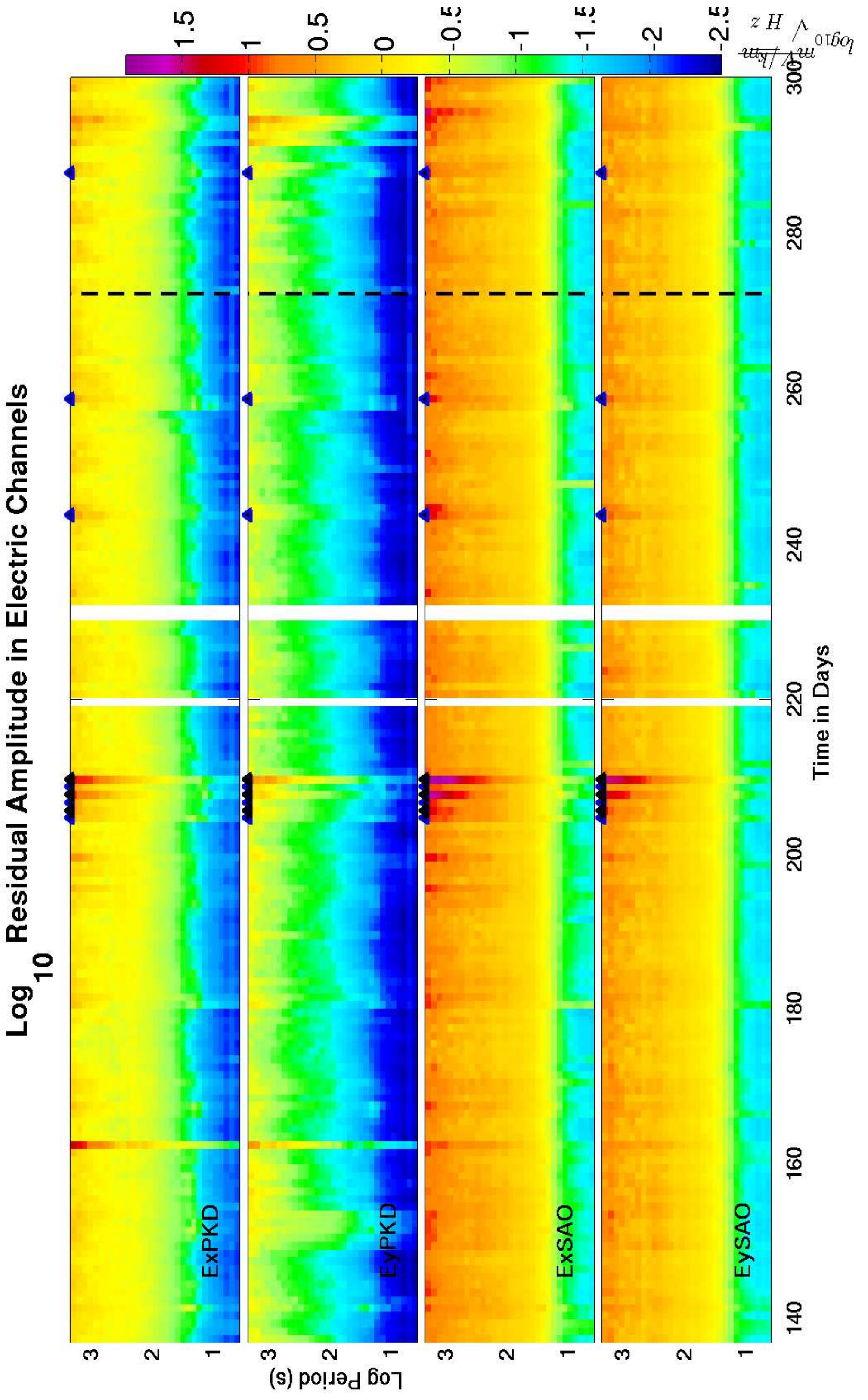


Figure 5.9: Residual amplitudes in $\log_{10}(\text{mV/kM}/\sqrt{Hz})$.

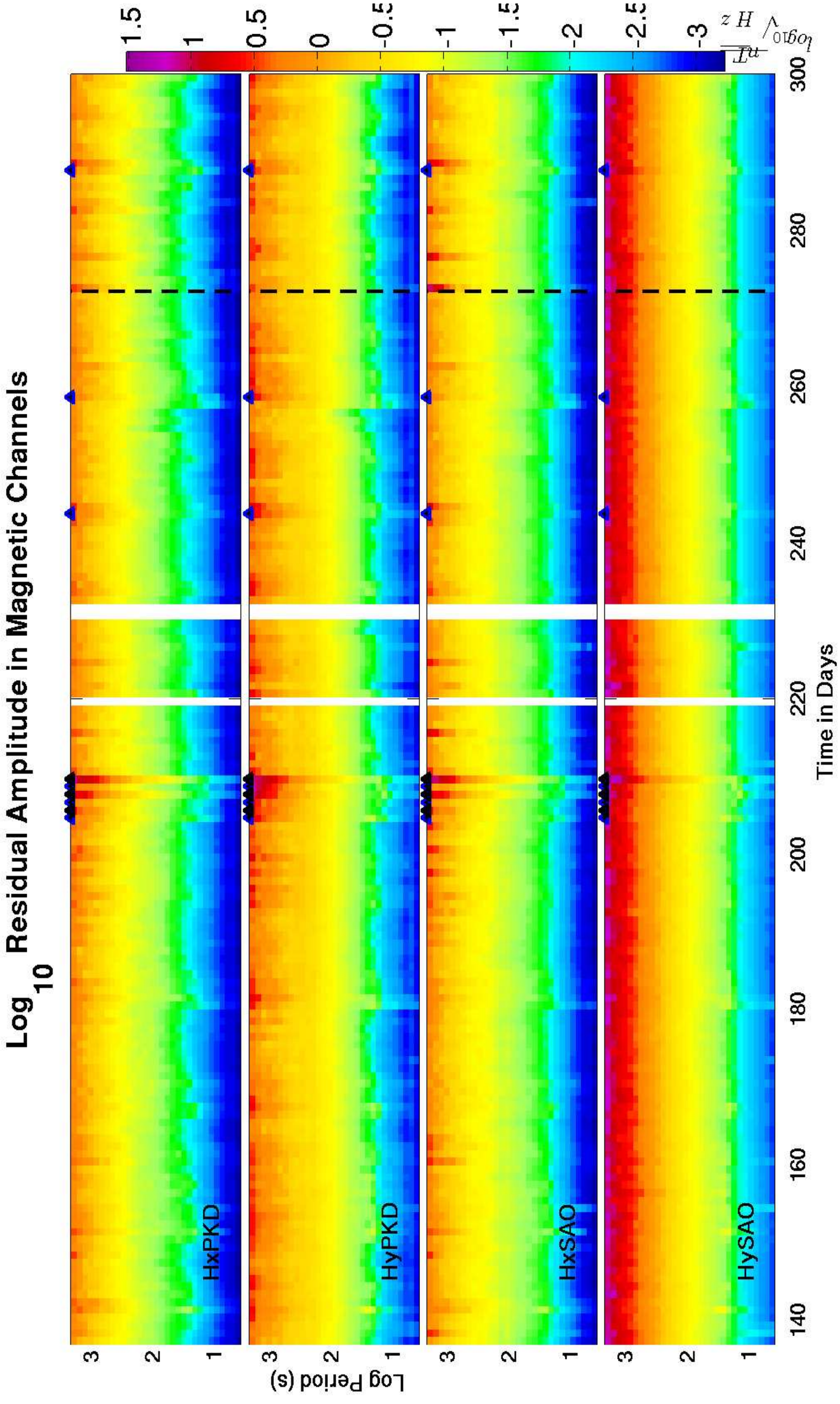


Figure 5.10: Residual amplitudes in $\log_{10}$(mV/kM/$\sqrt{Hz}$).

Chapter 6

Time of Day Residuals from Three-week Averaged TF

Plots of residuals similar to those shown in the last section are generated here, but with three key differences.

First, rather than calculating the TF directly from regression on the time-series of Fourier coefficients, the robust SDM is used to calculate the TF. This has the effect of stabilizing the TF estimates against outliers. Consider an arbitrary but fixed frequency bin centered at ω. We seek to predict the field values at from one group of channels by using another group. In this case, the groupings are Parkfield channels and Hollister channels, but the following steps would work for arbitrary channel groupings, provided the sets of channels have no intersection. Let one group of channels (for example the channels at PKD) be represented as the time series of Fourier coefficients **D**(t) (in this case **D**(t) has dimension 4xT), and similarly the other group as **M**(t) (in this case the channels at SAO). The channels can be ordered so that the SDM at the frequency ω, $\tilde{\mathbf{S}_\omega}$, hereafter called $\tilde{\mathbf{S}}$ is approximately

$$\tilde{\mathbf{S}} \approx \frac{1}{T}\left[\begin{array}{c|c} \mathbf{DD}^* & \mathbf{DM}^* \\ \hline \mathbf{M}^*\mathbf{D} & \mathbf{MM}^* \end{array}\right] = \frac{1}{T}\left[\begin{array}{c|c} \mathbf{S}_{1,1} & \mathbf{S}_{1,2} \\ \hline \mathbf{S}_{2,1} & \mathbf{S}_{2,2} \end{array}\right] \tag{6.1}$$

Where the approximate equality holds due to the robust SDM calculation. If the SDM were simply calculated by as a scaled covariance matrix equality would hold instead. By partitioning the SDM this way, the transfer function between the channels in **D** and the channels in **M** ca be calculated directly:

$$\mathbf{G} = \mathbf{S}_{1,2}\mathbf{S}_{2,2}{}^{-1} \tag{6.2}$$

This can be seen clearly if one lets **D** be a time series of only one channel. Then since one solves **d=GM** for **G** (in the absence of outliers) by ordinary least squares, i.e. $\mathbf{G}=\mathbf{dM}^*(\mathbf{MM}^*)^{-1}$ it is clear by direct substitution that the transfer function between $\mathbf{d}_i$(t) at PKD and all channels **M**(t) at SAO is exactly equal to that in 6.2. Equation 6.2 is slightly more general, as each row of **G** corresponds to a TF between one Parkfield channel, and all remote (Hollister) channels.

The second difference between the residuals in this section and those of the prior section is that the SDMs used to calculate the residuals are calculated over several time windows in the day. Specifically, the 24 hour day is partitioned into 13 non-overlapping, approximately equal sized windows, and the SDM is calculated for each of these 13 windows. Previously, for a given channel and frequency band, each pixel in the residual plot represented the channel's residual RMS amplitude over a whole day. In this case, there are 13 plots for each channel, and each pixel represents the residual RMS amplitude over an $\sim$ 2hr section of time.

Finally, the TFs are not calculated for each day of the 163-day section, but rather only over the first three weeks. Thus, for each of the w=1..13 windows of a day *d*, we get an estimate of the SDM at that day: $\mathbf{S}_{w,d}$. We then simply take:

$$\mathbf{S}_w = \sum_{d=1}^{21} \mathbf{S}_{w,d} \tag{6.3}$$

as the estimate for the robust SDM. Figures 6.1 through 6.13 depict the electric field residuals, and Figures 6.14 through 6.26 show the magnetic residuals.

A number of noteworthy features appear in these plots.

Figures 6.7 and 6.8 show an interesting weekly periodicity, characterized by a broadband decrease in amplitude of the residuals on Saturdays and Sundays. UT time is 7 hours ahead of PST in summer, thus these weekly periods are showing up between 4 and 10am local time. It is likely that much of this effect is due to the BART DC electric train. More evidence of this is shown in the next section. This effect however appears to be too broad in frequency range to be solely due to the BART, and may also be related to other EM phenomena. Ex at SAO in Figure 6.7 shows a clear decrease in residual power over all frequencies not in the dead band.

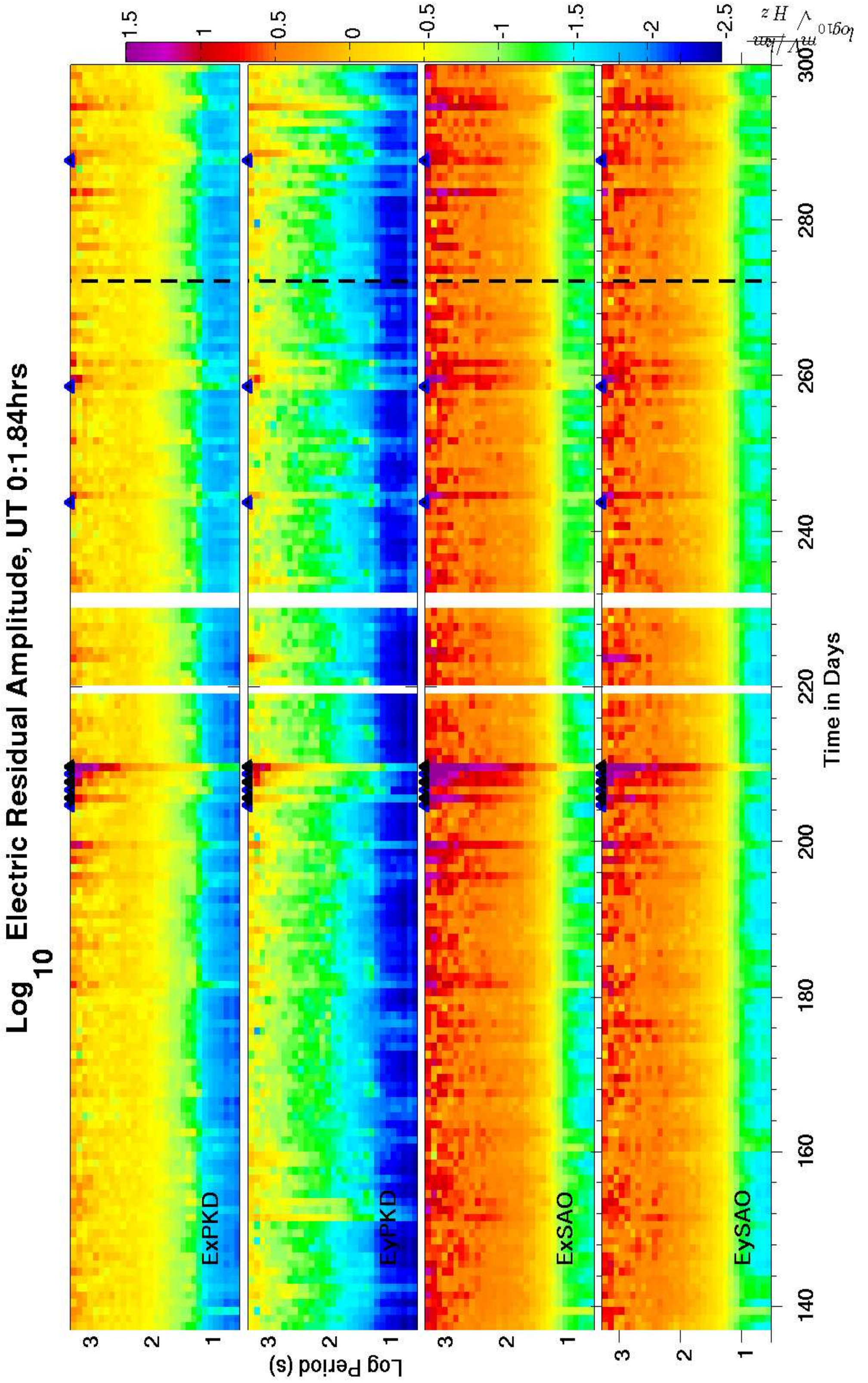


Figure 6.1: Residual amplitudes in $\log_{10}$(mV/kM/$\sqrt{Hz}$).

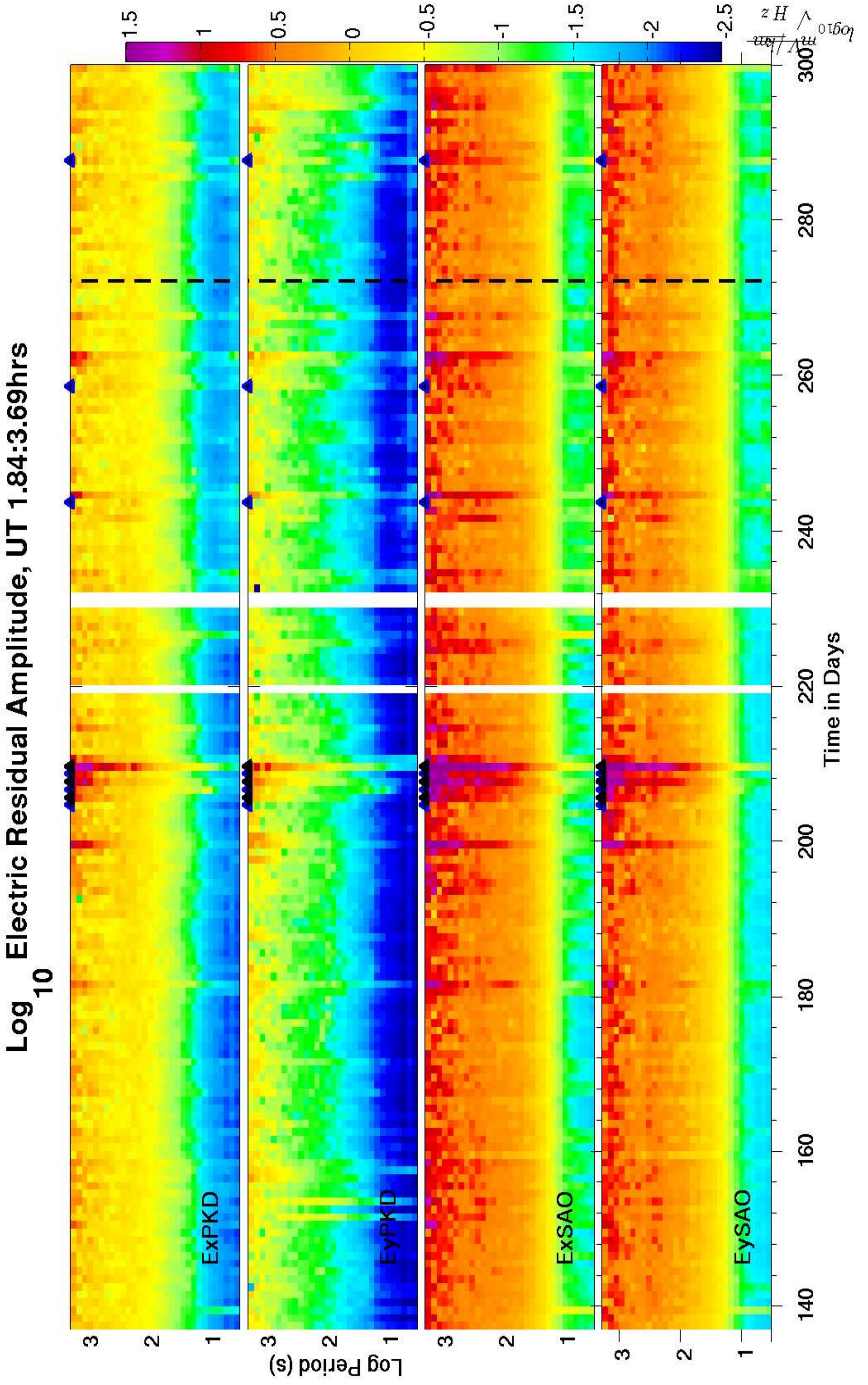


Figure 6.2: Residual amplitudes in $\log_{10}$(mV/kM/$\sqrt{Hz}$).

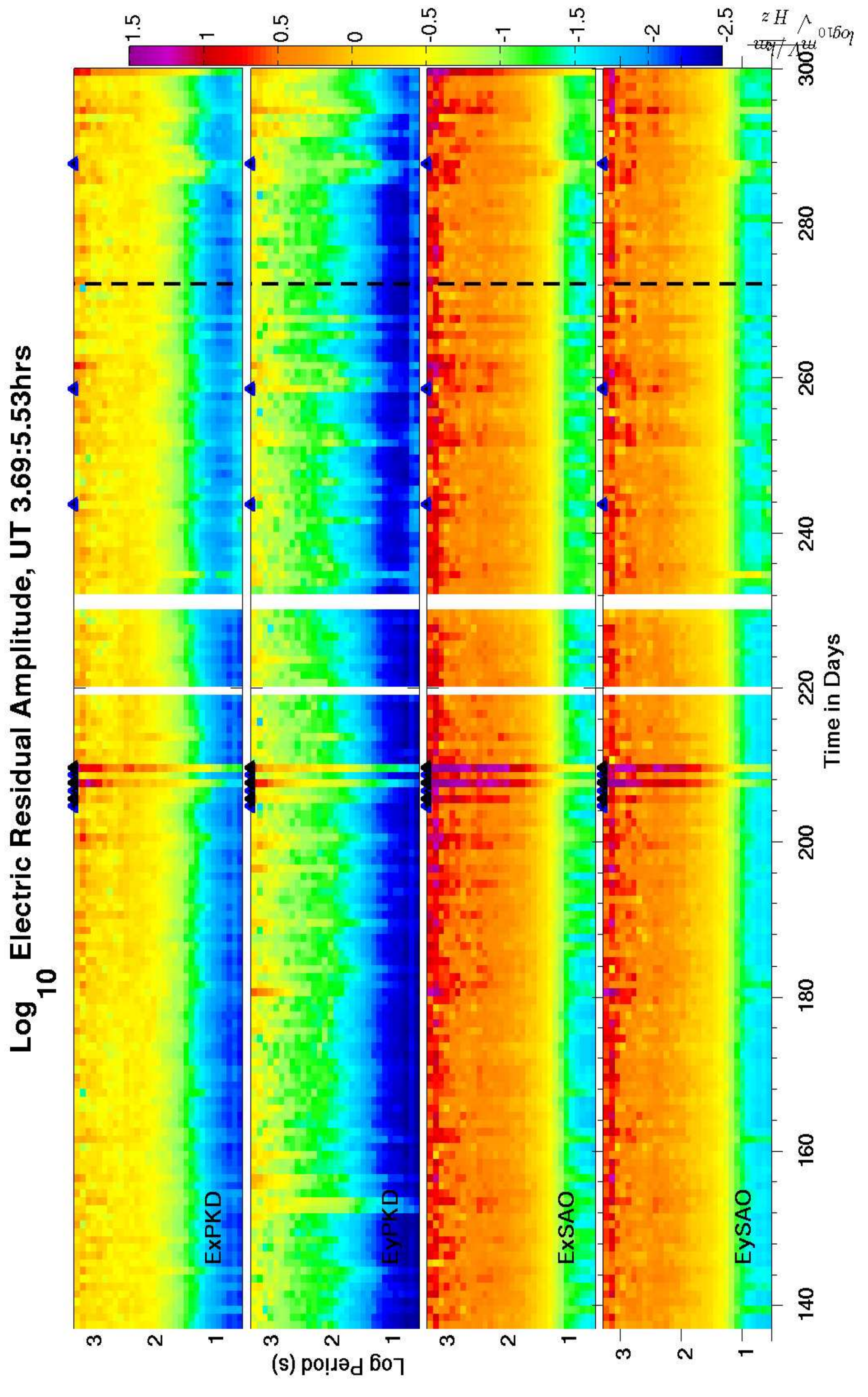


Figure 6.3: Residual amplitudes in $\log_{10}(\text{mV/kM}/\sqrt{Hz})$.

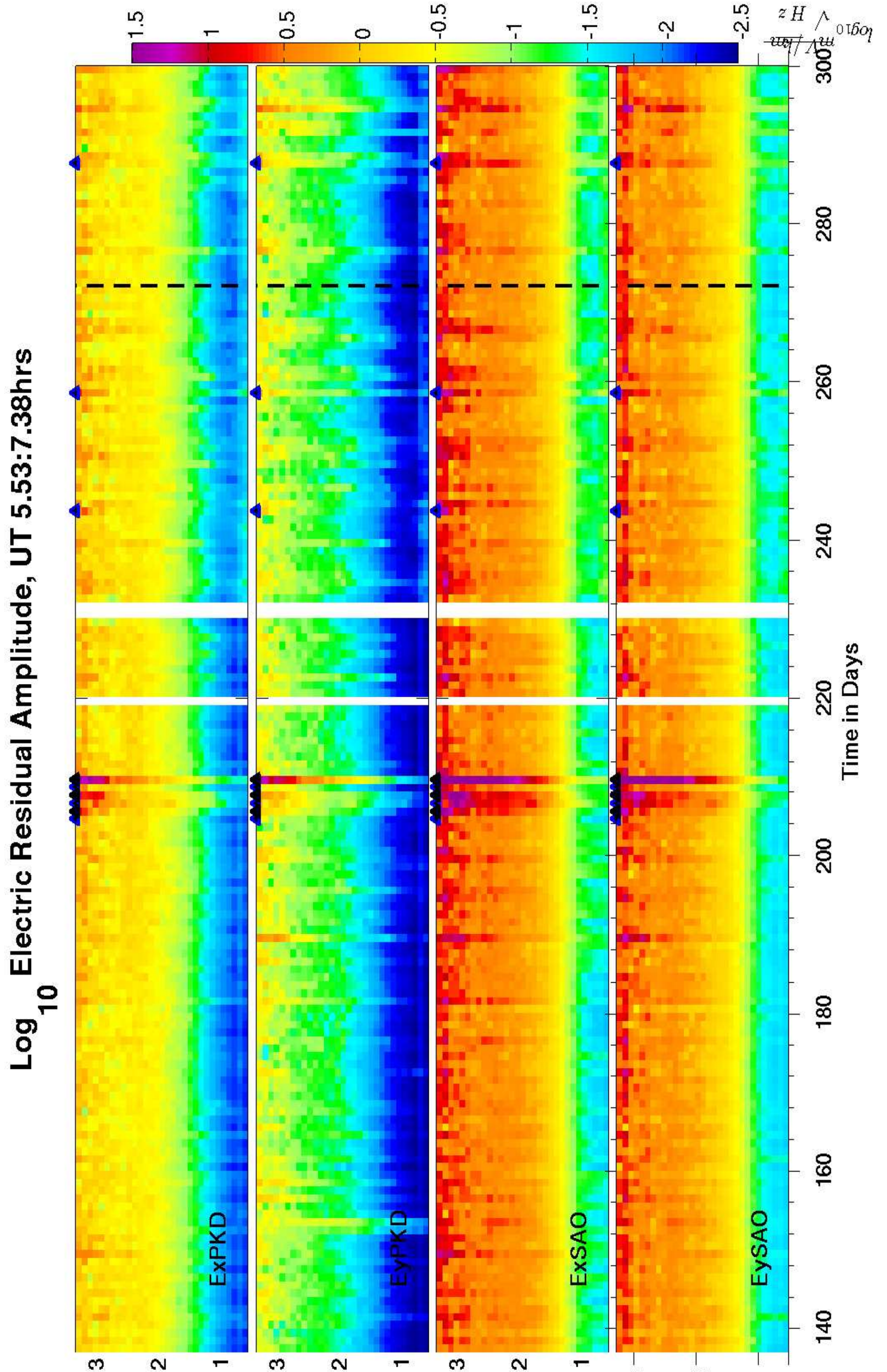


Figure 6.4: Residual amplitudes in $\log_{10}(\mathrm{mV/kM}/\sqrt{Hz})$.

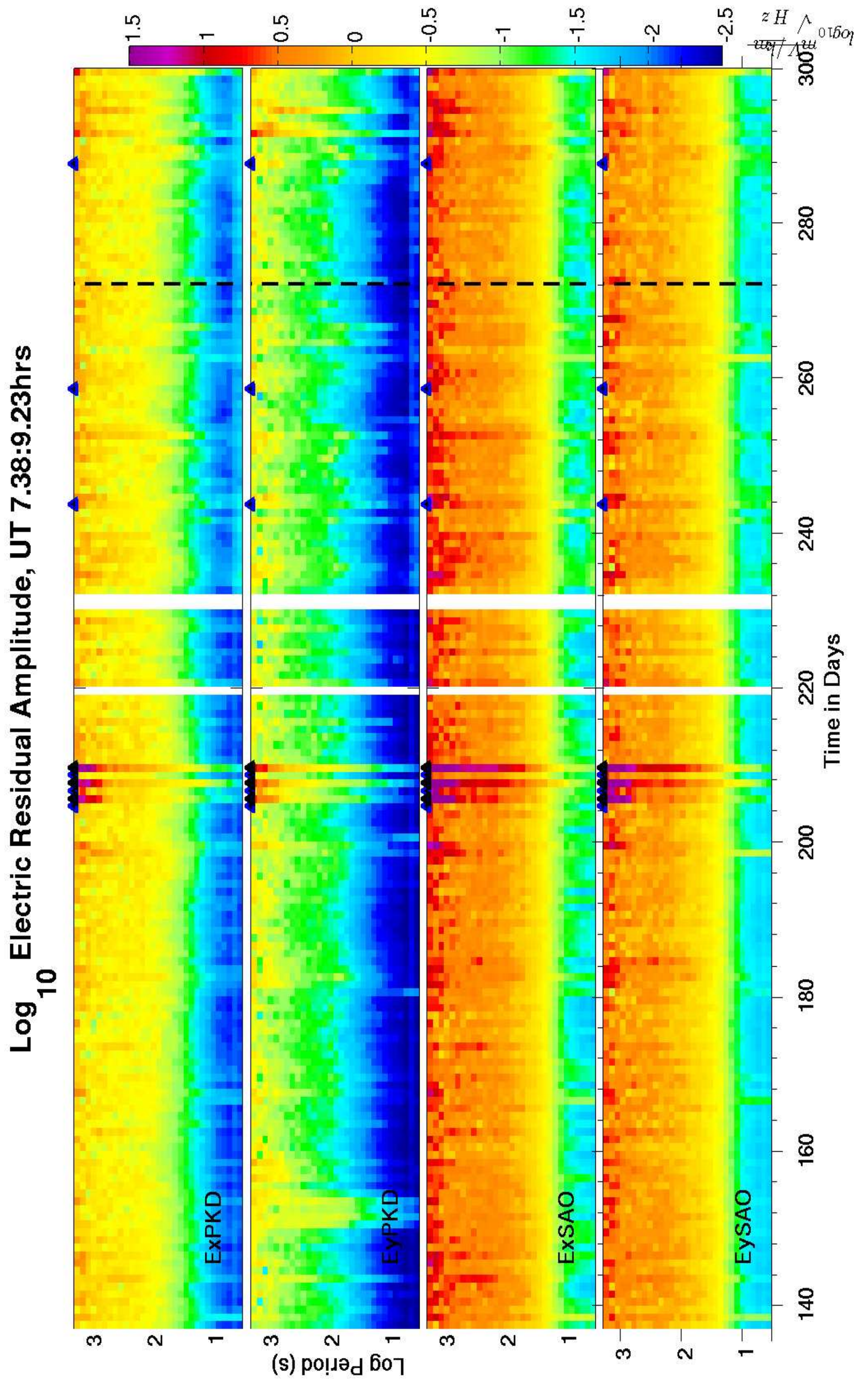


Figure 6.5: Residual amplitudes in $\log_{10}$(mV/kM/$\sqrt{Hz}$).

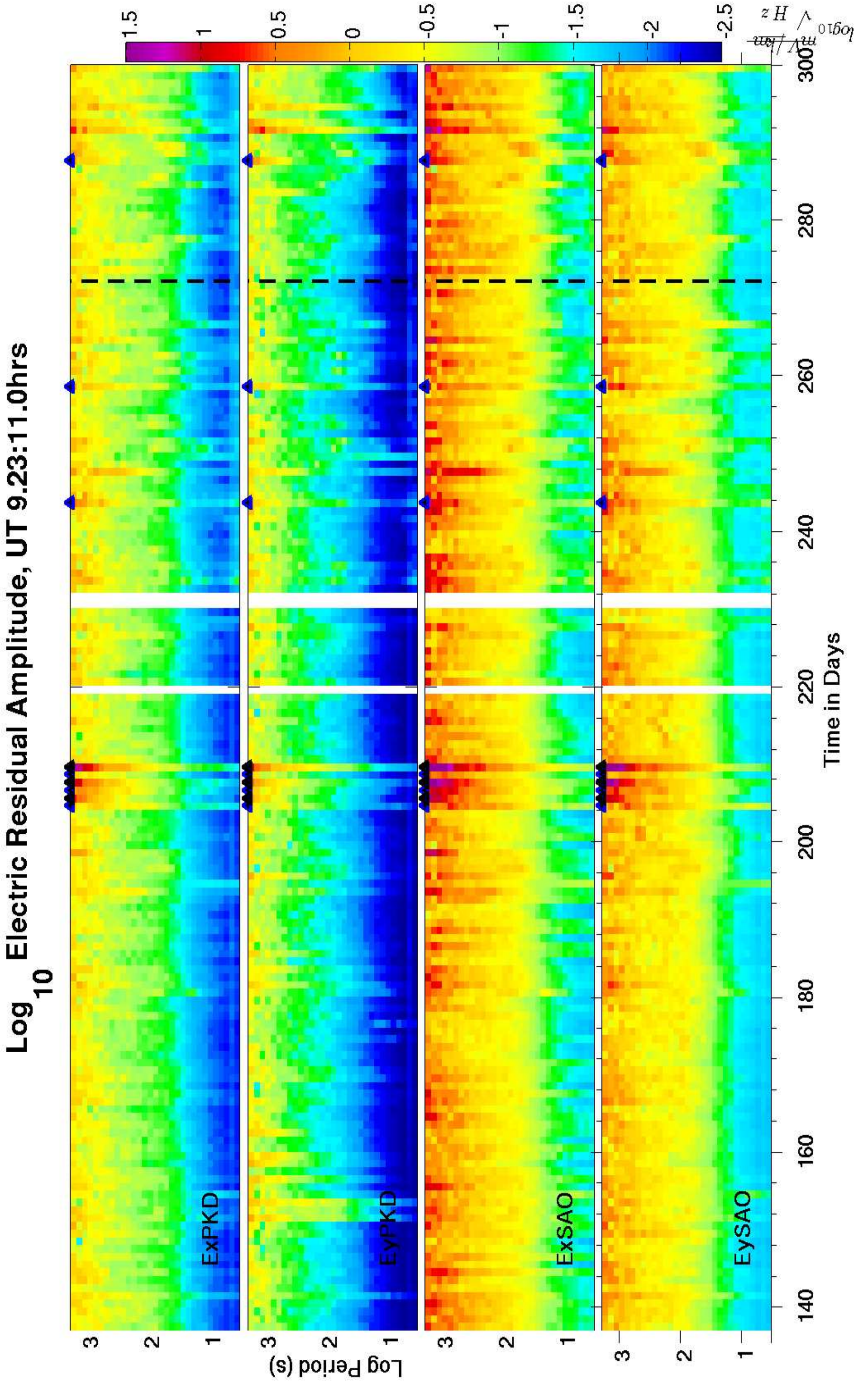


Figure 6.6: Residual amplitudes in $\log_{10}(\mathrm{mV/kM}/\sqrt{Hz})$.

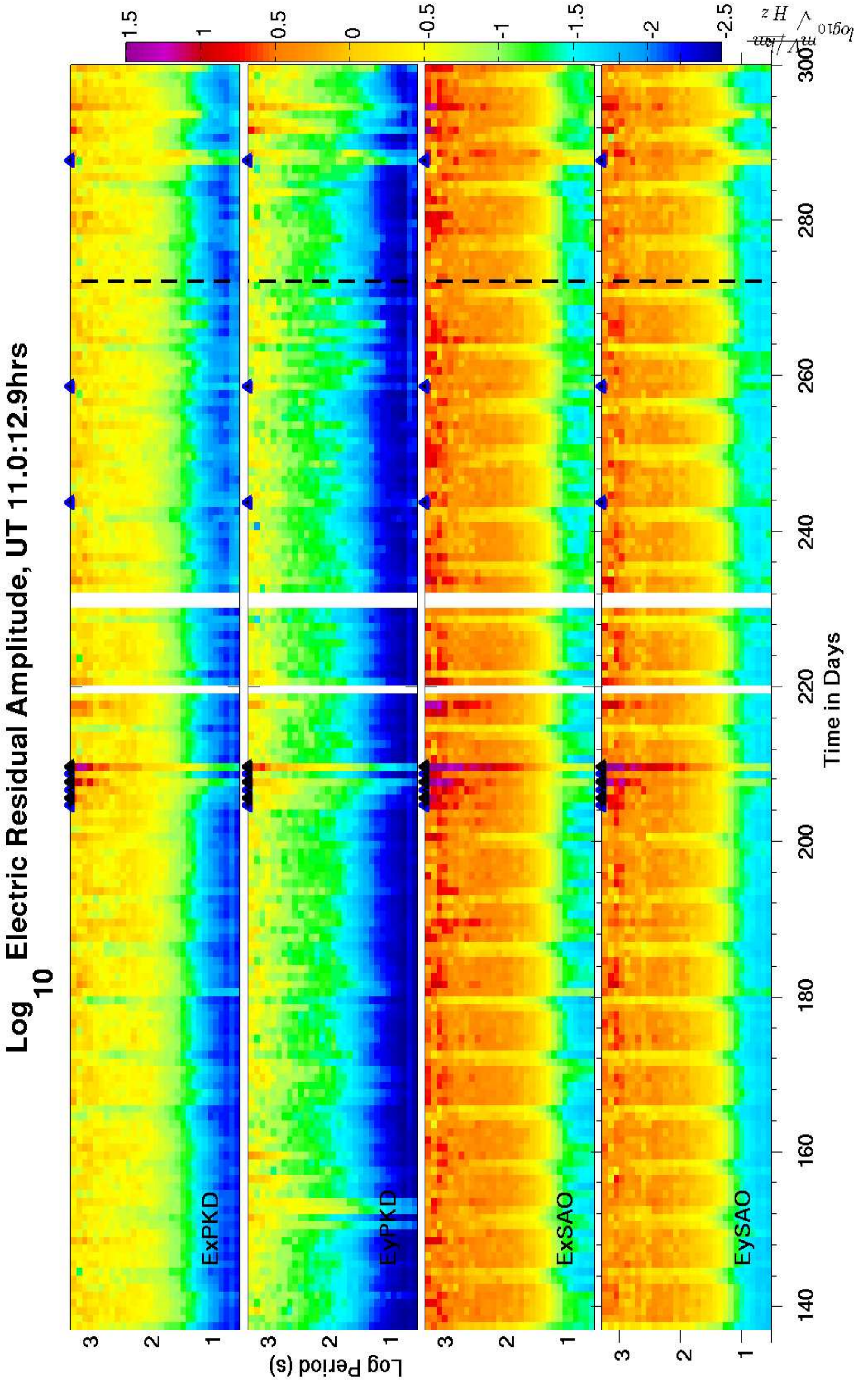


Figure 6.7: Residual amplitudes in $\log_{10}(\mathrm{mV/kM}/\sqrt{Hz})$.

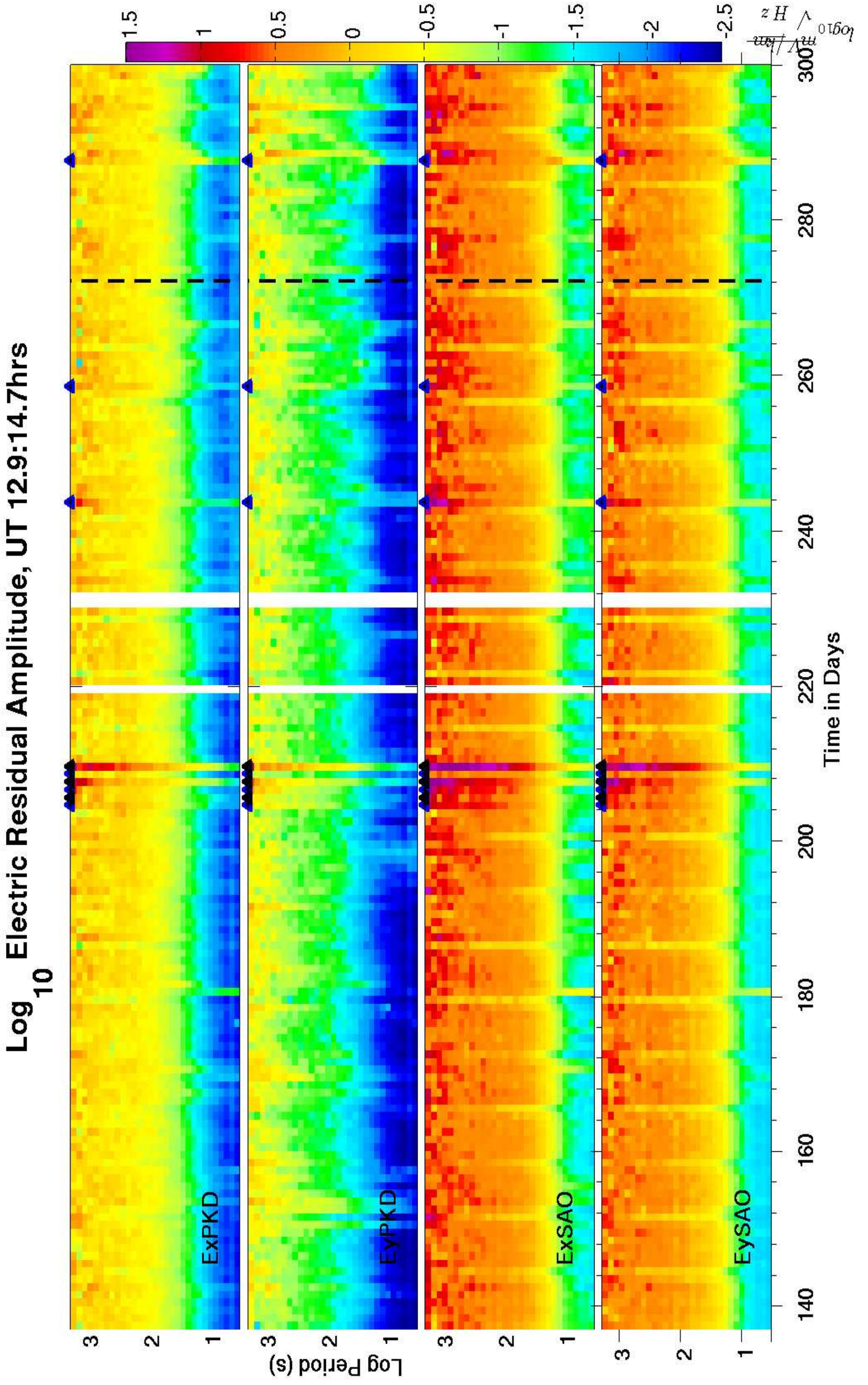


Figure 6.8: Residual amplitudes in $\log_{10}$(mV/kM/$\sqrt{Hz}$).

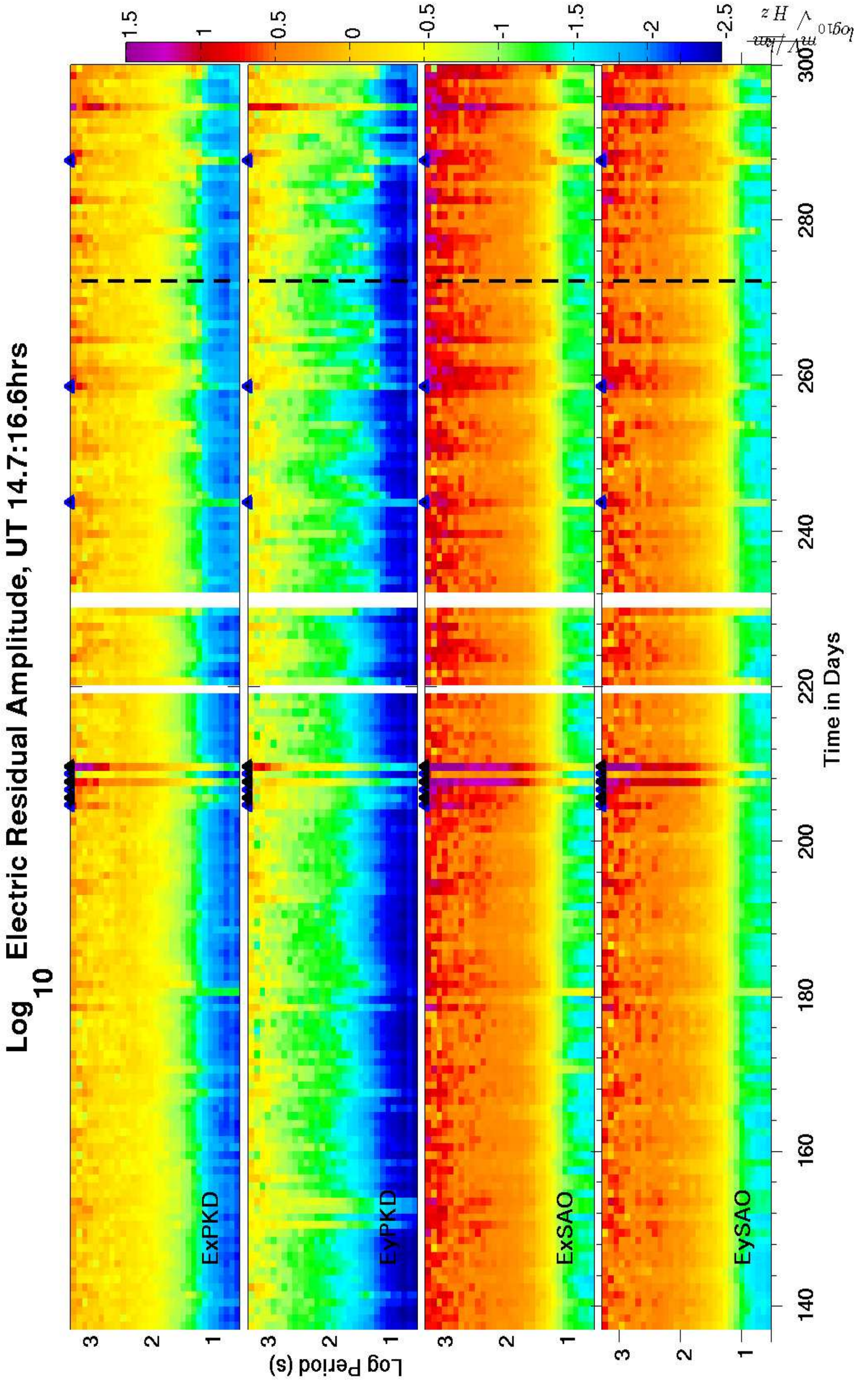


Figure 6.9: Residual amplitudes in $\log_{10}(\mathrm{mV/kM}/\sqrt{Hz})$.

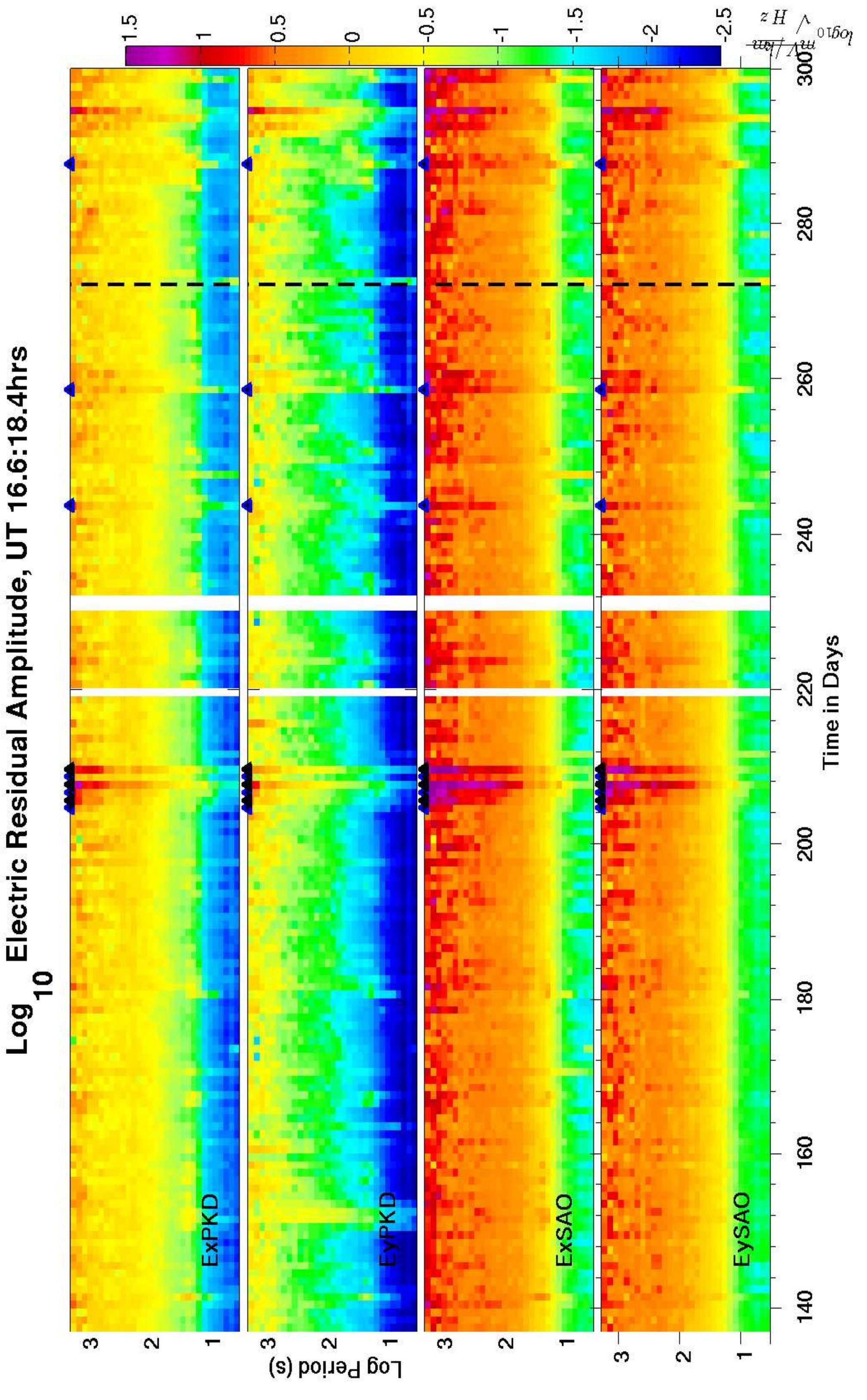


Figure 6.10: Residual amplitudes in $\log_{10}(\mathrm{mV/kM}/\sqrt{Hz})$.

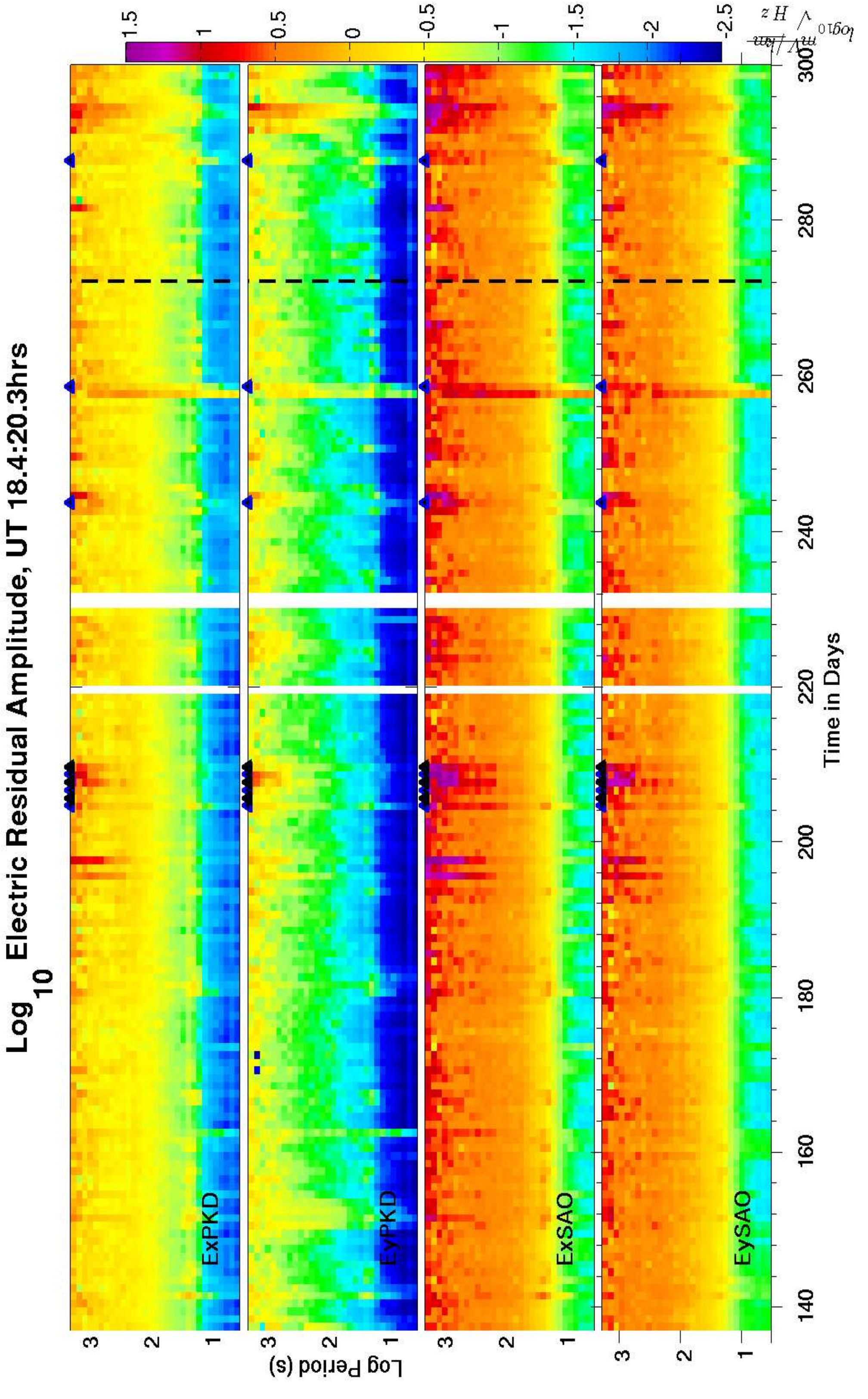


Figure 6.11: Residual amplitudes in $\log_{10}$(mV/kM/$\sqrt{Hz}$).

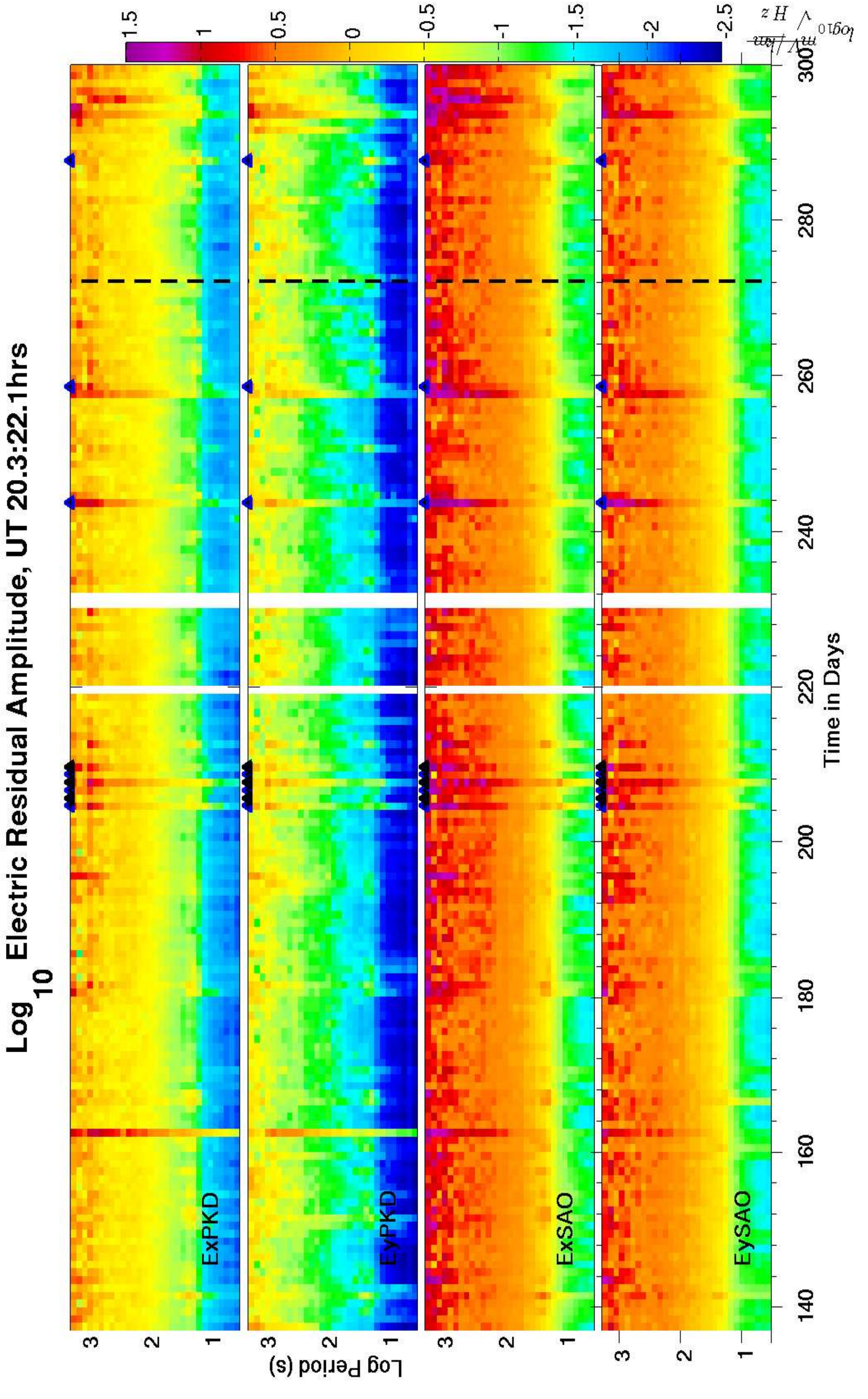


Figure 6.12: Residual amplitudes in $\log_{10}(\text{mV/kM}/\sqrt{Hz})$.

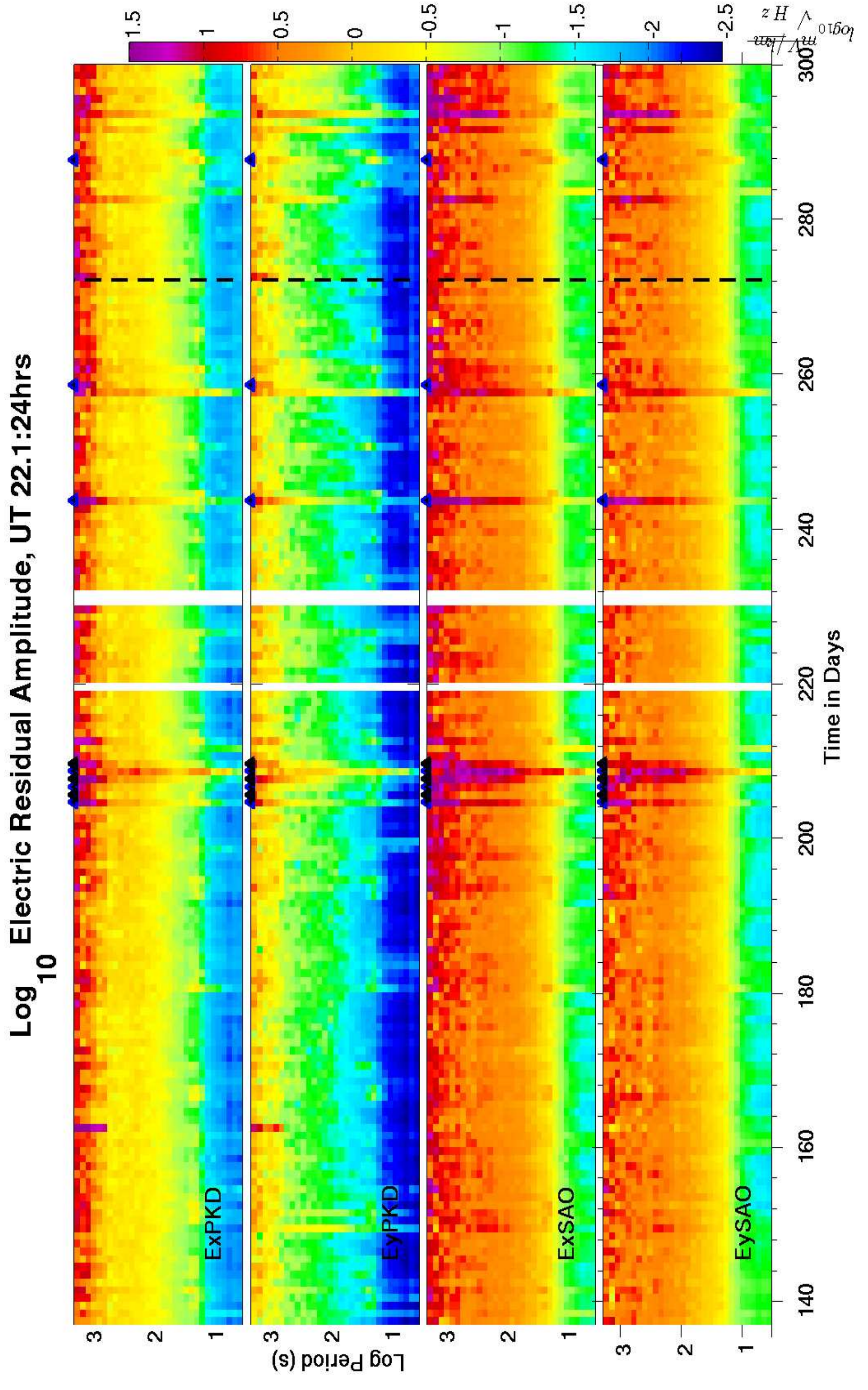


Figure 6.13: Residual amplitudes in $\log_{10}$(mV/kM/$\sqrt{Hz}$).

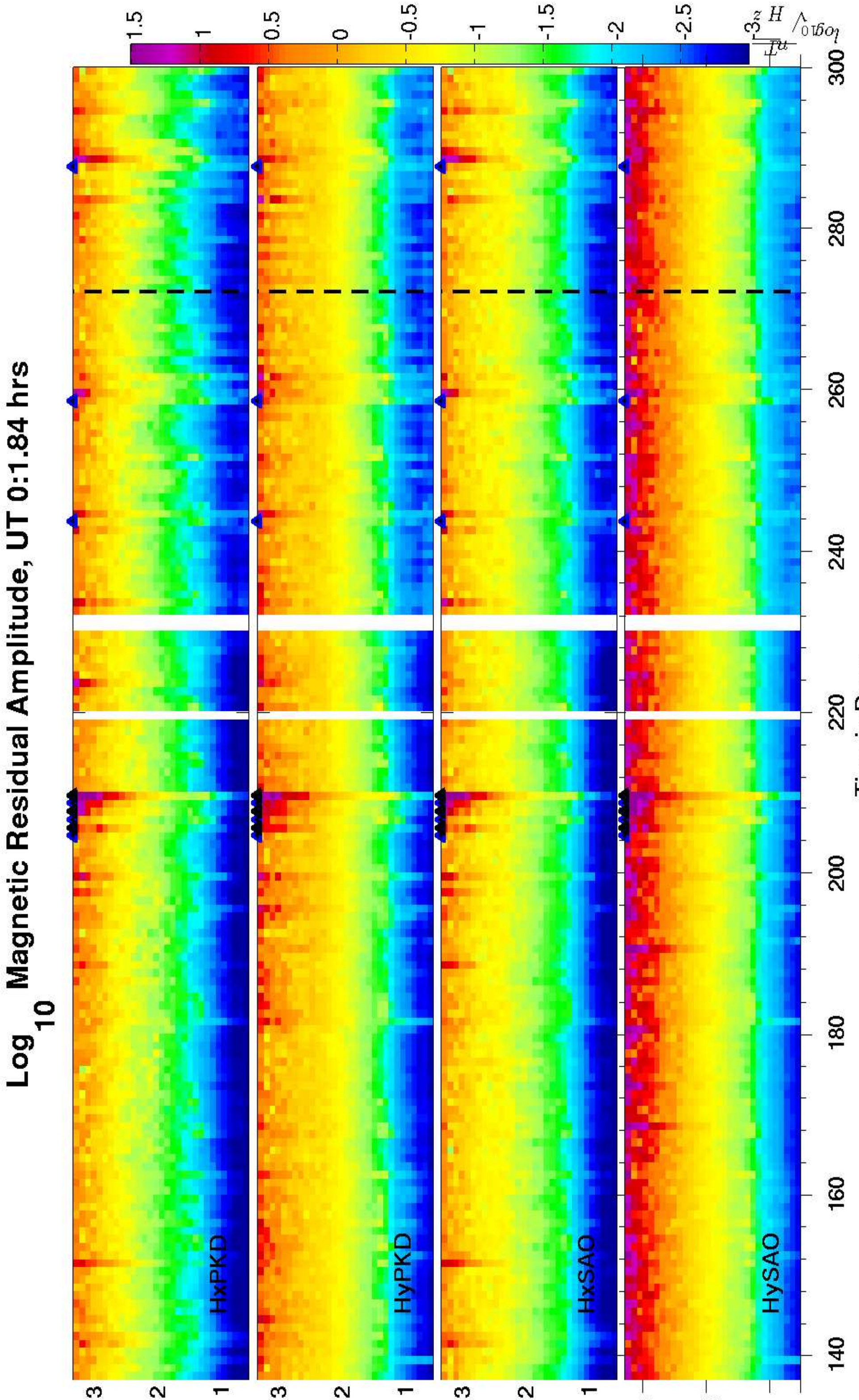


Figure 6.14: Residual amplitudes in $\log_{10}(\mathrm{nT}/\sqrt{Hz})$.

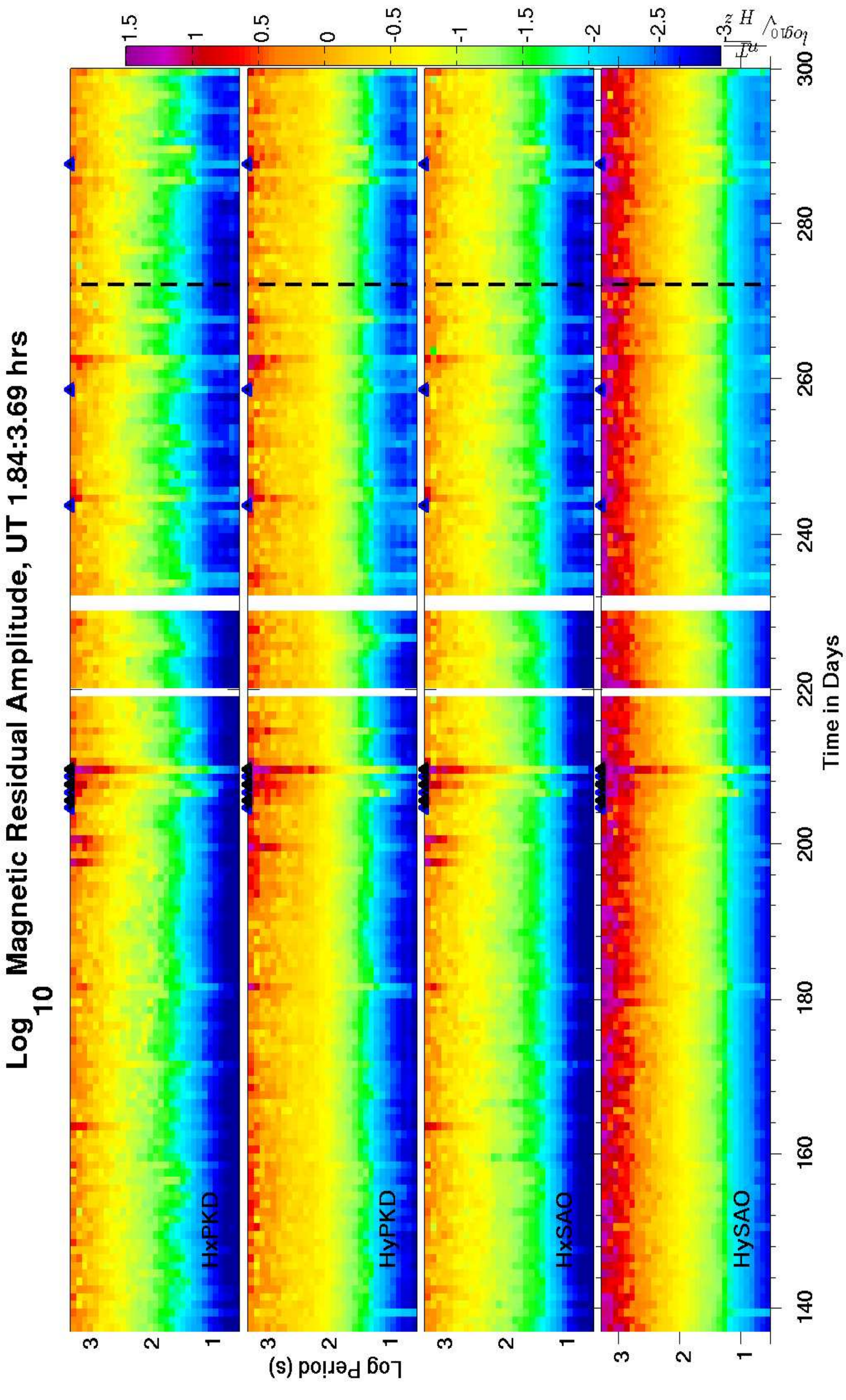


Figure 6.15: Residual amplitudes in $\log_{10}(\mathrm{nT}/\sqrt{Hz})$.

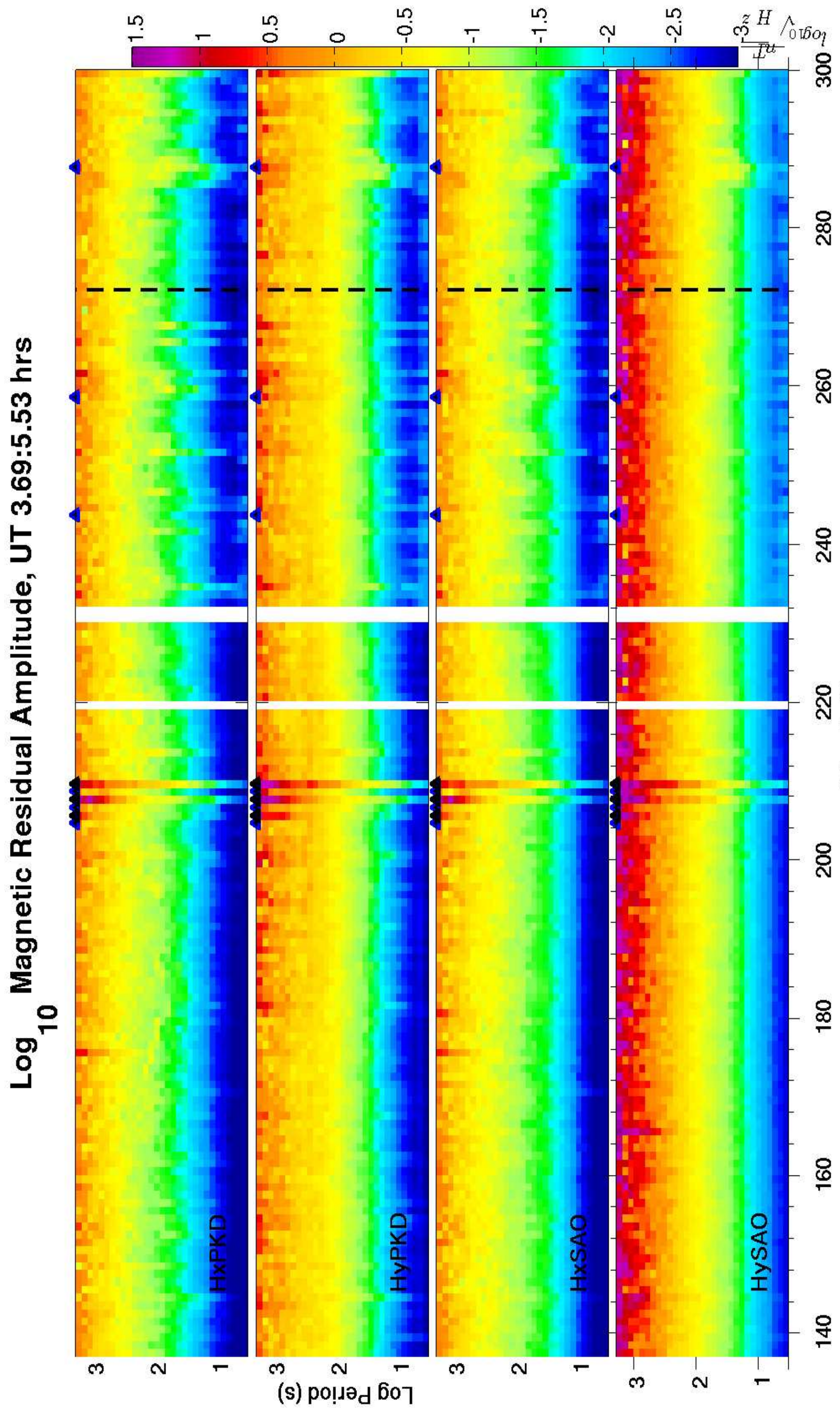


Figure 6.16: Residual amplitudes in $\log_{10}(\mathrm{nT}/\sqrt{Hz})$.

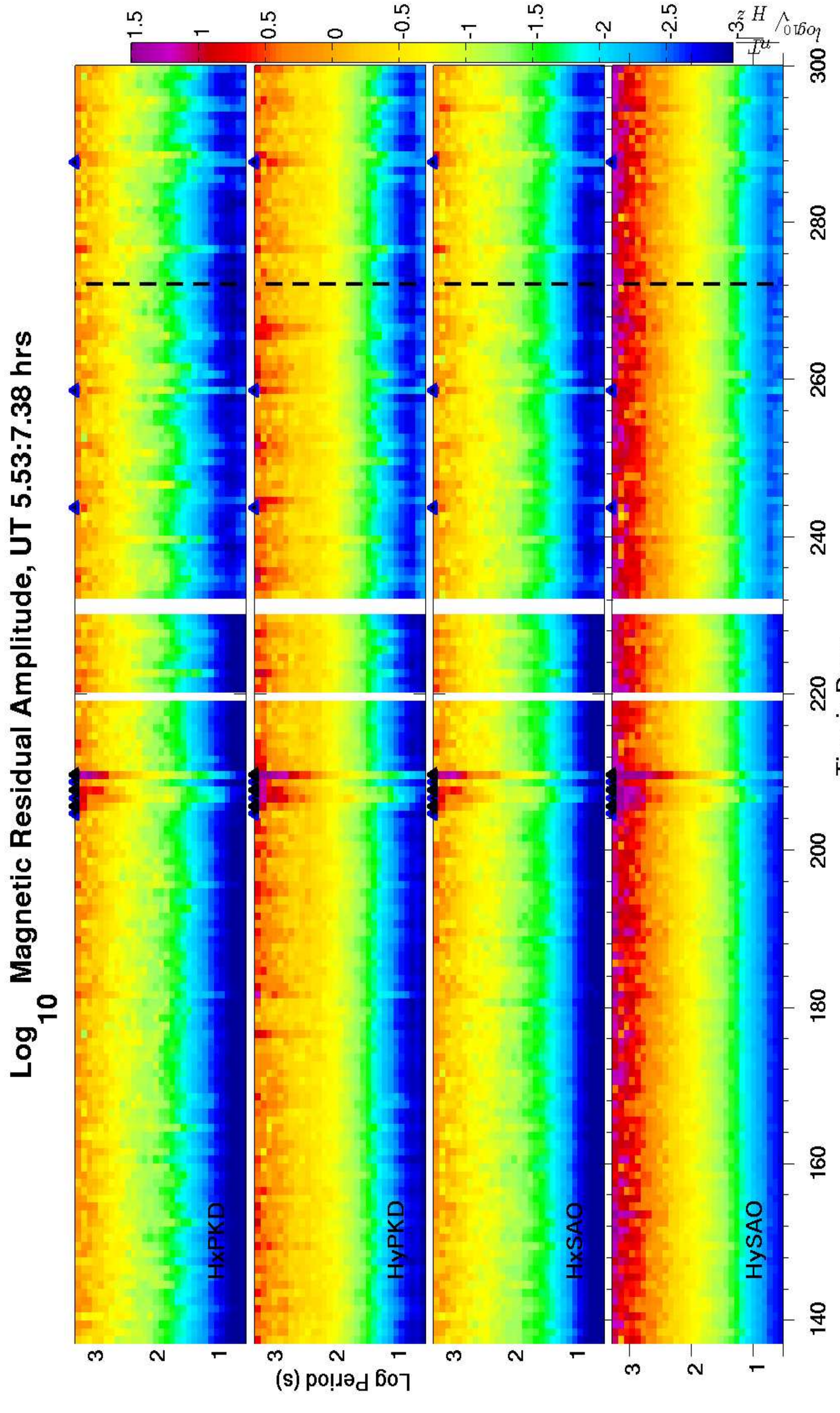


Figure 6.17: Residual amplitudes in $\log_{10}(\mathrm{nT}/\sqrt{Hz})$.

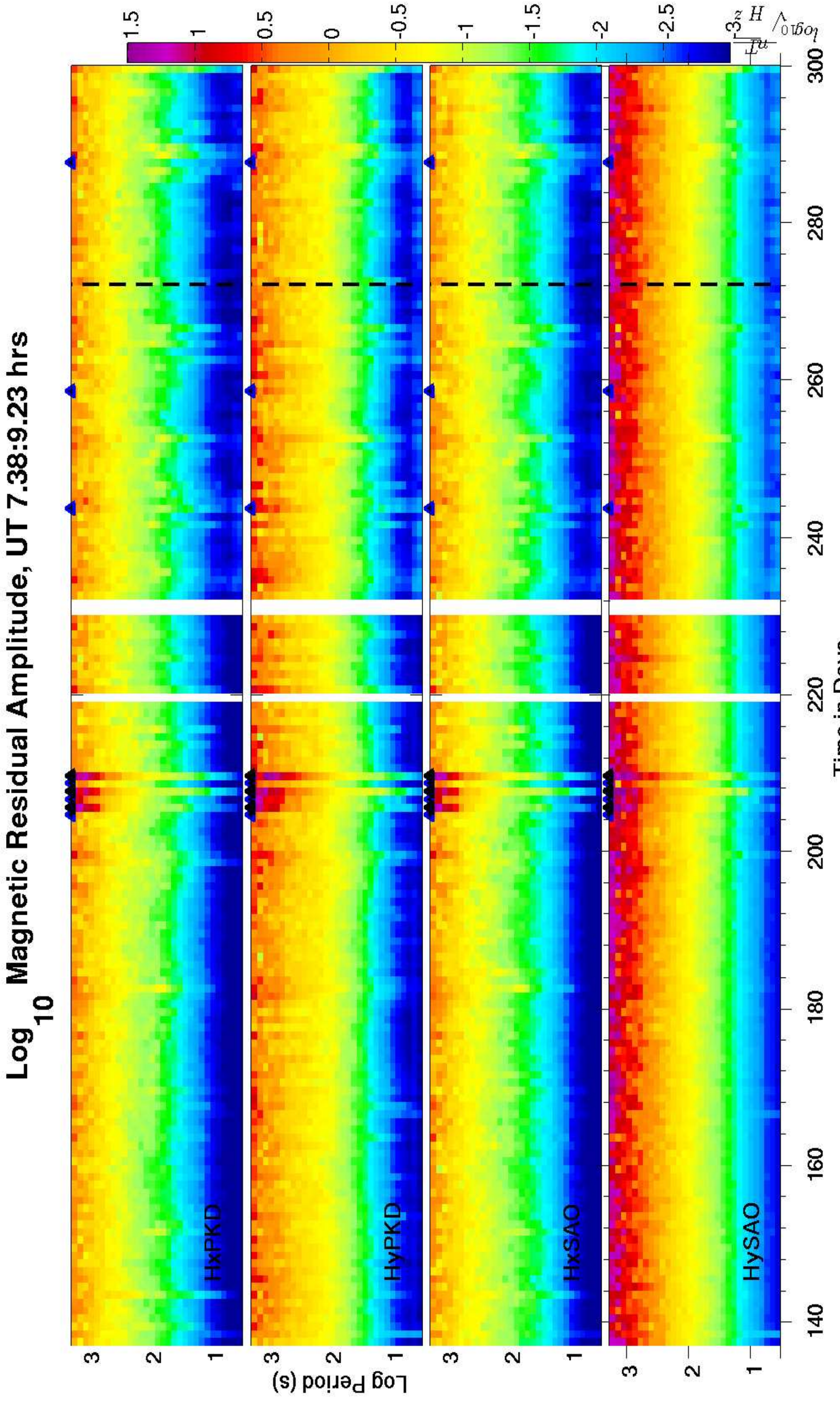


Figure 6.18: Residual amplitudes in $\log_{10}(\mathrm{nT}/\sqrt{Hz})$.

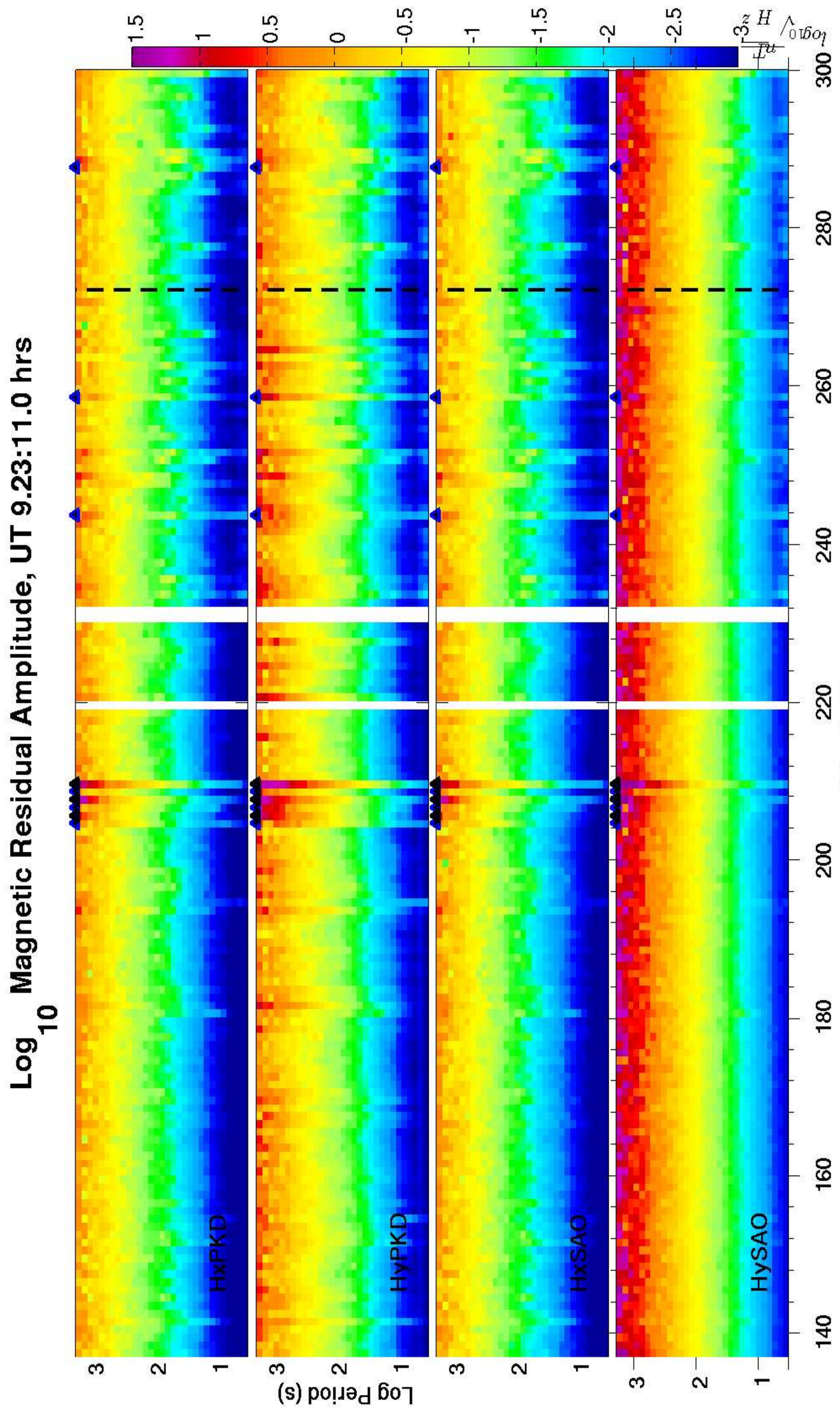


Figure 6.19: Residual amplitudes in $\log_{10}(\mathrm{nT}/\sqrt{Hz})$.

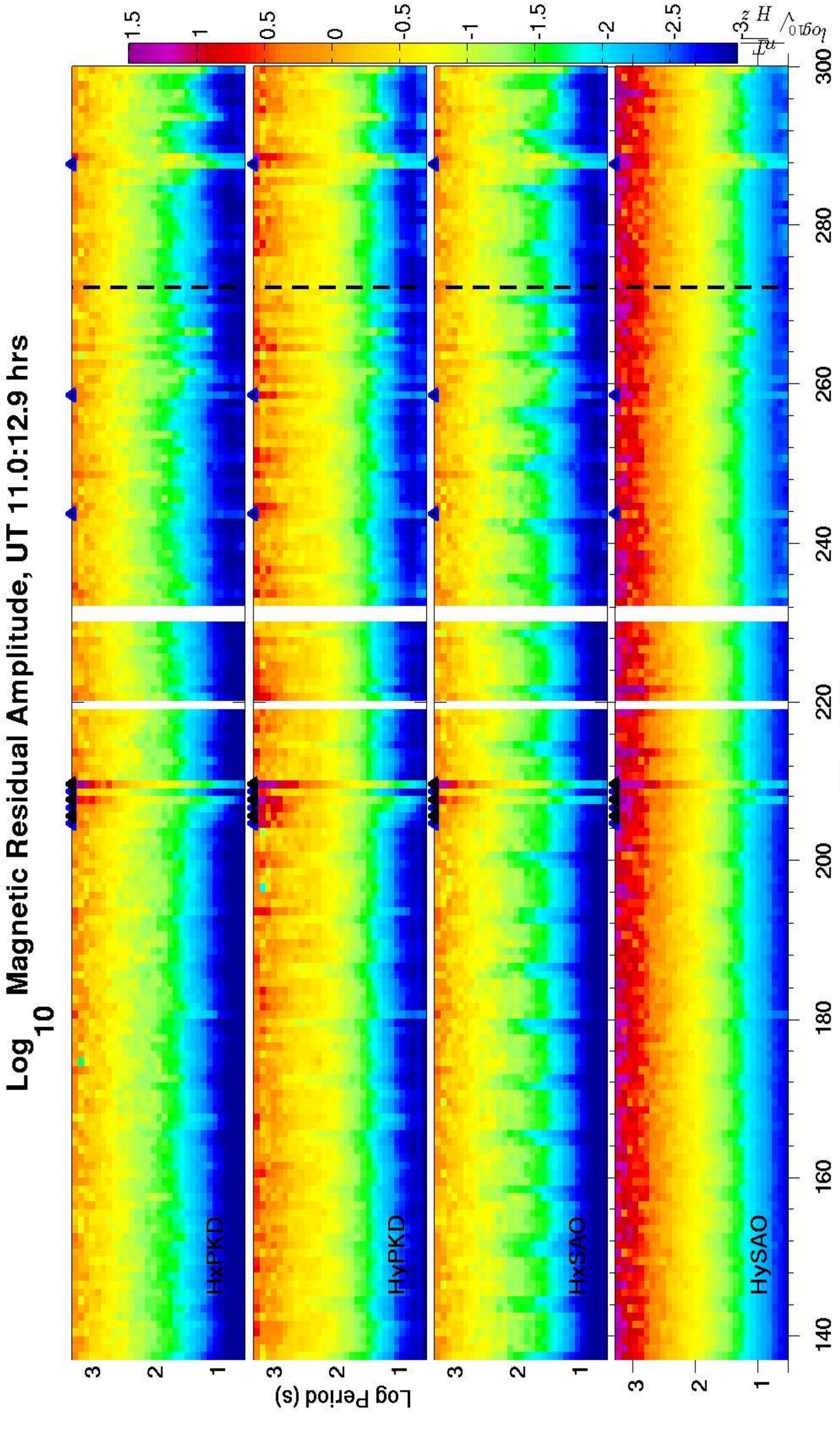


Figure 6.20: Residual amplitudes in $\log_{10}(\mathrm{nT}/\sqrt{Hz})$.

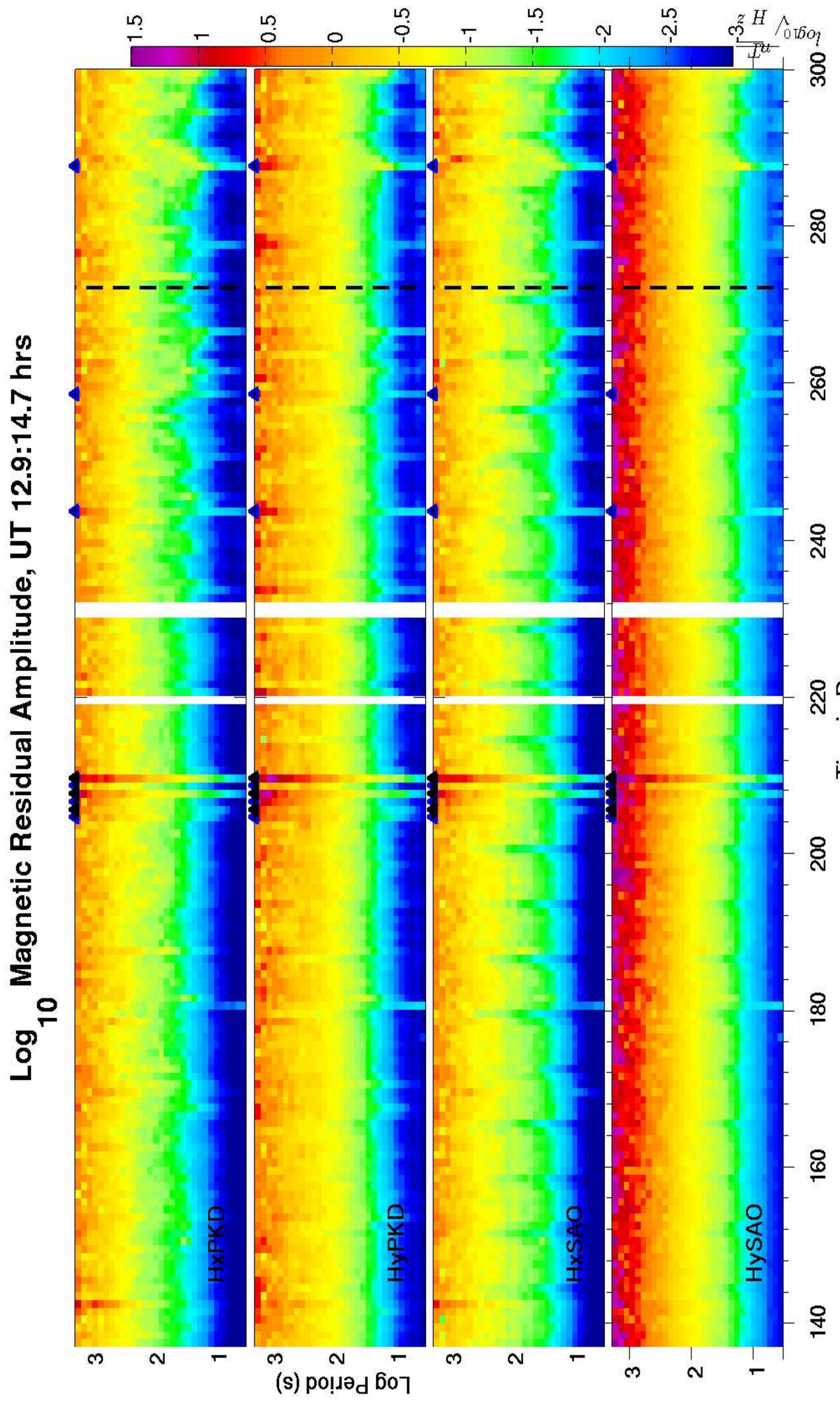


Figure 6.21: Residual amplitudes in $\log_{10}(\mathrm{nT}/\sqrt{Hz})$.

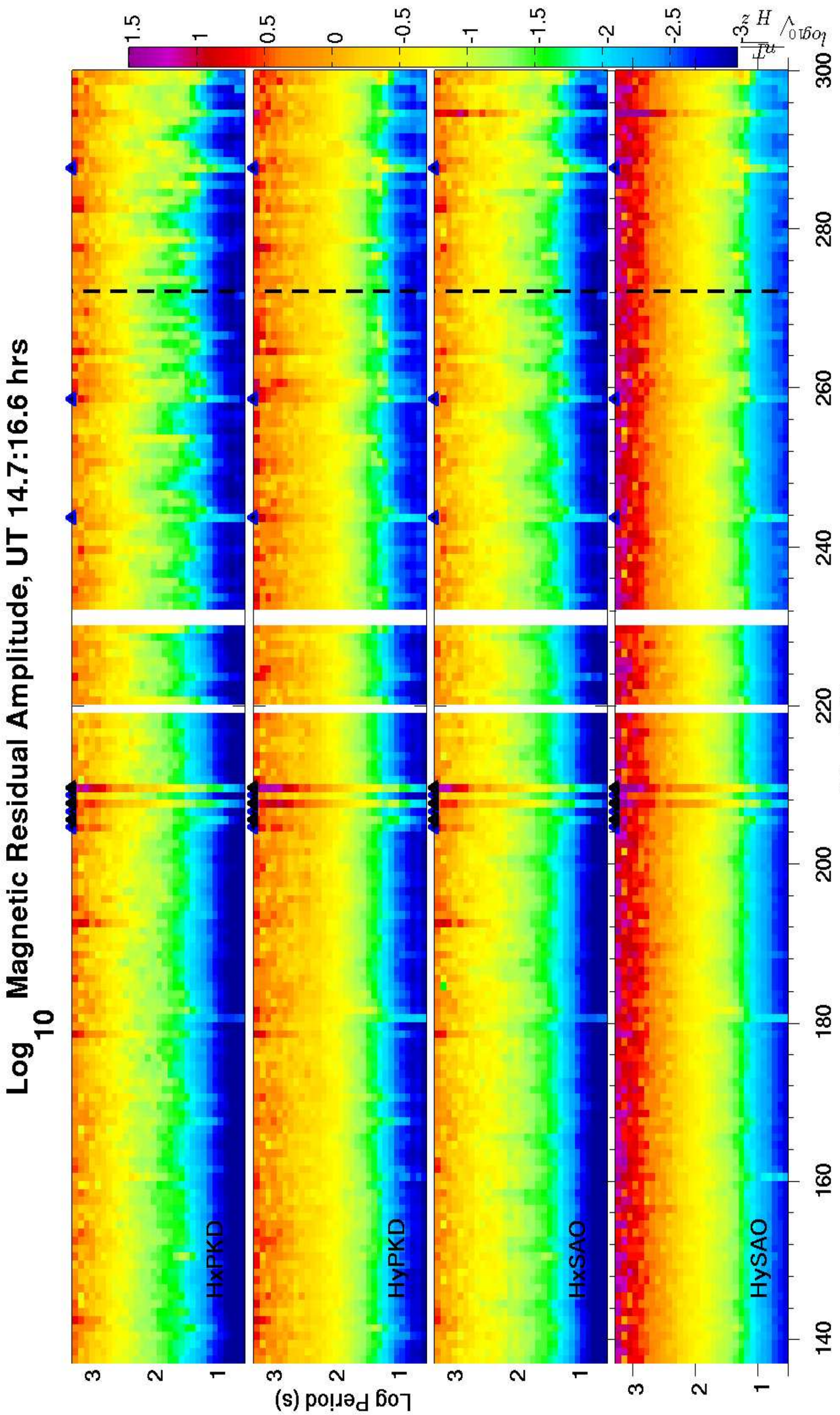


Figure 6.22: Residual amplitudes in $\log_{10}$(nT/$\sqrt{Hz}$).

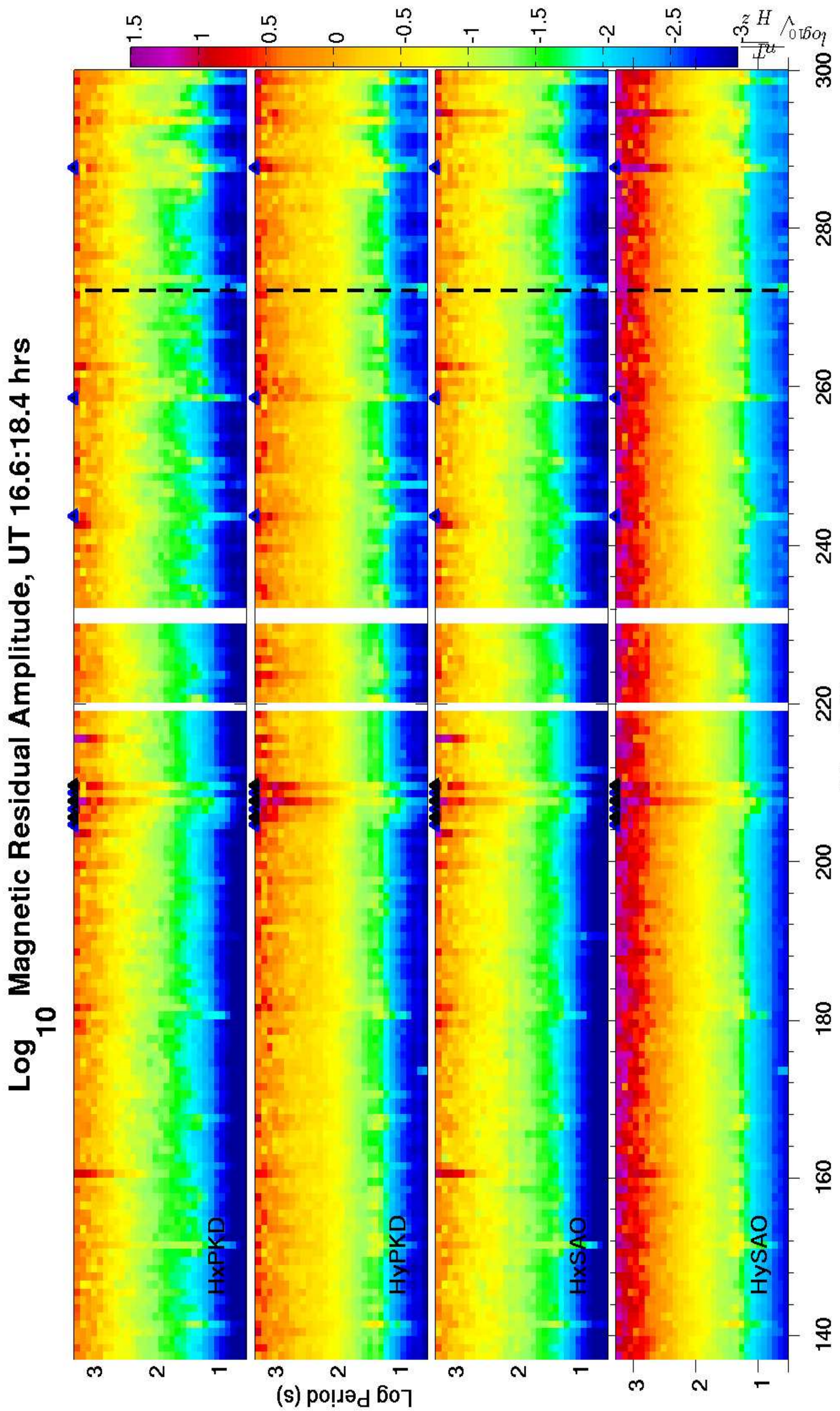


Figure 6.23: Residual amplitudes in $\log_{10}(\mathrm{nT}/\sqrt{Hz})$.

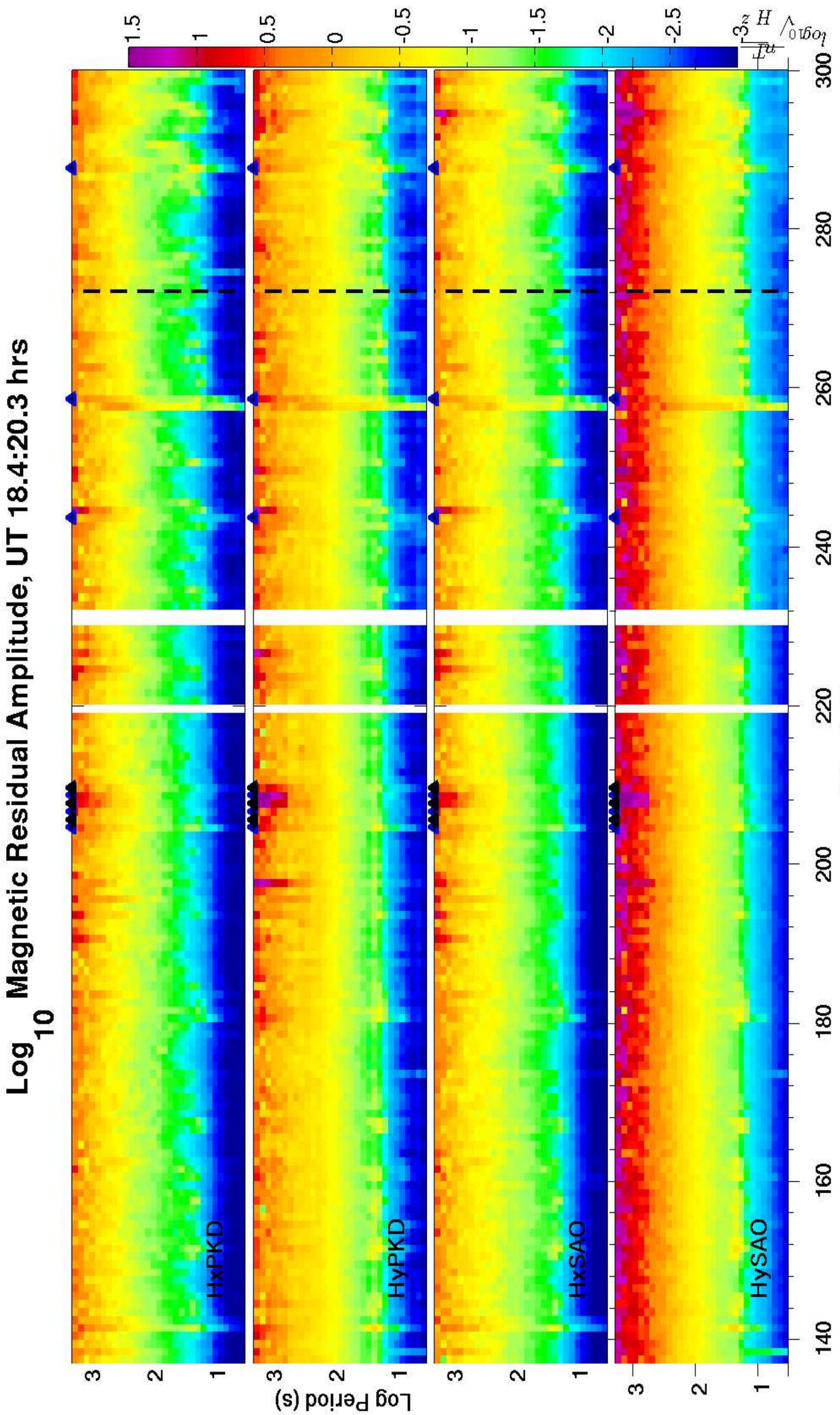


Figure 6.24: Residual amplitudes in $\log_{10}(\text{nT}/\sqrt{Hz})$.

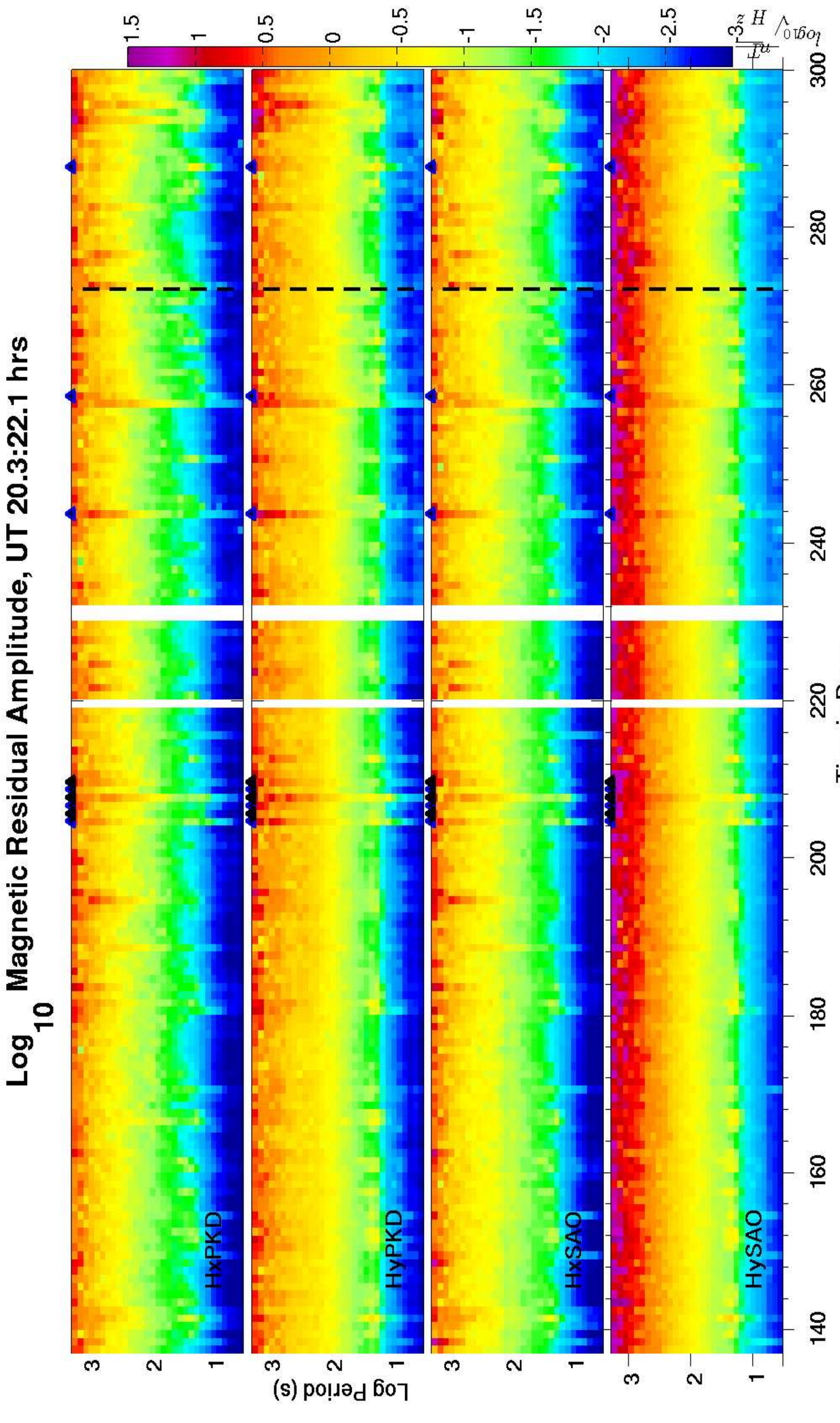


Figure 6.25: Residual amplitudes in $\log_{10}(\mathrm{nT}/\sqrt{Hz})$.

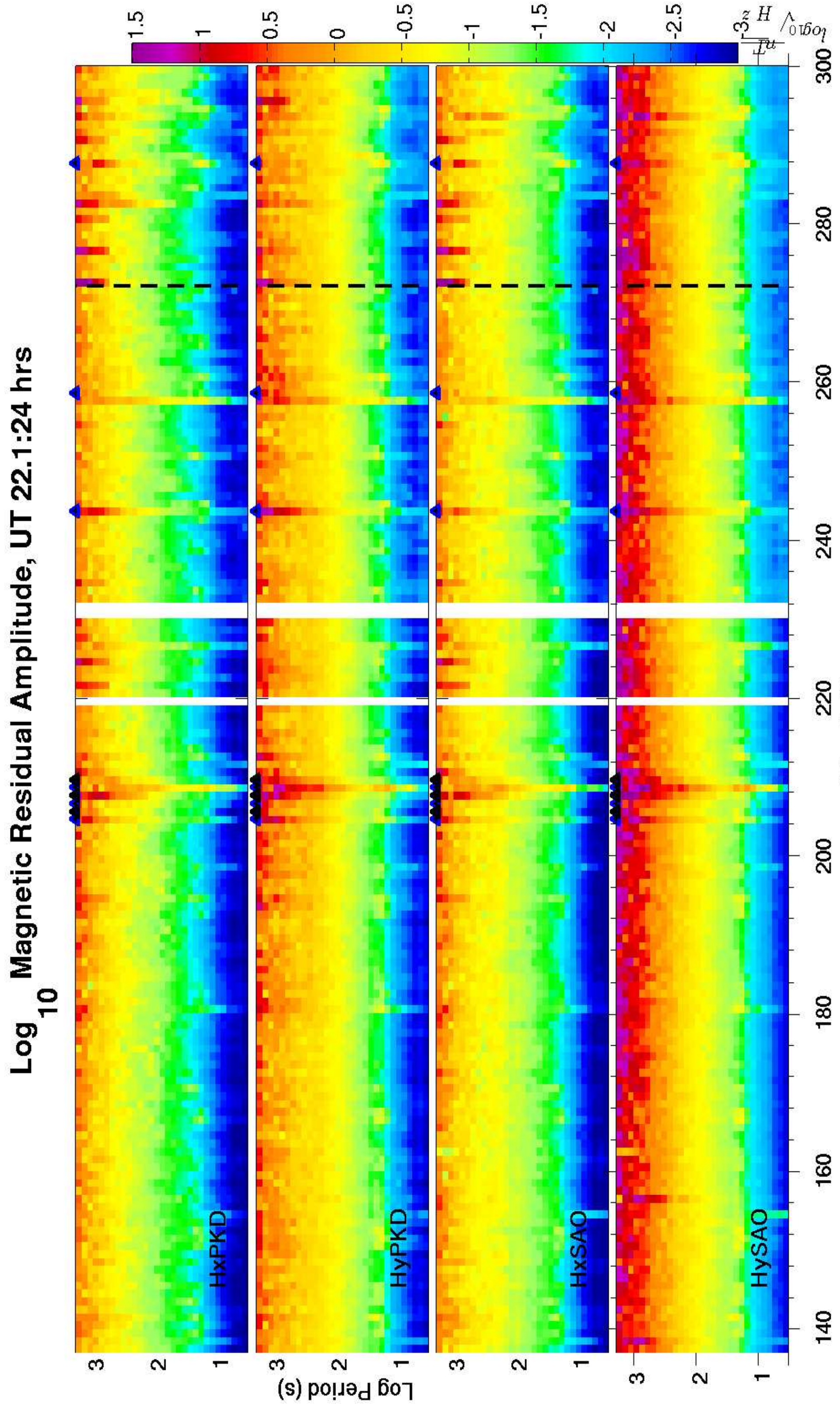


Figure 6.26: Residual amplitudes in $\log_{10}$(nT/$\sqrt{Hz}$).

Chapter 7

Eigenvalues of the SDM

The robustly-estimated spectral density matrices (SDMs) expressed in units of signal to noise ratio and discussed in Chapter 4 can be used to estimate the number of uncorrelated sources present in the array data. Typically, the MT field is a two dimensional entity at mid-latitudes [*Egbert and Booker*, 1989], and it is this field which is the dominant source of the ambient fields when cultural sources are far away. If a subsurface process were occurring in association with the earthquake, it is difficult to imagine that such a process would be correlated with the ambient geomagnetic variations. This assumption motivates the decomposition of the SDMs in each frequency band into their principal components, and an examination of the time series of eigenvalues associated with each principal component.

Formally, for an arbitrary but fixed combination of day t and frequency band ω, the daily averaged SDM $\mathbf{S}_\omega$(t) is expressed as

$$\mathbf{S}_\omega(t) = \mathbf{U}_\omega(t)\mathbf{D}_\omega(t)\mathbf{U}^*_\omega(t) \tag{7.1}$$

where **U** is the unitary matrix whose columns are the normalized eigenvectors of **S**, and **D** is a diagonal matrix whose non-zero elements are the eigenvalues of **S** (*see Appendix 1 for a discussion of the eigenvectors of covariance matrices*). Assuming that **D** is ordered so that the diagonal entries λ_i^2 are in descending order, the dominant mode of the data distribution which generated **S** is given by $\mathbf{u}_{:,1}$ (i.e. the first column of **U**). The coordinate system in which **u** is described is $\mathbb{C}^8$, such that each axis is associated with one of the eight sensors, and where one graduation on the i^{th} axis equals the standard deviation of the noise of the i^{th} channel. The dominant mode ^{1}M of the daily time series **X**(t) is then described by the time series:

$$^1M(t) = \sum_{i=1}^{8} u_{i,1}\mathbf{X}_{i,:}(t) \tag{7.2}$$

where $u_{1,j}$ is the j^{th} element of the 1st column of **U**, and $\mathbf{X}_{i,:}$(t) denotes the ith instrument channel making up **X**.

The first four ordered eigenvalues of **S**, D$_{i,i}$ are plotted in Figure 7.1. This plot shows that there are two dominant sources of energy, but that a non-trivial third source is clearly present especially in the bands around 30 to 100s. This is mostly noise form the BART (Bay Area Rapid Transit) trains, as shown in *Egbert et al.* (2000). The 163 day window around the earthquake is shown in Figure 7.2.

There is a clear increase in all eigenvalue amplitudes in the dead band on day 272. This appears to be the signal produced by the earthquake itself, and is discussed later in the section on future work. Evidence that this anomaly is the actual 'co-seismic signal' [*Kappler et al.*, 2006] is shown in Figures 7.3 and 7.4. Figures 7.3 show significant signals observed co-seismically in both the electric and magnetic recordings. Figure 7.4 displays the linear combination of channels (scaled into units of SNR) which make up the third eigenmode at sample Fourier coefficients in a few frequency bands. Figure 7.4(a,b) clearly show the co-seismic signal at 1715h UT, which is an order of magnitude larger than the other signal in (a). This one spike effectively more than doubles the norm of the time series. In case (b) at $\sim$ 7s the earthquake is still visible, but does not change the norm of the time series significantly. By the time the period gets to 13s, the co-seismic signal is buried beneath the ambient noise. The co-seismic signal is not removed by the despiker because the seismic waves only take around 45 seconds to propagate between the sites and hence the 256 second variance windows (described in Chapter 3) see a spike at both sites, avoiding spurious flagging of this window.

We see that there is an energy increase energy in the third eigenmode of the 4s period around the time of the earthquake. This band has shown anomalous increases in energy in the residuals as well, most visible in Ey at Parkfield. Three noteworthy aspects of this plot are: 1. Most of the energy in the system is in the top two eigenvalues. This implies that the condition number of the SDM is quite high. On days when there are solar storms more energy can be seen in the lower order modes; 2. The third eigenvalue shows a significant source which is uncorrelated with the two dominant modes in the band around 30-100s; 3. No anomalous energy (besides the narrow band phenomena in the third eigenvalue around 4s) is present around the time of the earthquake in this plot.

7.1 Dominant modes

The colour axis of the eigenvector plots shown in Figures 7.1 and 7.2 represents the lengths of the principal axes of the data distribution in the eight dimensional complex space $\mathbb{C}^8$. The principal axes of the data distribution have not, thus far, been shown to be stable. It is unclear whether the shape of the data distribution varies from day to day in its eccentricity (i.e the ratios of the various eigenvalues to one another). Also, the orientation of the distribution could vary, i.e. the distribution of data on one day may be rotated with respect to the distribution on another day. Considering that the dominant sources of the field (below 1Hz) are the quasi-stationary pulsations of the magnetosphere induced by the solar wind flowing past, it seems plausible that the modes of the daily distributions are similar from day to day.

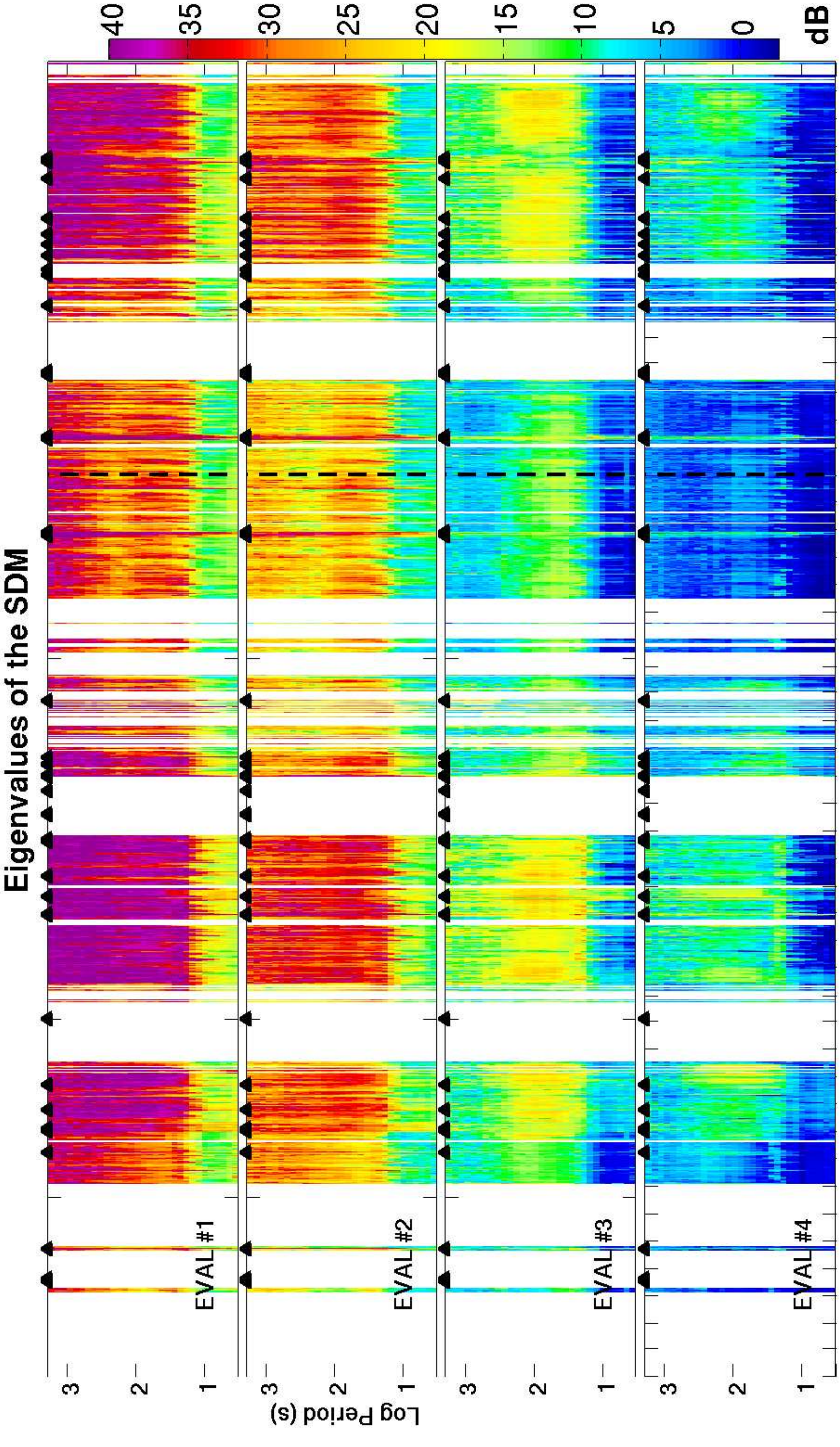


Figure 7.1: Dominant four eigenvalues of the SDM in dB plotted for the 2002-2005 time interval for all frequencies.

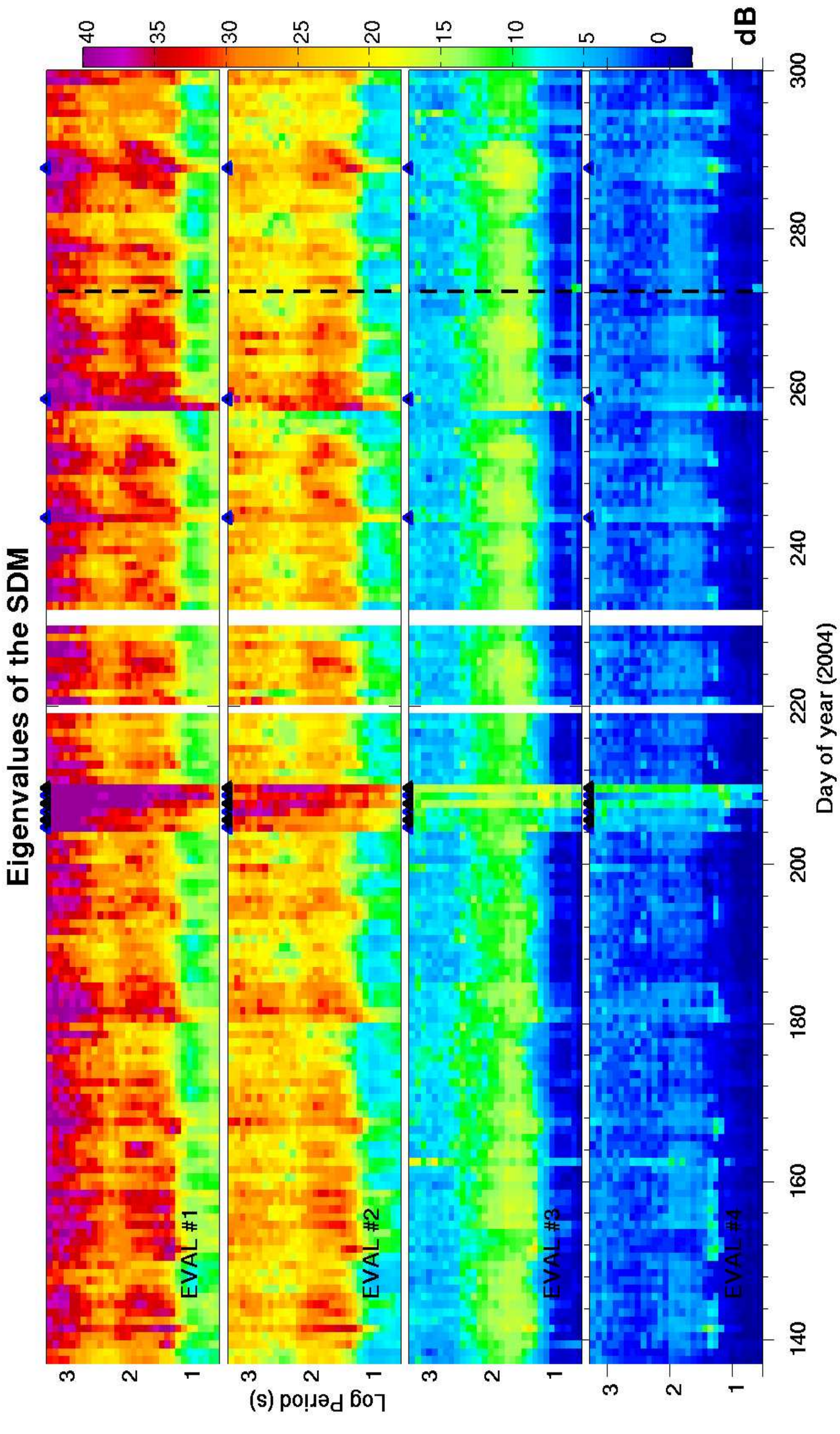


Figure 7.2: Dominant four eigenvalues of the SDM in dB plotted for the 163-day interval surrounding the 2004 Parkfield earthquake.

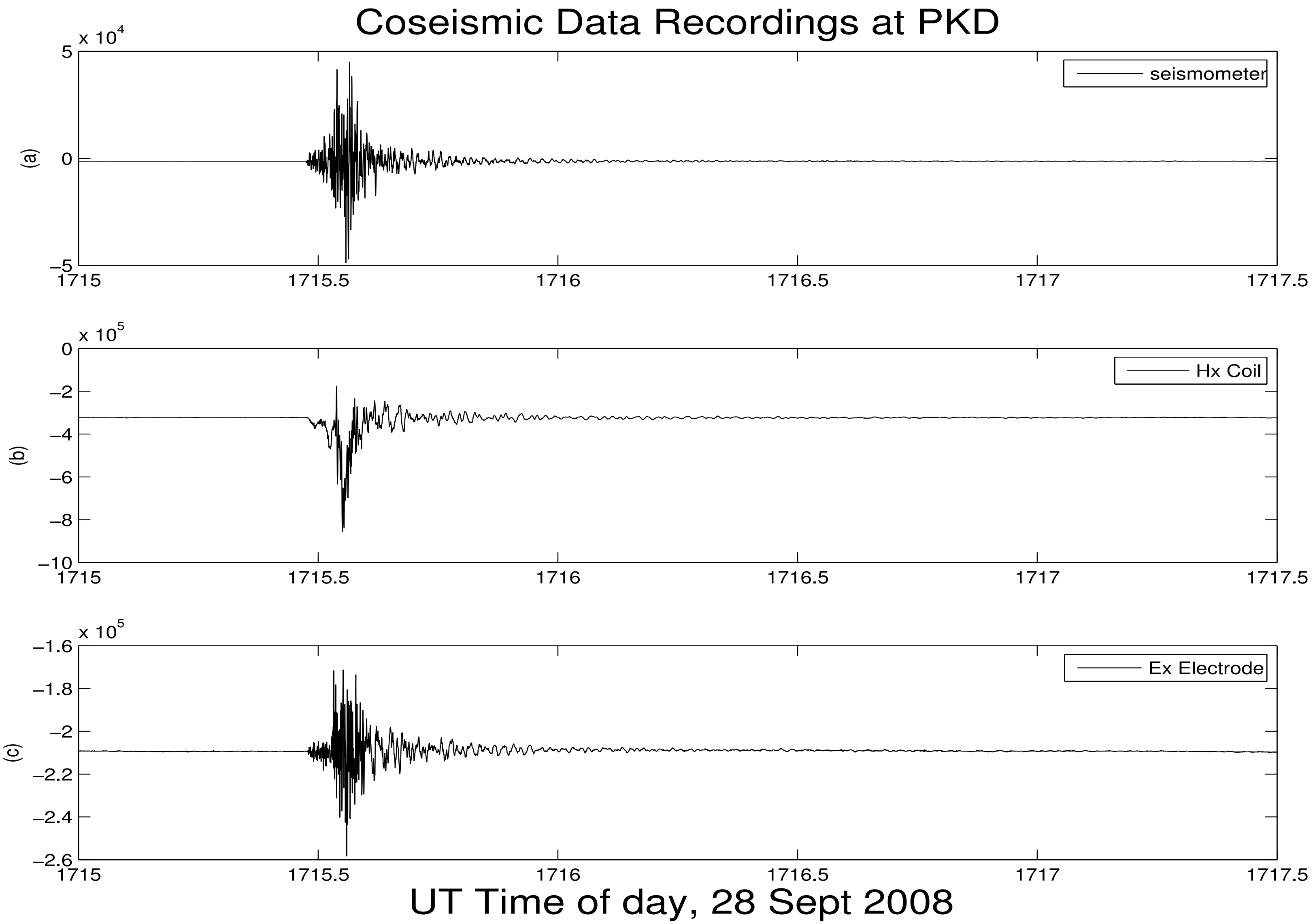


Figure 7.3: Co-seismic signals registered by the PKD seismometer (a), coils (b), and electrode (c). All y-axis units are in raw datalogger counts.

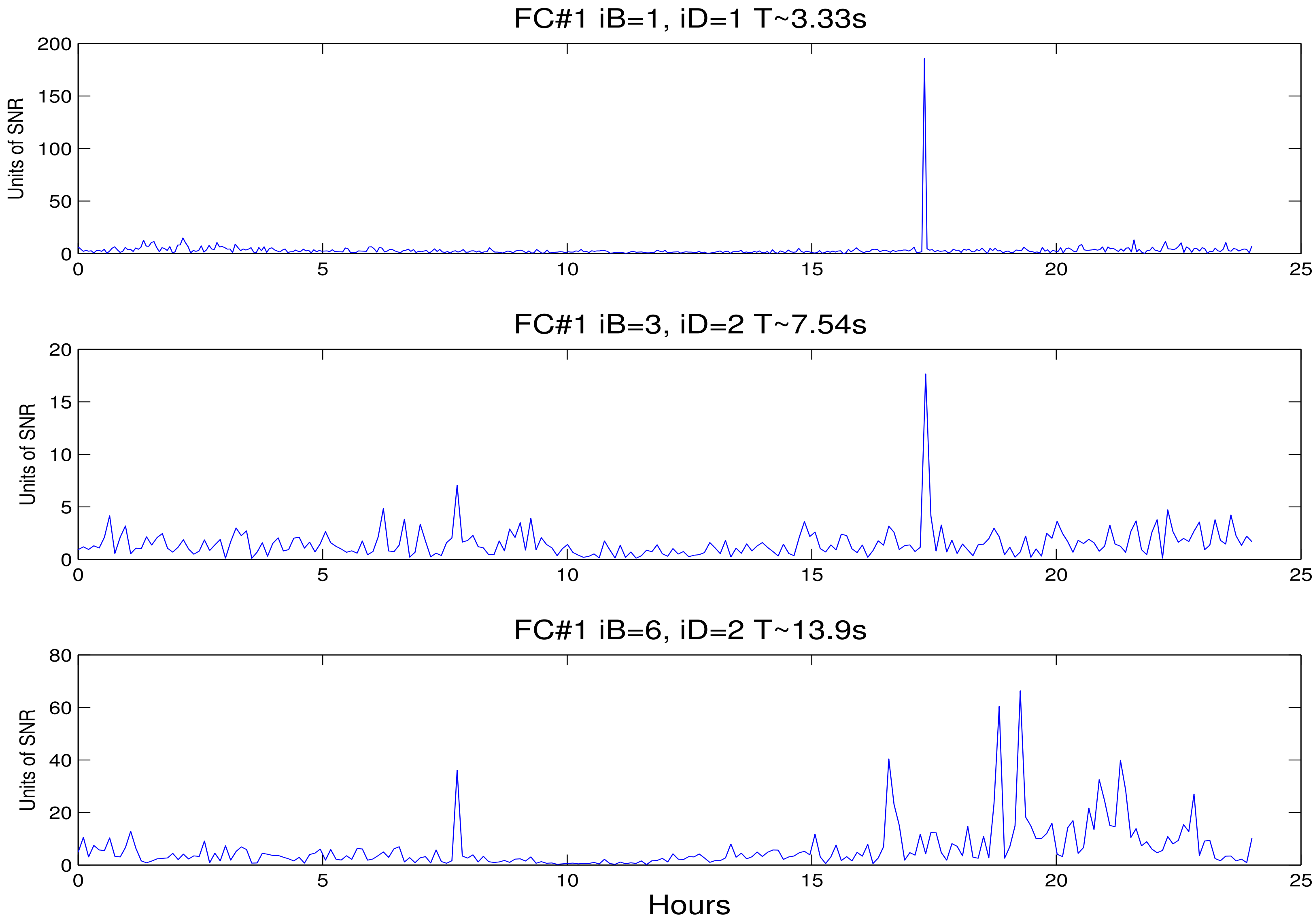


Figure 7.4: Fourier Coefficients for Hx at PKD, September 28, 2004. The spike at 1715UT corresponds the the earthquake.

To measure the tendency of the principal axes of the data distribution to vary or 'wiggle' from day to day, we look at the variation in angle of the Nth ordered, daily average eigenvectors, at frequency ω with respect to some reference unit vector $\hat{\mathbf{r}}_{N,\omega}$.

Formally, for each day t and frequency band ω, define $\mathbf{U}_{\omega}$(t) as the unitary matrix whose columns are the ordered, normalized eigenvectors of $\mathbf{S}_{\omega}$(t). For each of the eight modes, define $\mathbf{u}_{N,\omega}$(t) (N=1..8) as the Nth unit eigenvector of **S** or column of **U**. By inspecting Figure 7.2 together with some of the noise level plots in the previous sections, we can select out a few days that are low in noise and show no remarkable storm activity. For example days 154 through 161 of 2004 have an average Ap of 9.6, and show fairly low noise levels, especially at PKD. The reference vector, $\hat{\mathbf{r}}_{N,\omega}$ is then made by averaging together the $\mathbf{u}_{N,\omega}$(t) for t in [154, 161].

There are several ways to define the mean of a complex number. Here we store the complex numbers in Cartesian form, and simply average over the real parts and the imaginary parts independently. Therefore, the expectation value E of a complex variable **x** is estimated by:

$$E(\mathbf{x}) = E(Re(\mathbf{x})) + i * E(Im(\mathbf{x})). \tag{7.3}$$

Unless otherwise stated, any averaging over complex numbers is done via equation 7.5. Thus:

$$\hat{\mathbf{r}}_\omega = \frac{E(\mathbf{x}_\omega)}{\|E(\mathbf{x}_\omega)\|} \tag{7.4}$$

The cosine of the angle between the Nth average mode $\mathbf{r}_{N,\omega}$, and the Nth daily observed mode $\mathbf{u}_{N,\omega}$(t) is then simply given by

$$\| < \mathbf{r}_{N,\omega}, \mathbf{u}_{N,\omega}(t) > \| \tag{7.5}$$

Here $<,>$ denotes standard inner product. The cosines associated with each observed daily mode and reference vector are plotted as a function of frequency and time in Figure 7.5. The colour scale saturates at $\sqrt{2}$ corresponding to a cutoff at 45 degrees. Note that the cosines are typically near 1 in the first and second modes, indicating stability in eigenvector orientation. Interestingly, the modes which are consistently co-linear occur in the third eigenvector in the 30-100s range. This is undoubtedly the effect of the BART train. It is interesting to note that in the smallest eigenvalues (7 and 8) in the dead band there is also a consistent mode whose orientation is highly stable. This may be the instrument noise itself.

Motivated by evidence that the dominant eigenvectors are stable, we stack out some of the jitter in the modes by averaging the daily SDMs. If we average over all of the days in the 163 day window, we obtain Figure 7.6. Notice that most of the energy which was present in the fourth mode in Figure 7.2 is no longer present in this plot, implying that the orientation

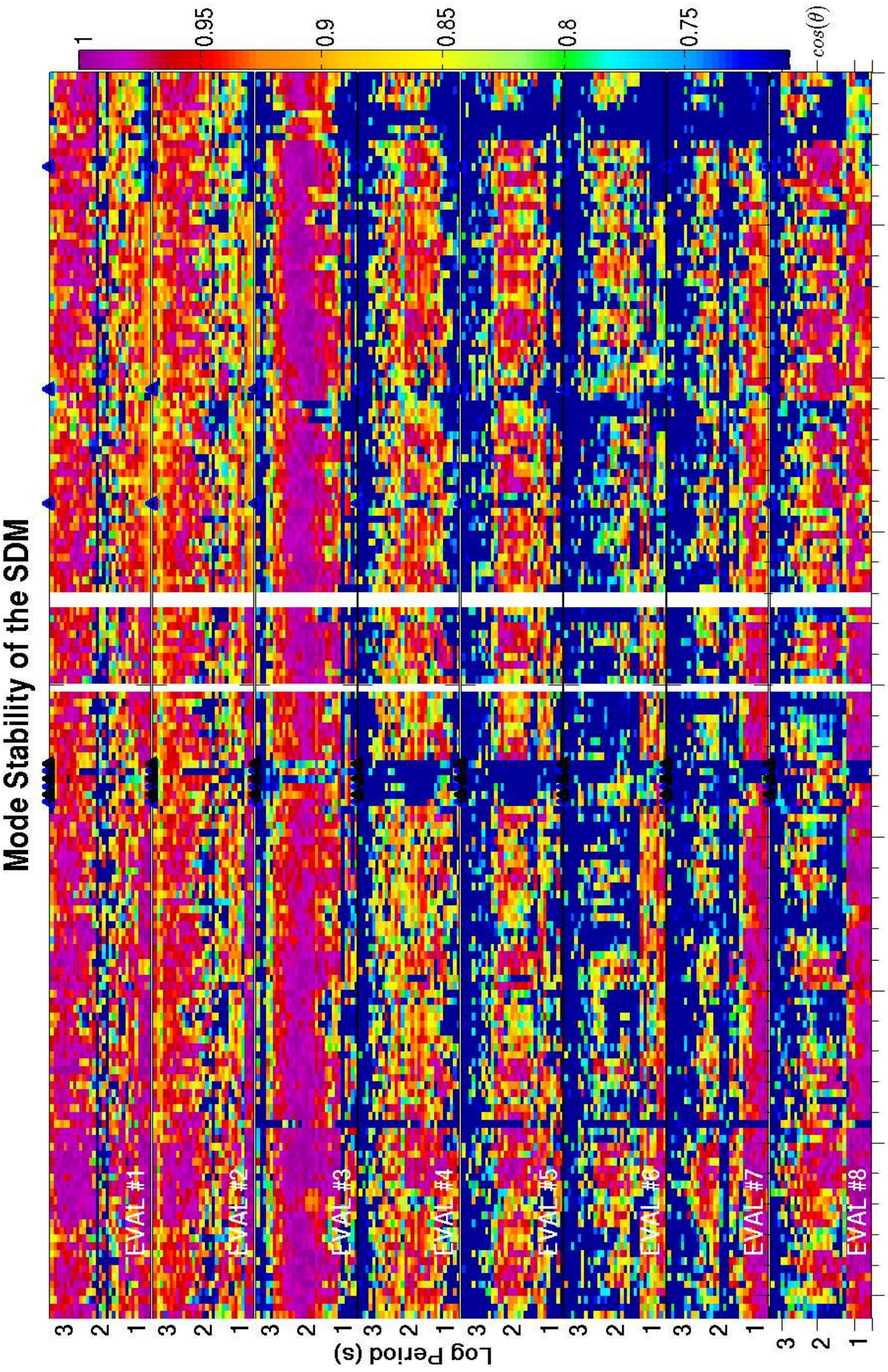


Figure 7.5: Cosines of the angles between observed modes and reference vectors.

of mode four is not particularly stable.

There is some evidence for weekly periodicity in the bands around 30s period in the third eigenvalue of Figure 7.6. This is explored further in Figures 7.7 through 7.9. A persistent phenomena in the 4th average mode around 10s period, first becomes apparent in these plots.

Note that in Figure 7.2 much energy is injected into modes 3 and 4 when storms are occurring. This energy clearly skews the 3rd and 4th eigenvectors, as can be seen in Figure 7.5. To account for this effect, an averaged SDM is calculated only over days when neither minor nor major storm activity is present (Ap<29). Projections onto the modes of this SDM are shown in Figure 7.7, and reveal some weekly periodicity in the 'BART' band. The times of reduced amplitude in the 3^{rd} mode correspond to Sundays when the BART train runs a reduced schedule. Respecting the caveat that some subtle components in the data distribution may be present around the time of the earthquake, Figure 7.7 is reproduced in Figure 7.8, but in this case only non-storm days which occurred more than four weeks prior to the earthquake are included in the averaged SDM. Noon on Sunday is denoted by a thin, black, solid vertical line.

Averaged SDMs can be calculated over any set of days. It is not necessary that the days which generate an average SDM $\bar{S}$ are in the set of daily SDMs which are summarily projected. Figure 7.9, is made by averaging only weekdays, in particular days [806-811] and [816-820], although the projected (plotted) days correspond to days [867-1030]. Note

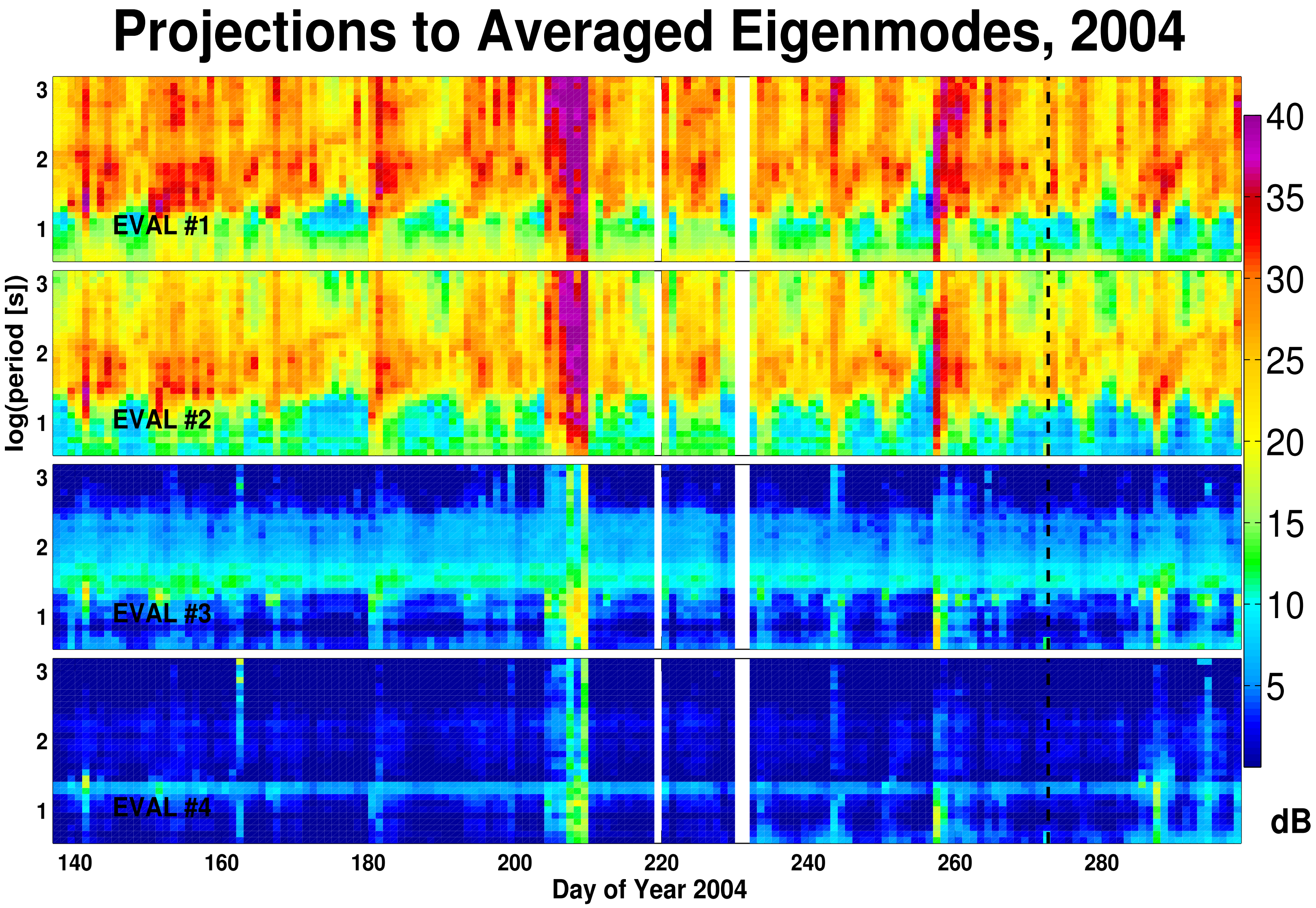


Figure 7.6: Projection of daily eigenvectors onto modes of averaged SDM calculated over all days in the 163 day window shown

that in all of the 'averaged modes' plots, the narrow band signal, previously observed in the third eigenvalue at $\sim$ 4seconds in Figure 7.2 is not visible. This implies that the signal was not typically present on the days used to create the averaged modes, i.e. that that component is not typically part of the data distribution. Figures 7.6 through 7.9, also demonstrate some sensitivity to the choice of days used in the average, but as long as days in which the distribution is highly abnormal (such as major storm days) are not used in calculating the averaged SDM, the plots retain the same basic characteristics.

The plots presented in this chapter show several subtle phenomena in the ULF EM fields, but provide no information about the location of a particular phenomenon. In the following section, we explore a technique that can be used to increase our confidence that an observed anomaly is physically plausible by determining if it is coherent between E and H, and to determine if the phenomenon is present at a single site, or both sites, by determining if it is coherent at sensors present at both sites.

Figure 7.10 is a plot of the eigenvalues of the SDMs calculated from the residual time series from Chapter 5. The residual SDM is calculated by simply substituting the residual time series in for the original time series in Equation 4.8. A comparison of Figure 7.10 with Figure 7.2 clearly shows that the two dominant modes have been almost completely removed by calculating the residuals. Indeed, the main source of energy in Figure 7.10 is the BART.

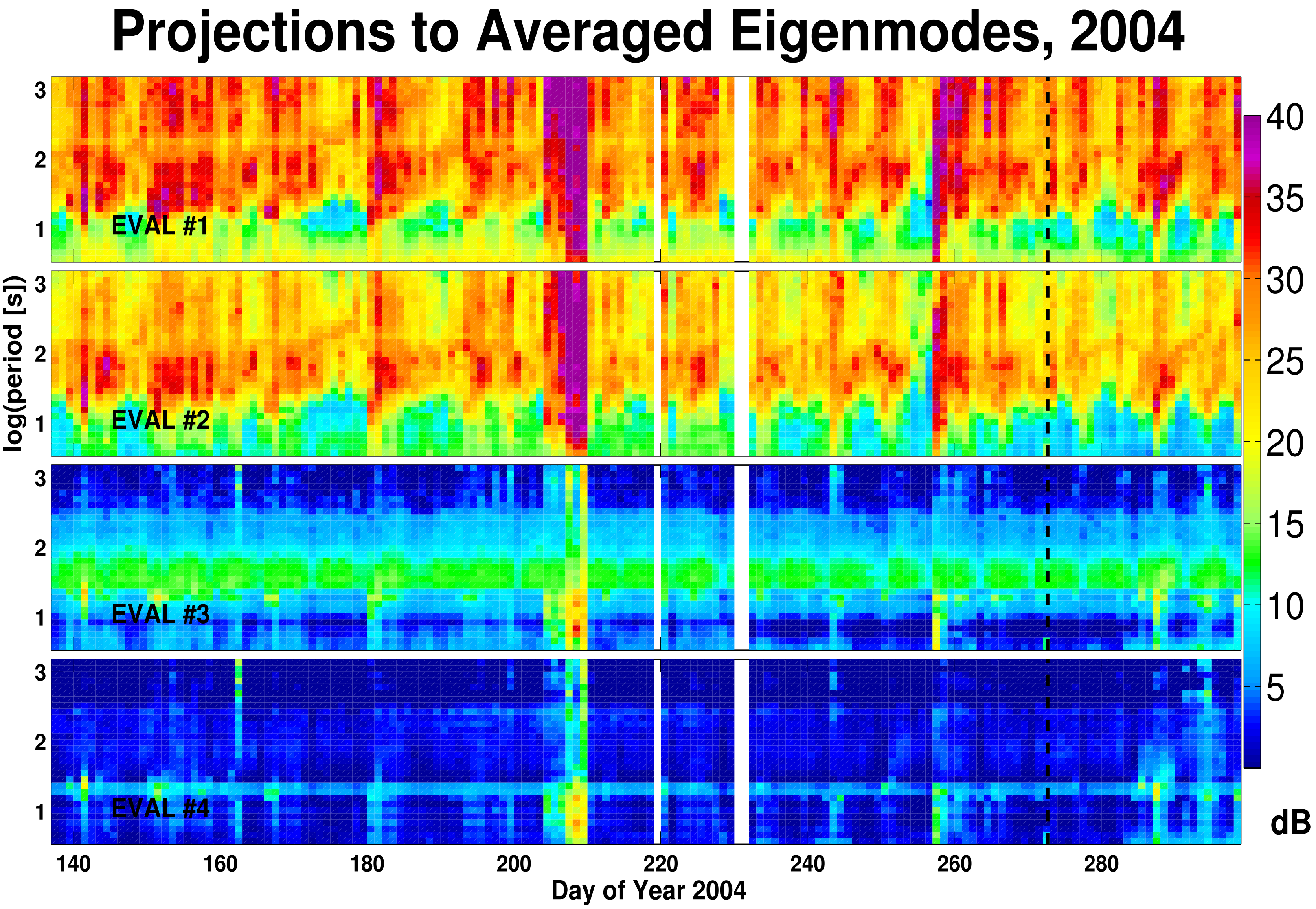


Figure 7.7: Projection of daily eigenvectors onto averaged modes in the case where SDM is averaged over all non-storm days in the 163 day window.

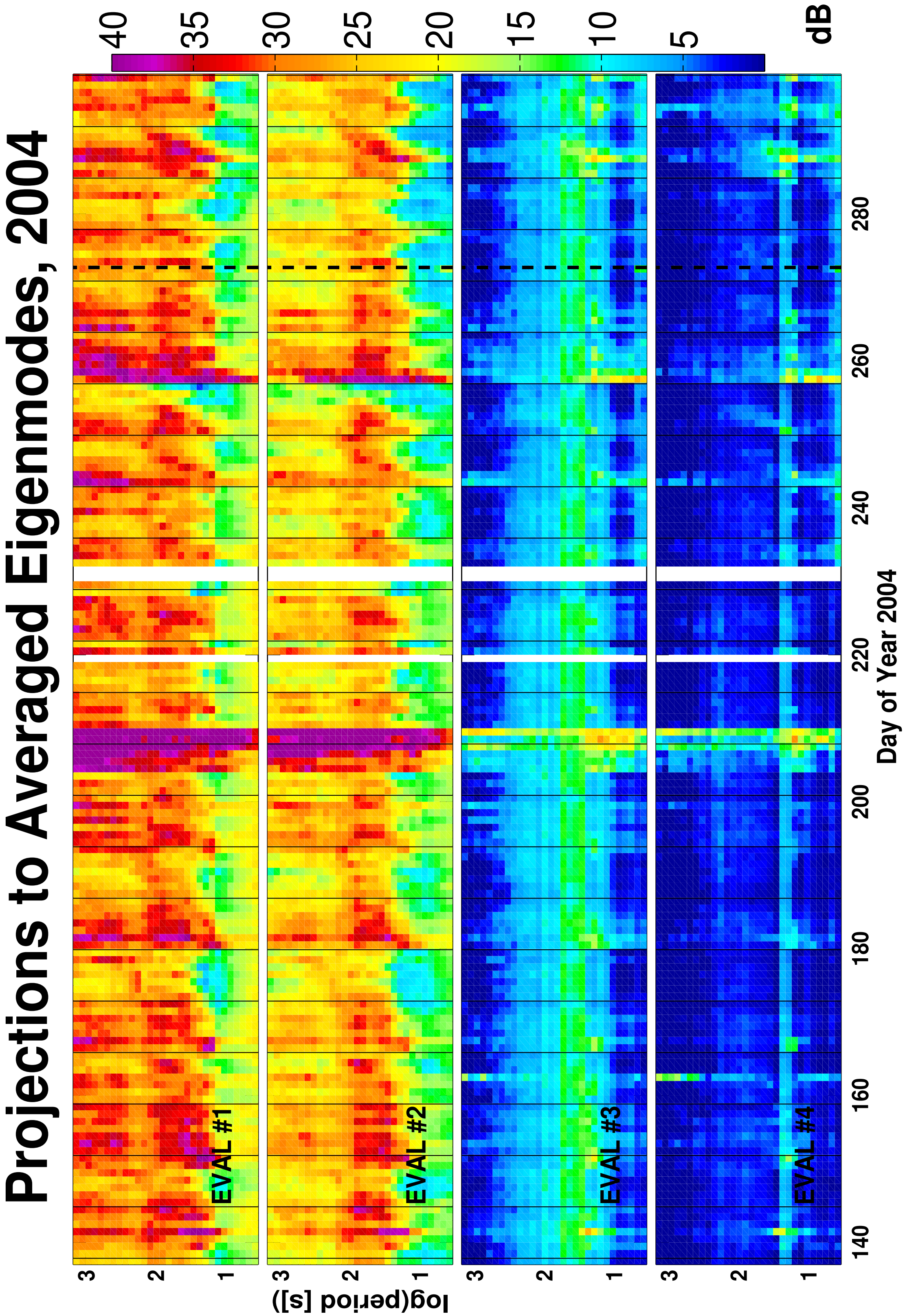


Figure 7.8: Projection of daily eigenvectors onto averaged modes. SDM averaged over all non-storm days in window which are more than four weeks prior to September 28th.

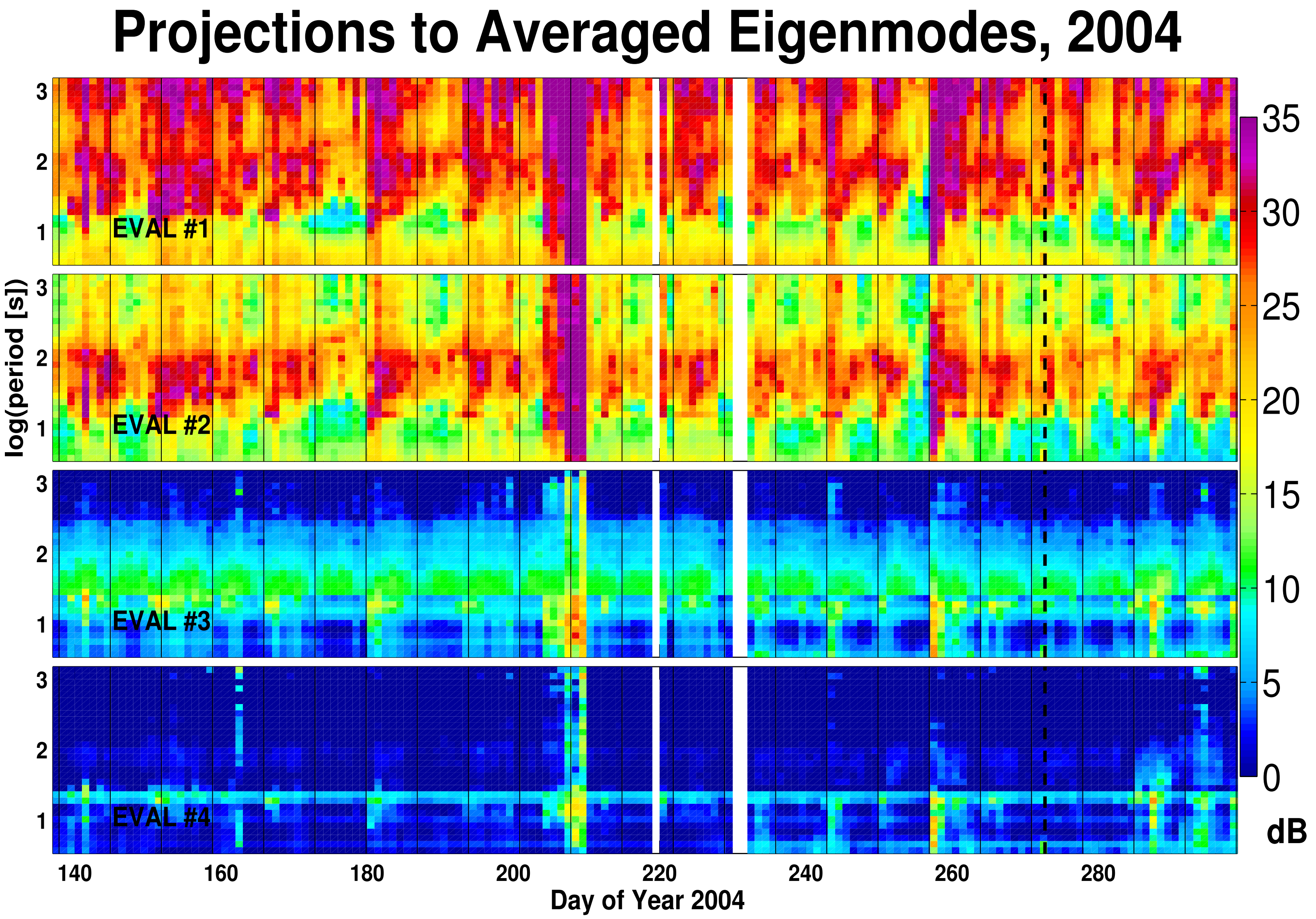


Figure 7.9: Projection of daily eigenvectors onto averaged modes in the case where SDM is averaged over weekdays [806-811], and [816-820].

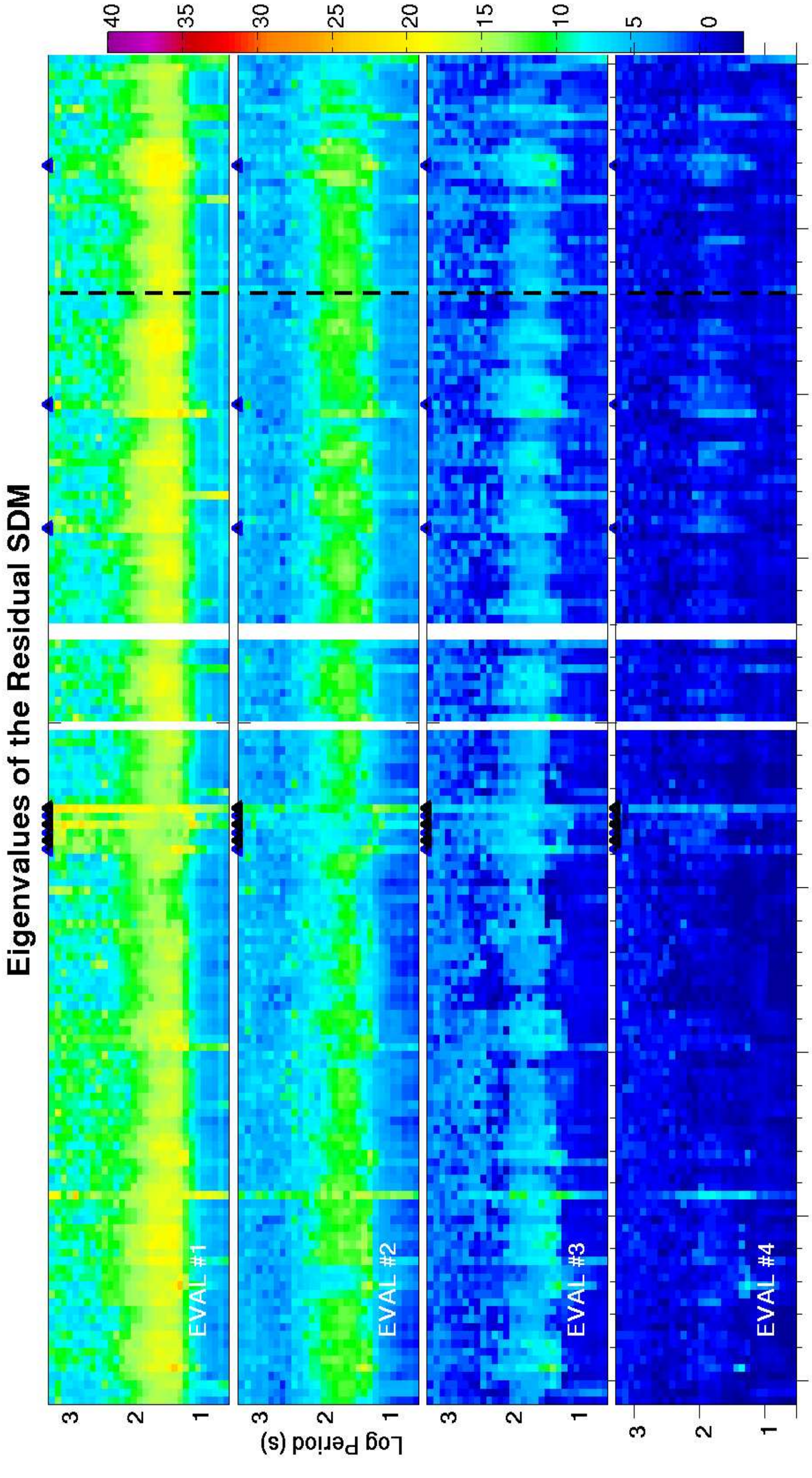


Figure 7.10: Eigenvalues of the residual SDM. Colour bar is in dB with respect to noise level

Chapter 8

Canonical Coherences

8.1 The method of Canonical Coherences

The Canonical Coherence (CC), sometimes referred to as Canonical Correlations, method [*Hotelling*, 1935; *Brillinger*, 1969] has been applied to long term geophysical monitoring by *Lyubushin* (1998). Here, it is a frequency dependent measure of the coherence between two vector-valued time series. We use the method to explore the extent to which two disjoint groupings of array channels are correlated. A time series of Fourier coefficients for all array channels $\mathbf{Z}$(t), is broken into two time series, $\mathbf{X}$(t) and $\mathbf{Y}$(t) where there is no intersection between the set of channels making up $\mathbf{X}$ and the set of channels making up $\mathbf{Y}$. We seek to best approximate the behaviour of a linear combination of channels $\mathbf{X}$(t) using a linear combination of channels in the other group **Y**(t). Let dim(**X**)=m and dim(Y)=n, and without loss of generality, $m \leq n$. Then, there exist m ordered pairs of row vectors $\{\mathbf{p}_1,\mathbf{s}_1\}...\{\mathbf{p}_m,\mathbf{s}_m\}$, where $\mathbf{p}_i$ and $\mathbf{s}_i$ are m and n dimensional respectively with the following properties:

P1: the correlation between the canonical variates $\mathbf{p}_i^*\mathbf{X}$ and $\mathbf{s}_i^*\mathbf{Y}$ is maximal in the space defined by the projections of **X**(t) and **Y**(t) orthogonal to the first *i-1* canonical variates. (for i=1 no projection is required).

P2: the canonical variates within **X** or **Y** are uncorrelated, i.e. $<\mathbf{p}_i^*\mathbf{X},\mathbf{p}_j^*\mathbf{X}>=0=<\mathbf{s}_i^*\mathbf{Y},\mathbf{s}_j^*\mathbf{Y}>$ for $i \neq j$

The calculation of the first first canonical variates is shown below: If **X** and **Y** are mean subtracted, then the correlation Γ of the vectors $\mathbf{p}_1^*\mathbf{X}$ and $\mathbf{s}_1^*\mathbf{Y}$ is given by:

$$\Gamma = \frac{<\mathbf{pX},\mathbf{sY}>}{\sqrt{<\mathbf{pX},\mathbf{pX}><\mathbf{sY},\mathbf{sY}>}} = \frac{\mathbf{pXY}^*\mathbf{s}^*}{\sqrt{(\mathbf{pXX}^*\mathbf{p}^*)(\mathbf{sYY}^*\mathbf{s}^*)}} \tag{8.1}$$

Adopting standard covariance matrix notation (Σ_{XY}=$\mathbf{XY}^*$), equation 8.1 can be written:

$$\Gamma = \frac{\mathbf{p}\Sigma_{XY}\mathbf{s}^*}{\sqrt{\mathbf{p}\Sigma_{XX}\mathbf{p}^*\mathbf{s}\Sigma_{YY}\mathbf{s}^*}} \tag{8.2}$$

Now, define two vector-valued variables:

$$\rho = (\Sigma_{XX}^{1/2})\mathbf{p}^* \qquad \sigma = (\Sigma_{YY}{}^{1/2})\mathbf{s}^* \tag{8.3}$$

The square-root matrices above are well defined as long as the covariance matrices are non-singular. Note, if the condition number of either covariance matrix is too large to reliably calculate the square roots of the covariance matrices, one will need to remove extraneous dimensions using principal components or similar techniques.

Substituting the variables of Equations 8.3 into Equation 8.2 yields:

$$\Gamma = \frac{\rho \Sigma_{XX}^{-1/2} \Sigma_{XY} \Sigma_{YY}^{-1/2} \sigma^*}{\sqrt{(\rho\rho^*)(\sigma\sigma^*)}} \tag{8.4}$$

Now considering the numerator of Equation 8.4 as the inner product between the two vectors **a** and **b**, defined as

$$\mathbf{a} = \rho \Sigma_{XX}^{-1/2} \Sigma_{XY} \Sigma_{YY}^{-1/2} \qquad \mathbf{b} = \sigma^* \tag{8.5}$$

and substituting Equation 8.5 into the following form of the Cauchy-Schwarz inequality:

$$< \mathbf{a}, \mathbf{b} > \leq \sqrt{< \mathbf{a}, \mathbf{a} >< \mathbf{b}, \mathbf{b} >} \tag{8.6}$$

we obtain the following inequality

$$\rho\Sigma_{XX}^{-1/2}\Sigma_{XY}\Sigma_{YY}^{-1/2}\sigma^* \leq (\rho\Sigma_{XX}^{-1/2}\Sigma_{XY}\Sigma_{YY}\Sigma_{YX}\Sigma_{XX}^{-1/2}\rho^*)^{1/2}(\sigma\sigma^*)^{1/2} \tag{8.7}$$

Dividing both sides of the above expression by the norms of ρ and σ, the left hand side becomes exactly the expression for Γ from 8.4 and we obtain:

$$\Gamma \leq \sqrt{\frac{(\rho\Sigma_{XX}^{-1/2}\Sigma_{XY}\Sigma_{YY}\Sigma_{YX}\Sigma_{XX}^{-1/2}\rho^*)}{\rho\rho^*}} \tag{8.8}$$

By viewing the five-matrix product in the center of the numerator in expression 8.8 as a single matrix:

$$\mathbf{A} = \Sigma_{XX}^{-1/2}\Sigma_{XY}\Sigma_{YY}\Sigma_{YX}\Sigma_{XX}^{-1/2} \tag{8.9}$$

we see that the expression within the square-root on the right hand side of 8.8 is actually a Rayleigh quotient, having form:

$$\frac{\rho\mathbf{A}\rho^*}{\rho\rho^*} \tag{8.10}$$

This ratio obtains its maximum value when ρ is the dominant eigenvector of **A** (note the expression is independent of any non-zero scalar applied to ρ). Thus, we have determined that Γ is bounded above by the value it attains for a specific ρ, which we can calculate from the data by performing an SVD on the matrix A. By definition 8.3, given ρ, we can calculate **p**. Calculation of **s** can be done similarly.

To obtain further canonical variates, an inductive argument can be used to subtract the projections of **X** and **Y** which lie along the first canonical variate, and then repeat the previous arguments, this time with *m* set to *m-1*. It turns out that the ordered eigenvectors of A provide the *m* vectors $\rho_1 ... \rho_m$ needed to calculate $p_1 ... p_m$, and the ordered eigenvalues $\lambda_1 ... \lambda_m$ provide the correlation coefficients of each canonical variate pair.

For now, we have shown that P1 is at least true for i=1;

If we take as the canonical variates the vectors $\mathbf{p}_i$ which are generated by A's eigenvectors ρ_i and equation 8.3 we can see that they obey property P2 as follows: The covariance of $\mathbf{p}_i\mathbf{X}$ and $\mathbf{p}_j\mathbf{X}$ is given by

$$\mathbf{p}_i \Sigma_{XX} \mathbf{p}_j^* = \rho_i \rho_j^* = \delta_{i,j} \tag{8.11}$$

where the first equality sign is verified by the definition of **p**, and the second follows from the properties of the unit eigenvectors of A. Clearly a similar argument can be applied to the $s_i\mathbf{Y}$. Defined as above, the 2nd pair of canonical variates has the greatest correlation of any vectors in the subspace orthogonal to the first vector. In fact the i^{th}th canonical variates are the maximally correlated linear combinations of $\mathbf{X}|_{i-1}$ and $\mathbf{Y}|_{i-1}$. Here $\mathbf{X}|_{k-1}$ refers to the projection of $\mathbf{X}$ onto the subspace orthogonal to $\{\mathbf{b}_{x,1}...\mathbf{b}_{x,k-1}\}$, and similar for $\mathbf{Y}$. We conclude that The eigenvalues of A yield the canonical correlation coefficients. Further exposition of the properties of canonical coherences is given in *Hardle and Simar* (2007).

8.2 Numerical results

With only two sites and two field types (E and H), there are two natural signal groupings. The channels are thus partitioned as:

1. Electrics in one group $\mathbf{X}$(t), magnetics in the other$\mathbf{Y}$(t);

2. Parkfield in one group, Hollister in the other.

Figures 8.1 and 8.2 show the four year behaviour of the four canonical coherence coefficients, for the E-H, PKD-SAO channel groupings respectively. The 163 day window surrounding the earthquake is shown for the same groupings in Figures 8.3 and 8.4

Note the increase in coherence of the 4s band in the third electric-magnetic canonical coherence around the time of the earthquake. A corresponding phenomena is not visible in the PKD-SAO plot. This implies that the phenomena is present at only one site. A closer

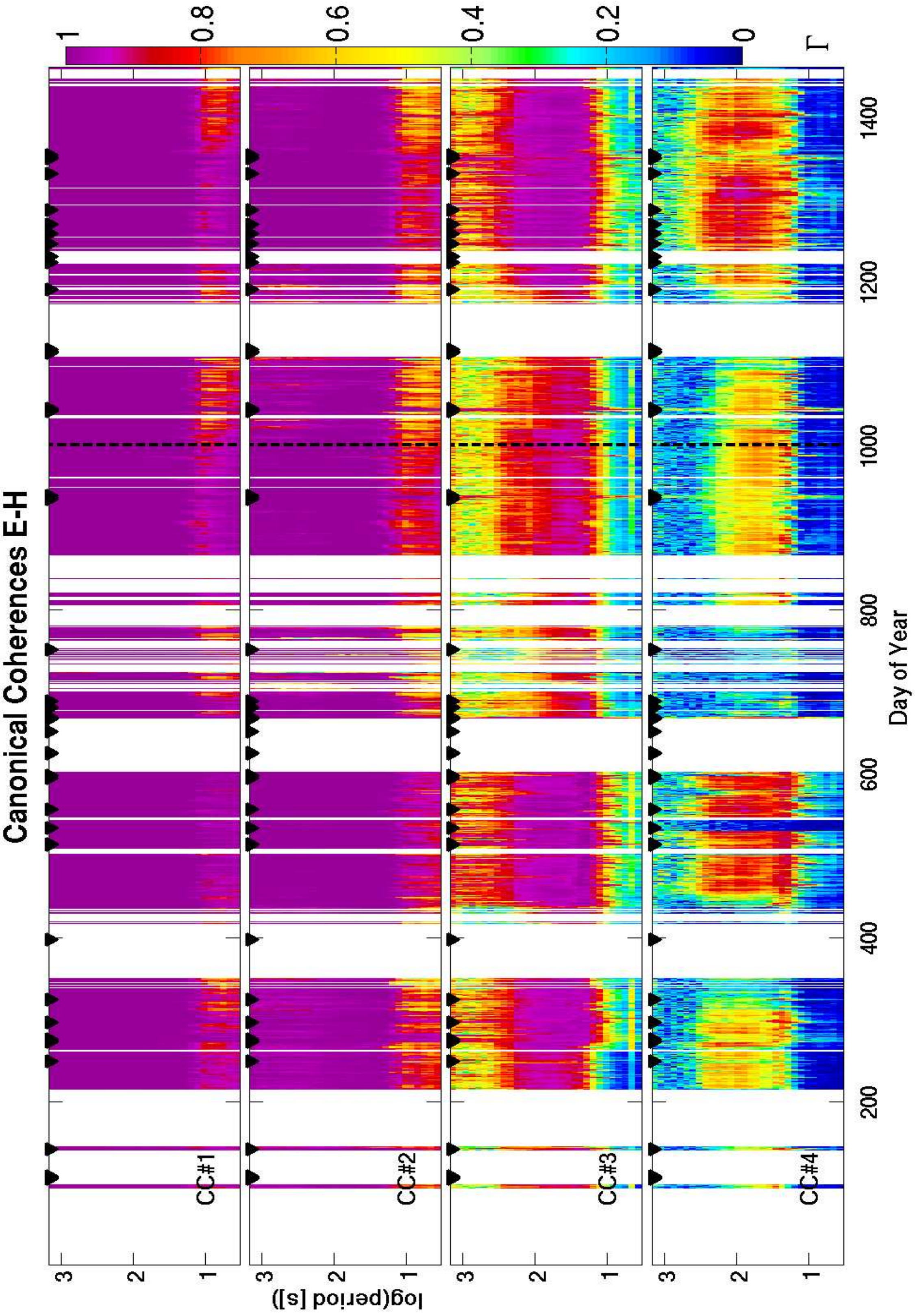


Figure 8.1: Canonical Coherences (E-H) for the 2002-2005 time interval. The colour axis is Γ. Black triangles indicate days of major magnetic storms.

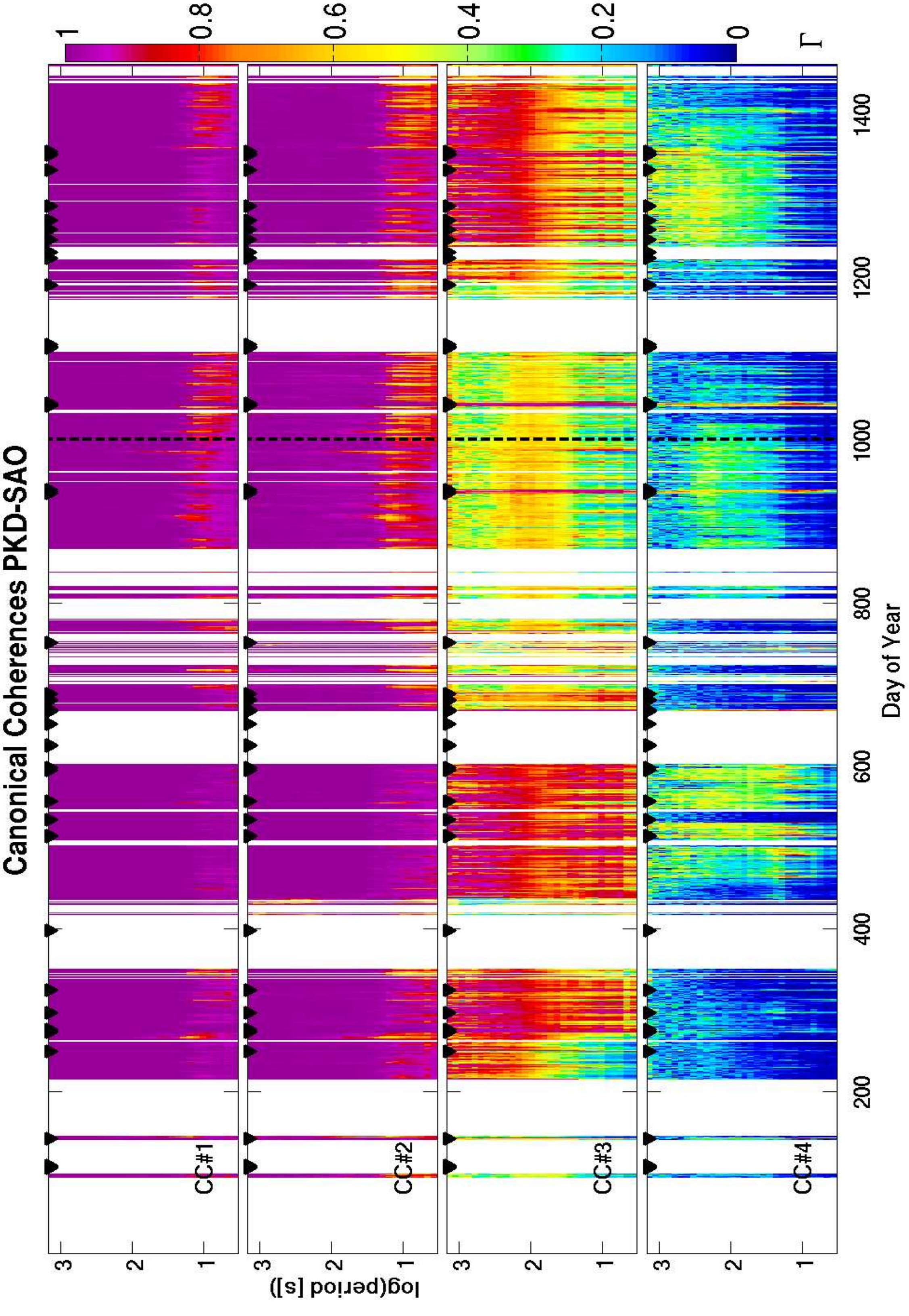


Figure 8.2: Canonical Coherences for the 2002-2005 time interval for intersite channel groupings. Colour axis is Γ. Black triangles indicate days of major magnetic storms.

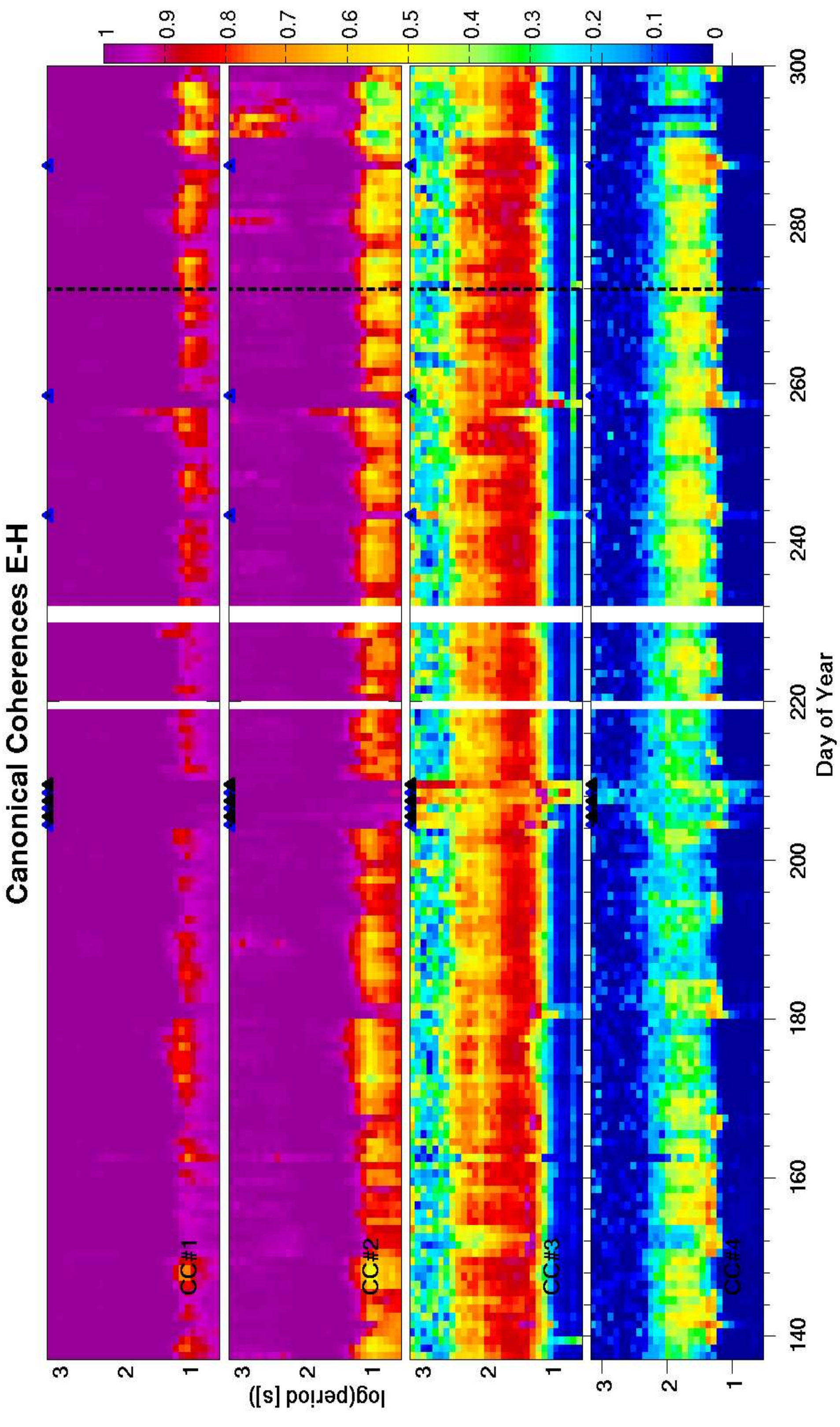


Figure 8.3: Canonical Coherences between electric and magnetic field channels in 2004.

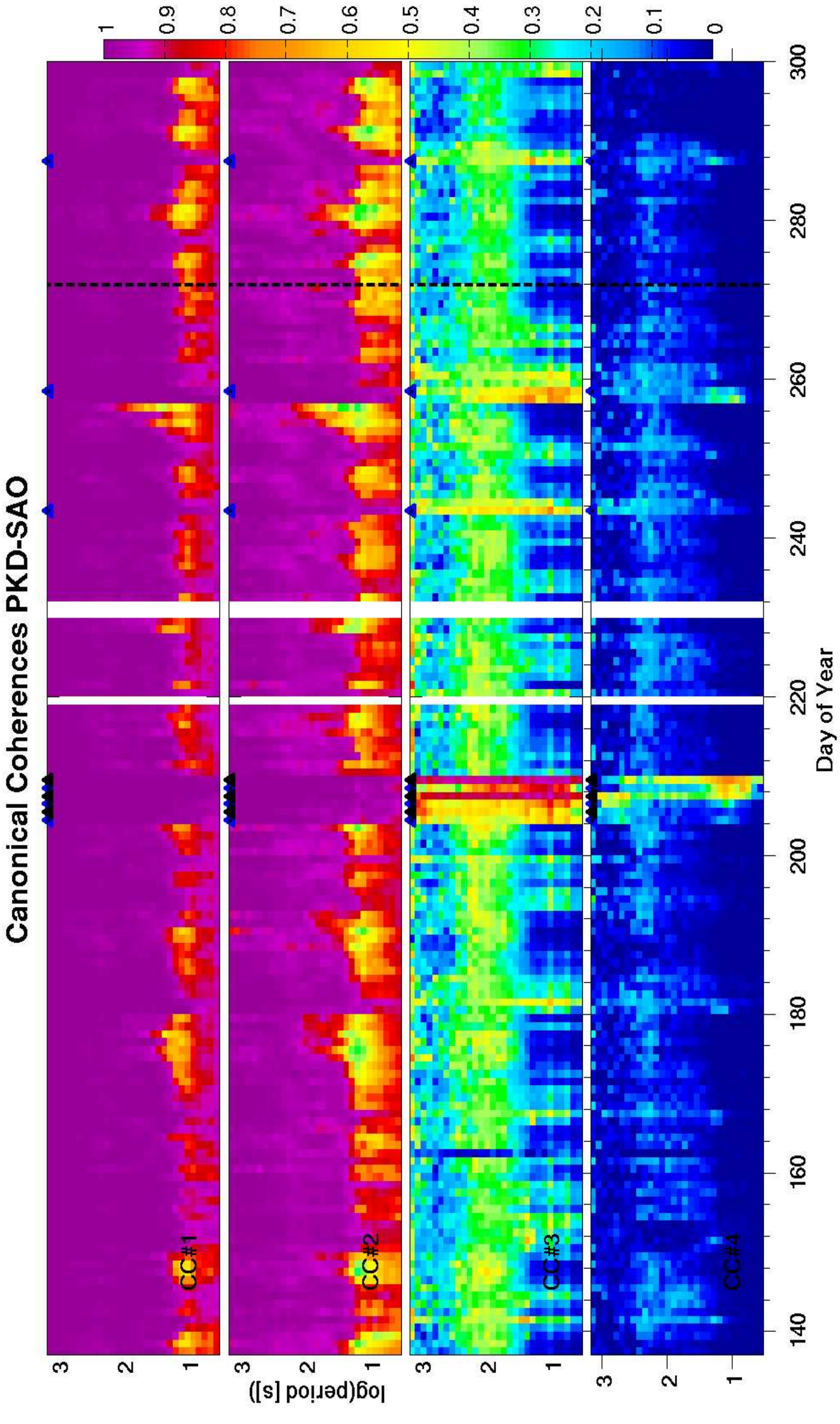


Figure 8.4: Canonical Coherences in 2004 between channel groupings: Parkfield, and Hollister.

inspection of this band is shown in Figure 8.5.

A four year plot of this 4s band, median filtered, is shown in Figure 8.6 (top) and again with storm days removed (bottom). The significance of the increased Γ value around the earthquake is diminished as one looks at this four-year time series. In an effort to reduce the appearance of variations in this time series which are also present in neighboring bands, we normalize the 3^{rd} CC coefficient in the the 4s band by the 3^{rd} CC coefficient in a neighboring band (3.3s). This results, after smoothing, in Figure 8.7. It is tempting to ascribe significance to the local maxima in the normalized coherence near the earthquake in Figure 8.7. The reality is that the large Γ observed in the 4s band of the 3^{rd} canonical coherence during Fall 2004 has a signature identical to the one observed in Spring of 2002. It is narrow-band and inhabits the exact same three Fourier coefficients of the band cantered around 4.09s. A plot of the Fourier coefficient and residual amplitudes, calculated according to the procedure outlined in Chapter 6 is shown in Figure 8.8, which compares April 5, 2002, against September 26, 2004. Note that although the observed field amplitudes are shifted from one time of observation to the next—as can be expected due to natural variability in field strength)— the anomalous signal is present at the same amplitude. No significant seismicity occurred at near PKD within a month of day 95, 2002. Clearly there is a real phenomenon responsible for these signals, and we discuss strategies for identifying the source in the section on future work. On the basis of Figure 8.8, however, it is reasonable to imagine that the source of the 2002 anomaly is the same as the source of the 2004 anomaly.

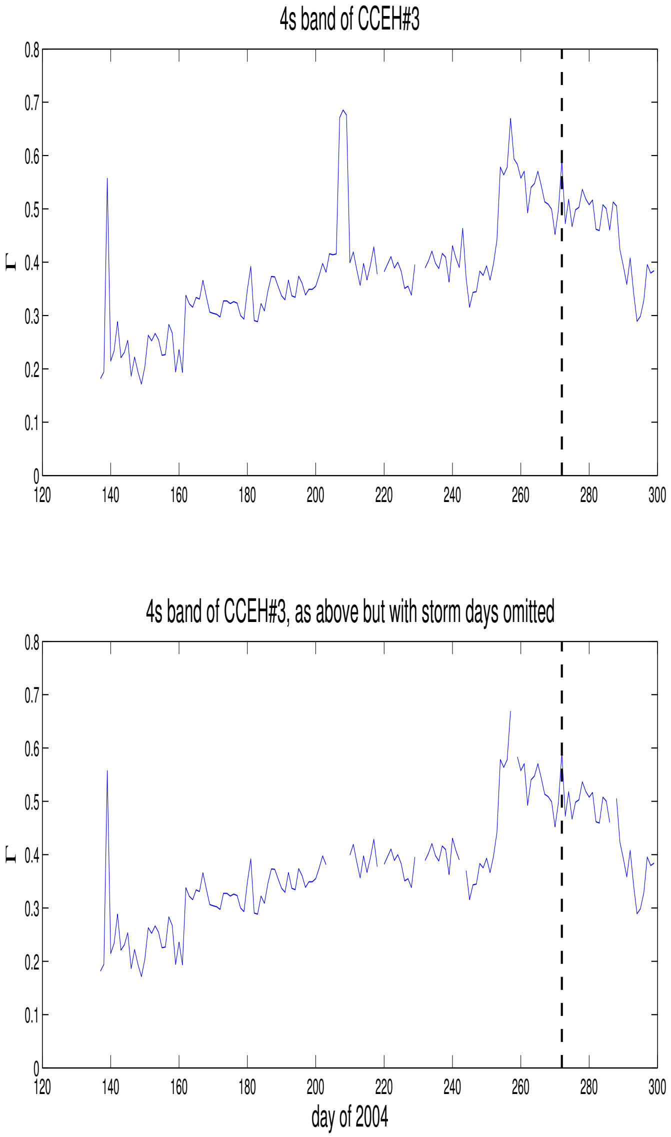


Figure 8.5: The anomalous band in CCHE for the 163-day interval surrounding the 2004 Parkfield earthquake. The top figure includes days with major magnetic storms, while these days are removed in the bottom figure.

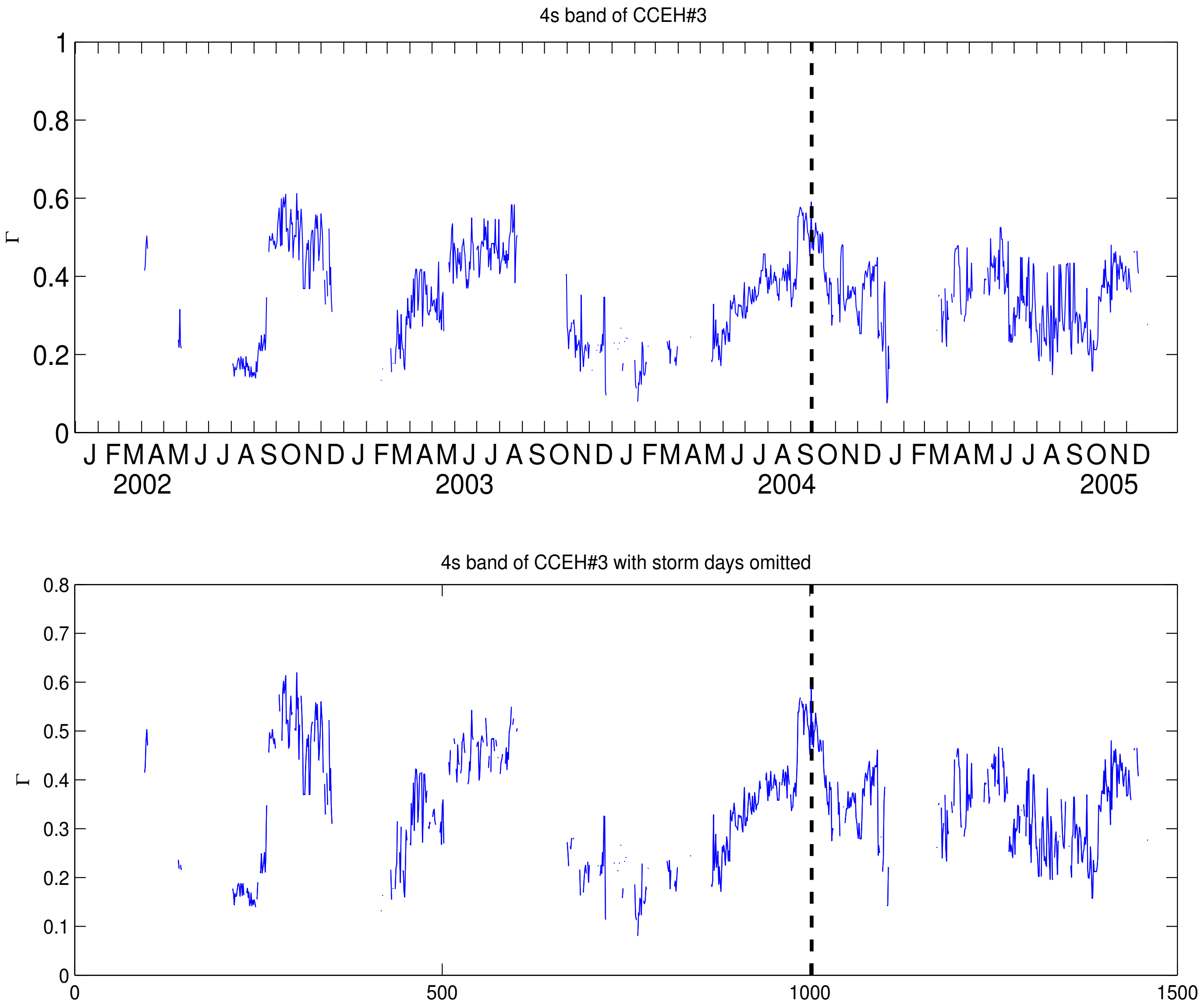


Figure 8.6: The anomalous band in CCHE for 2002-2005. Top figure includes days with major magnetic storms; these days are removed in the bottom figure.

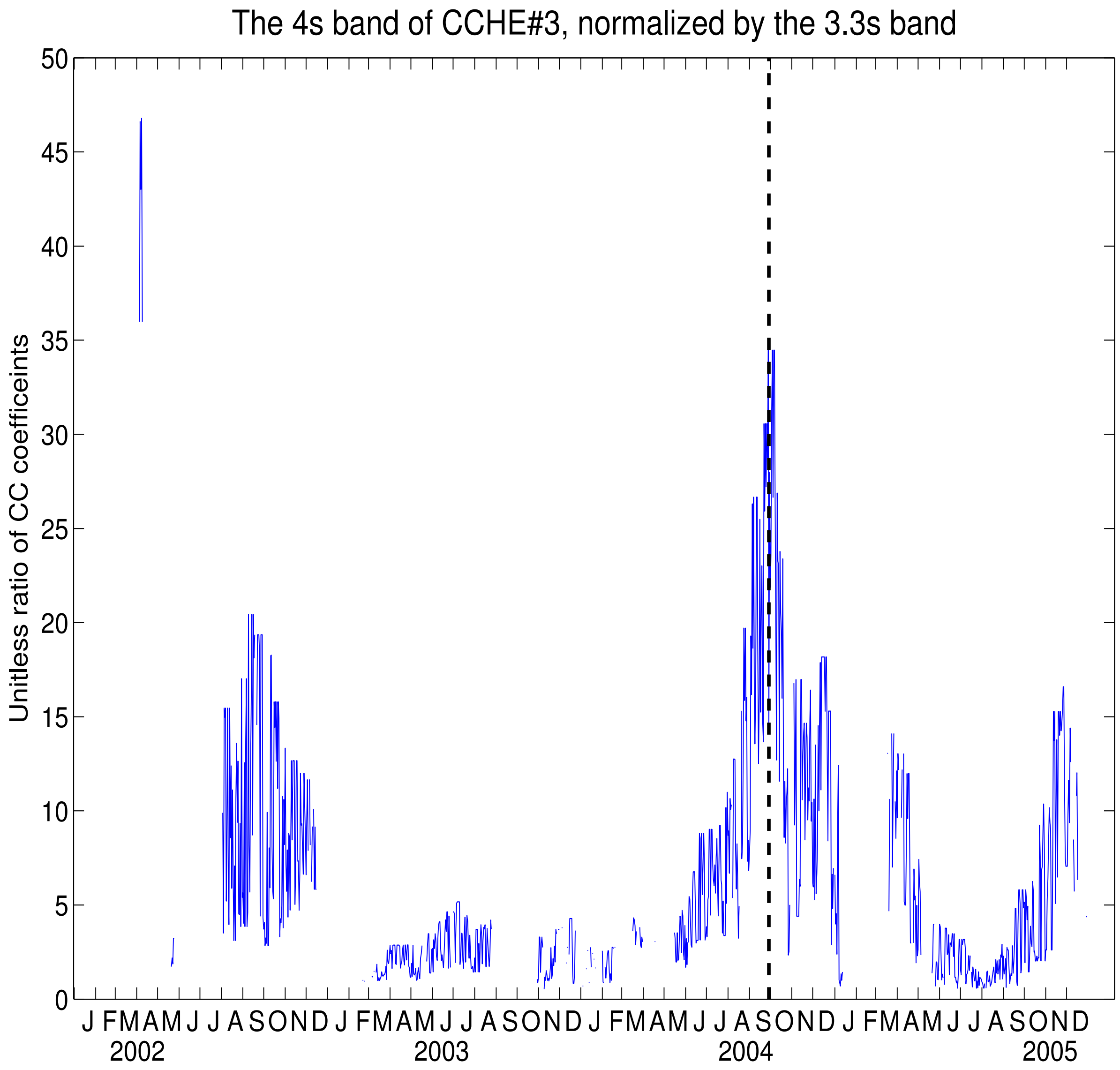


Figure 8.7: The importance of long-term monitoring is highlighted by this plot. Considered only the two years centered around the earthquake, the time series might be interpreted as being related to the earthquake. A wider look at the phenomena, however, shows anomalous variations in 2005, as well as a global maximum in spring of 2002.

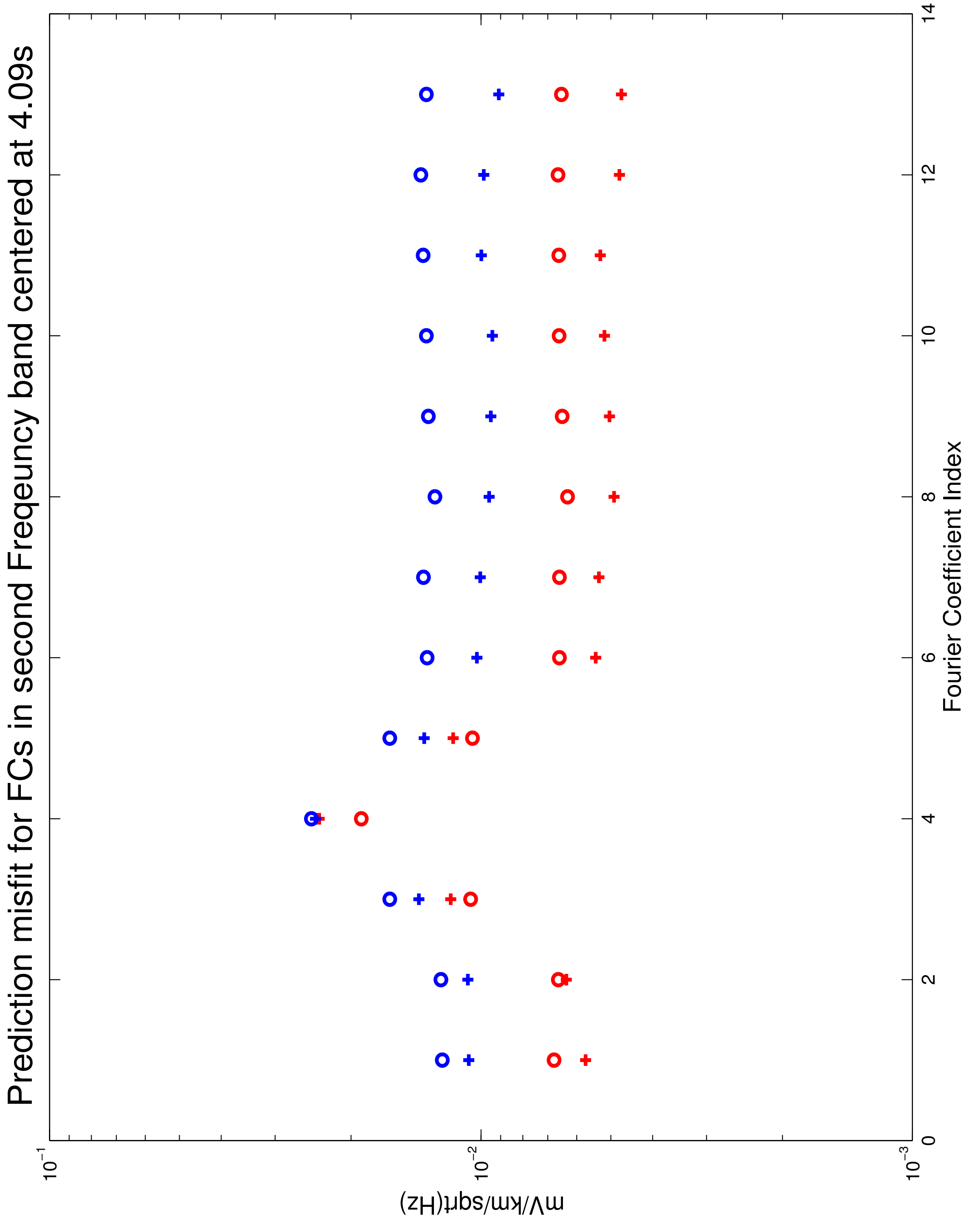


Figure 8.8: Misfit in a narrow frequency band centered at ∼4s period. Observed field FCs in blue, , residuals in red. X's are from day 95, 2002, O's are from day 270, 2004 (2 days prior to the PKD earthquake).

Chapter 9

Apparent Resistivity Variations

9.1 Estimation of the MT impedance tensor and apparent resistivity

The transfer function monitoring described in the array objectives (Chapter 1) generates stable estimates of apparent resistivity (ρ_a) which can be used to track time and frequency (depth) dependent variations in ground conductivity. The magnetotelluric impedance tensor **Z** has the following structure and relationship to measured fields E and H:

$$\begin{pmatrix} E_x \\ E_y \end{pmatrix} = \begin{pmatrix} Z_{xx} & Z_{xy} \\ Z_{yx} & Z_{yy} \end{pmatrix} \begin{pmatrix} H_x \\ H_y \end{pmatrix} \tag{9.1}$$

Here E and H are the locally measured fields.

Equation 3 can be shown to hold perfectly when the usual MT assumptions are in place, i.e. incident fields are plane waves of infinite horizontal extent. In practice, we know that this assumption is not exactly true [*Egbert and Booker*, 1989; *Egbert et al.*, 2000], but is a reasonable approximation at our site latitudes. The off-diagonal elements of the MT impedance tensor are in direct proportion to the square root of the bulk resistivity of the ground.

$$\rho_{ij} = \frac{1}{\mu\omega} \|Z_{ij}^2\| \qquad for \;\; i \neq j \tag{9.2}$$

The method of transfer function calculation is similar to the RMEV method used for SNR estimation. Our technique of calculating the MT impedance tensor, however, involves both local and remote field measurements called the Remote Reference (RR) estimate, which has been found to significantly stabilize estimates of **Z** [*Gamble et al.*, 1979]. The robust RR processing was suggested in *Chave and Thomson* (1989) and is outlined clearly in *Eisel and Egbert* (2002). Minor modifications account for remote reference processing and the inclusion of only other-field-type (OFT) channels.

Using RR processing, the impedance tensor is estimated iteratively by:

$$\mathbf{Z} = (\mathbf{RWH})^{-1}(\mathbf{RWE}) \tag{9.3}$$

The estimation process is initialized using the dataset cleaned during SNR processing in Chapter 4 as our **E**, **H**, and **R** data, where **R** corresponds to the remote site magnetic field data, and **W** is a weighting matrix initially set to **I**. At each step, a diagonal **W** is calculated which pulls outliers in **E**, **H**, **R** toward their expected values as described in *Eisel and Egbert* (2002). Here, 'errors' are measured using a hybrid norm, which is essentially an L^2 norm for residuals within 1.5 standard deviations of the residual probability-density-function, and an L^1 norm otherwise. The process effectively downweights leverage points [*Chave and Thomson*, 1989], and allows for outliers at the reference site. Note that careful, robust estimates for σ are also required at each iteration. Calculation details of these σ estimates can be found in the appendix of *Egbert and Booker* (1986).

The four-year median values for apparent resistivity are shown in Figure 9.1. Figure shows the raw apparent resistivity estimates returned by the PKDSAOg MT processing codes. The plots are in percent deviation from the median. Two features of these plots stand out to first order: evidence for gradual variation of apparent resistivity is present as are broadband offsets in the time series. Frequency independent shifts in the time series cannot be truly related to changes in subsurface conductivity, since each frequency has different depth

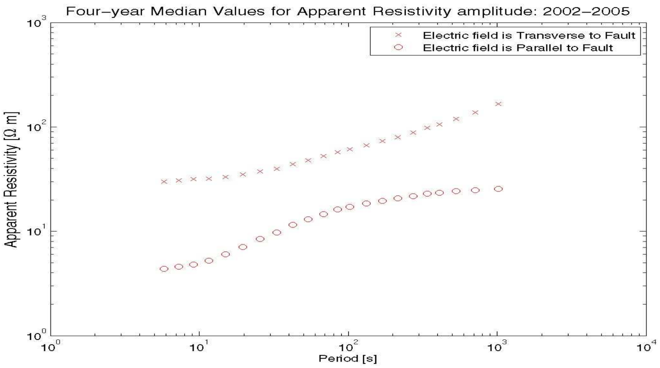


Figure 9.1: Median values of apparent resistivity, x denotes TE, o denotes TM

sensitivity. We attempt to account for these by applying a distortion analysis to the time series.

9.2 Distortion corrections

Frequency independent shifts in the data can be caused by incorrect gain settings on EFSCs, as well as by 'galvanic distortion effects'. These effects can be caused by inhomogeneities in conductivity structure. When quasi-DC electric fields cross boundaries in the conductivity structure, charges accumulate on the boundaries, and give rise to a secondary electric field, which perturbs or 'distorts' the primary field in the region around the inhomogeneity. Galvanic distortion is approximately a frequency independent effect as long as the skin depths associated with the frequencies considered remain large compared to the size of the perturbing inhomogeneities. Phases of the MT tensor are, however, not distorted by galvanic or DC effects. We seek to decouple the daily measured impedance tensors into a product of tensors, where one factor is the complex four-year median impedance tensor, and the other factor is a frequency independent (Real) perturbation (distortion) tensor we label **D**.

Formally, for each day t, and each frequency bin ω, we have an impedance tensor estimate $\mathbf{Z}_{t,\omega} \in \mathrm{M}_{2,2}(\mathbb{C})$. These 2x2 matrices can be concatenated over all frequencies to create $\mathbf{Z}_t \in \mathrm{M}_{2B,2}(\mathbb{C})$, where B is the number of frequency bands under consideration. Here we use B=23 log-linear spaced bands. Similarly, a four-year median impedance tensor at each frequency, which we denote by $\bar{\mathbf{Z}}_{t,\omega}$, can be concatenated over the B frequency bins to make $\bar{\mathbf{Z}} \in \mathrm{M}_{2B,2}(\mathbb{C})$.

We then define the daily distortion tensor $\mathbf{D}_t$ as the least squares solution to

$$\mathbf{0} = \|\mathbf{Z}_t - \bar{\mathbf{Z}}\mathbf{D}_t\| \tag{9.4}$$

where we constrain **D** to be in $\mathrm{M}_{2,2}(\mathbb{R})$.

By multiplying $\mathbf{Z}_t$ on the left by $\mathbf{D}_t^{-1}$, one obtains a a 'distortion-corrected' impedance tensor, which is nearest the long term median impedance tensor in a least-squares sense.

The distortion-corrected apparent resisitivities are plotted on the same scale as the uncorrected data in Figure 9.3, and with a finer scale in Figure 9.4. It is difficult to interpret the variations in Figure 9.4 in terms of earth processes as opposed to simply smoothed noise. Erring on the side of caution, we interpret these variations as noise. Note that though these variations are random from day to day, the four-year time series is not stationary. There seems to be a tendency for the apparent resistivity estimates to be less stable during the rainy season.

The distortion corrections for a few individual bands can be seen in Figures 9.5- 9.7. Most of the variability in the deviation from the median is removed by the distortion corrections. Figure 9.8 shows that the distortion tensor absorbs most of the seasonal variability present in the apparent resistivity time series.

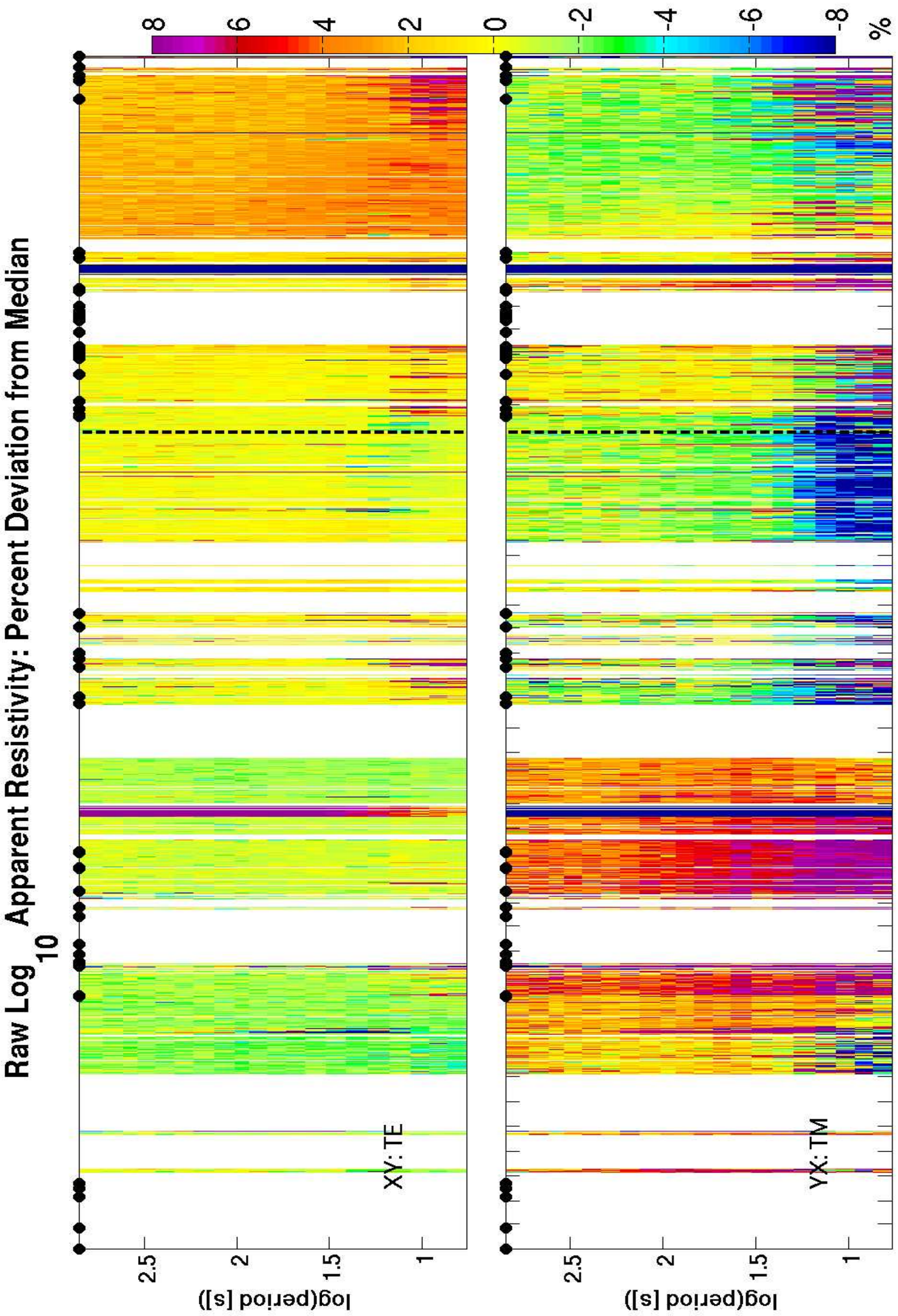


Figure 9.2: Raw apparent resistivity data in percent devition from log-median. TE mode is shown on top, and TM mode on the bottom. Days of significant (> 0.1 inch) rainfall are marked with black circles.

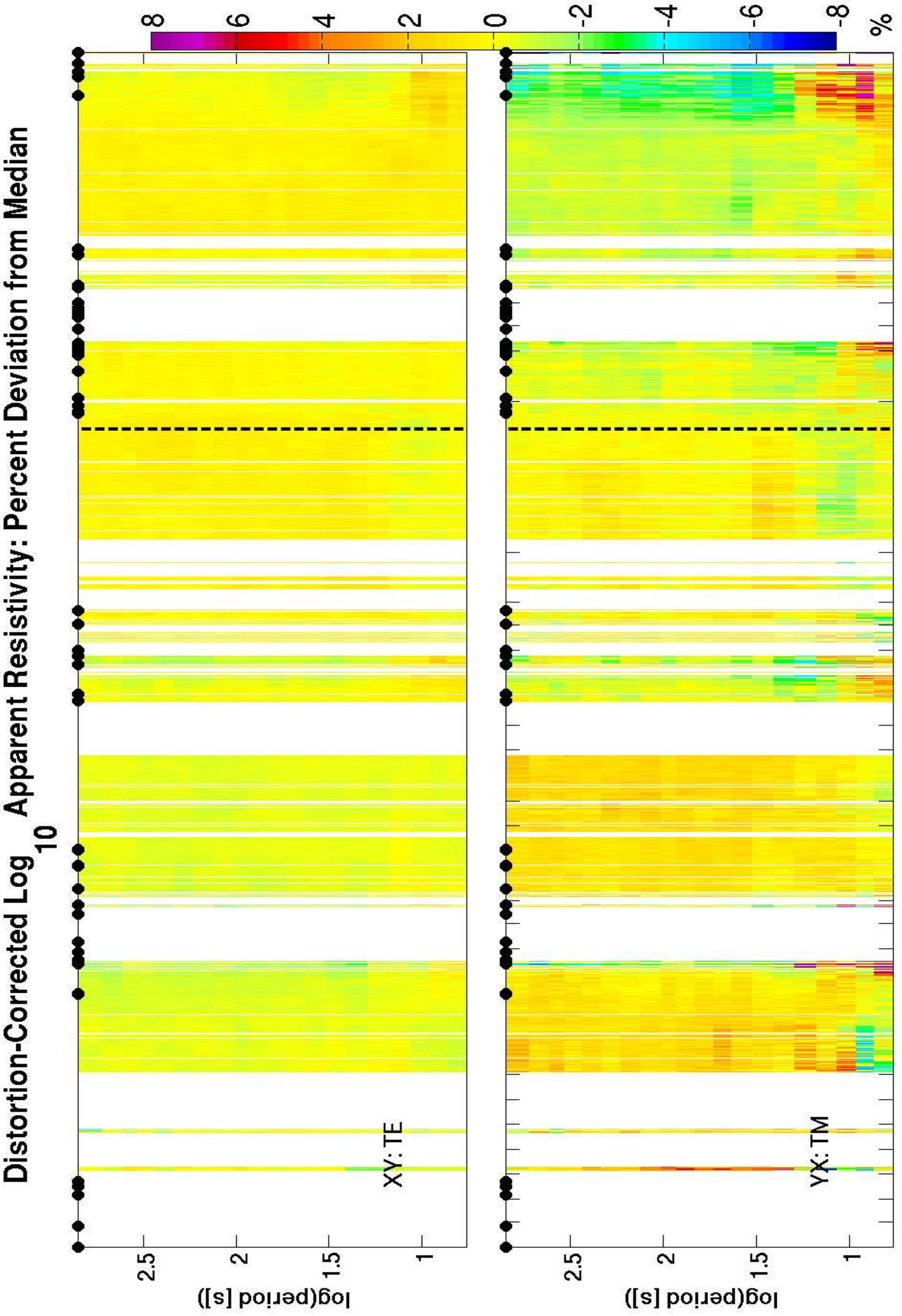


Figure 9.3: Smoothed, distortion-corrected apparent resistivity data from Figure 9.2. The y-axis is $\log_{10}$ of the period in seconds.

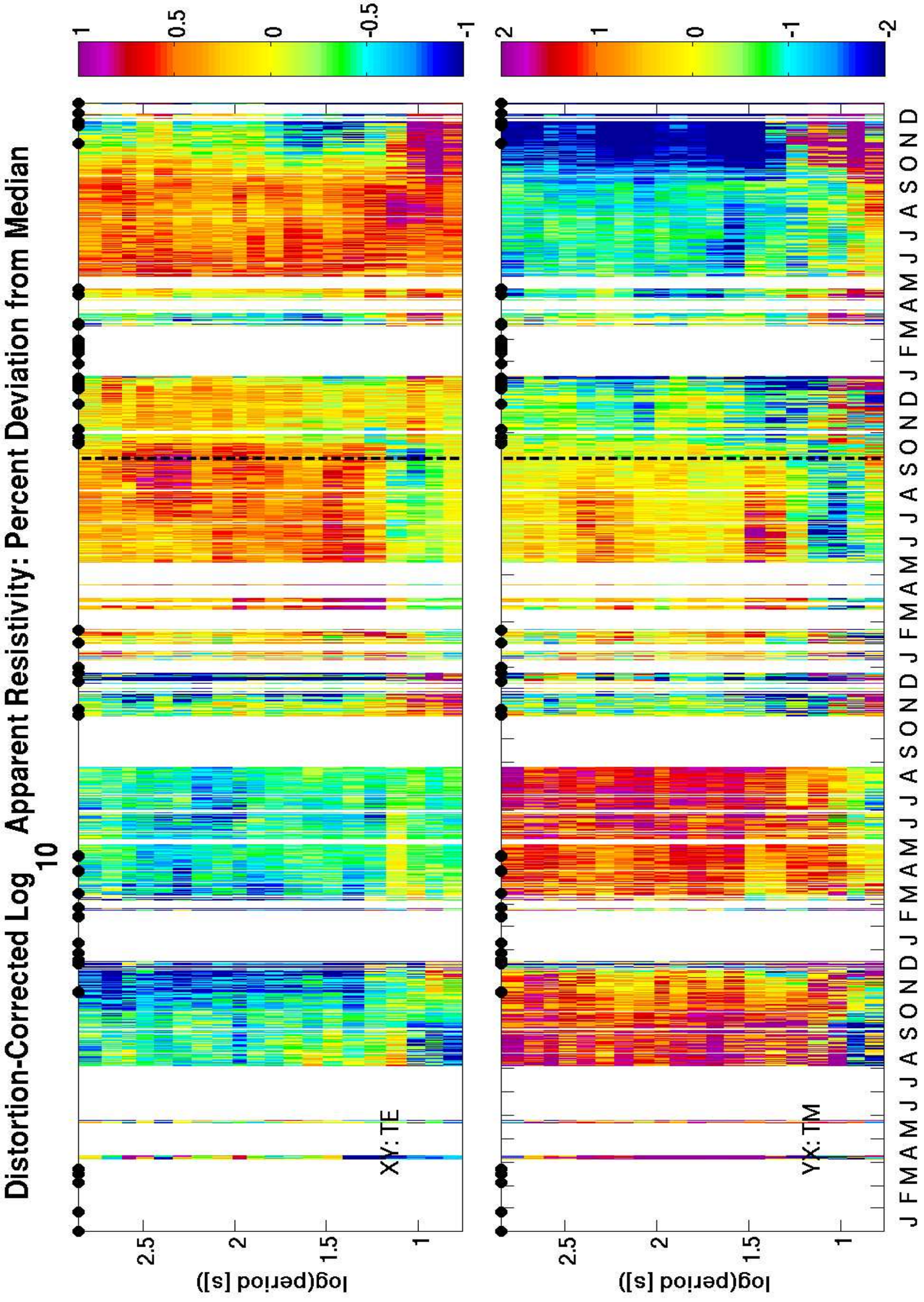


Figure 9.4: Smoothed distortion-corrected apparent resistivity data from Figure 9.3. Colour scale has been contracted from the previous plot, but is still in percent deviation from the median.

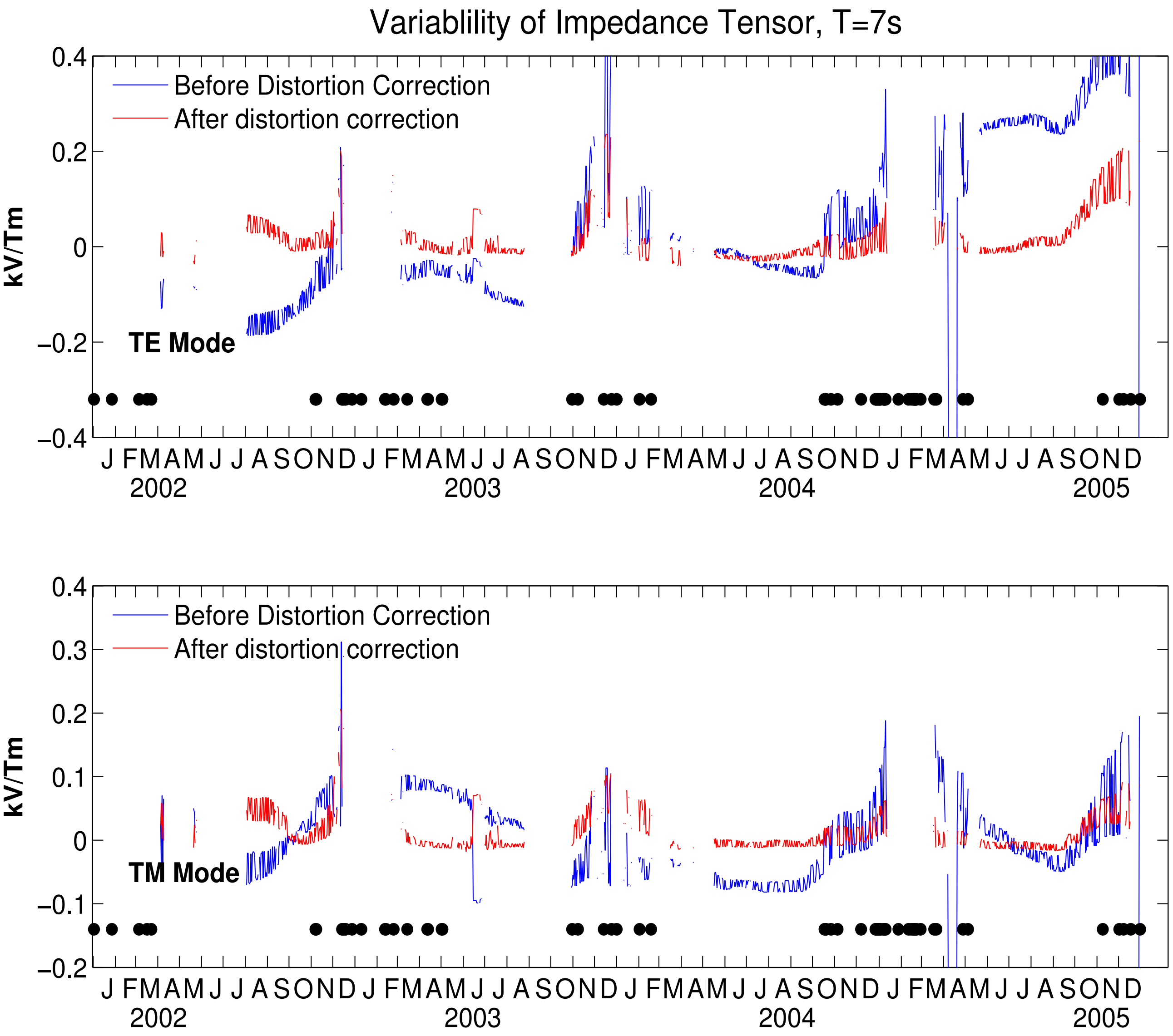


Figure 9.5: Time series of apparent resistivity deviation from median, and distortion-corrected time series deviation from the median. Band centered at 7sec period.

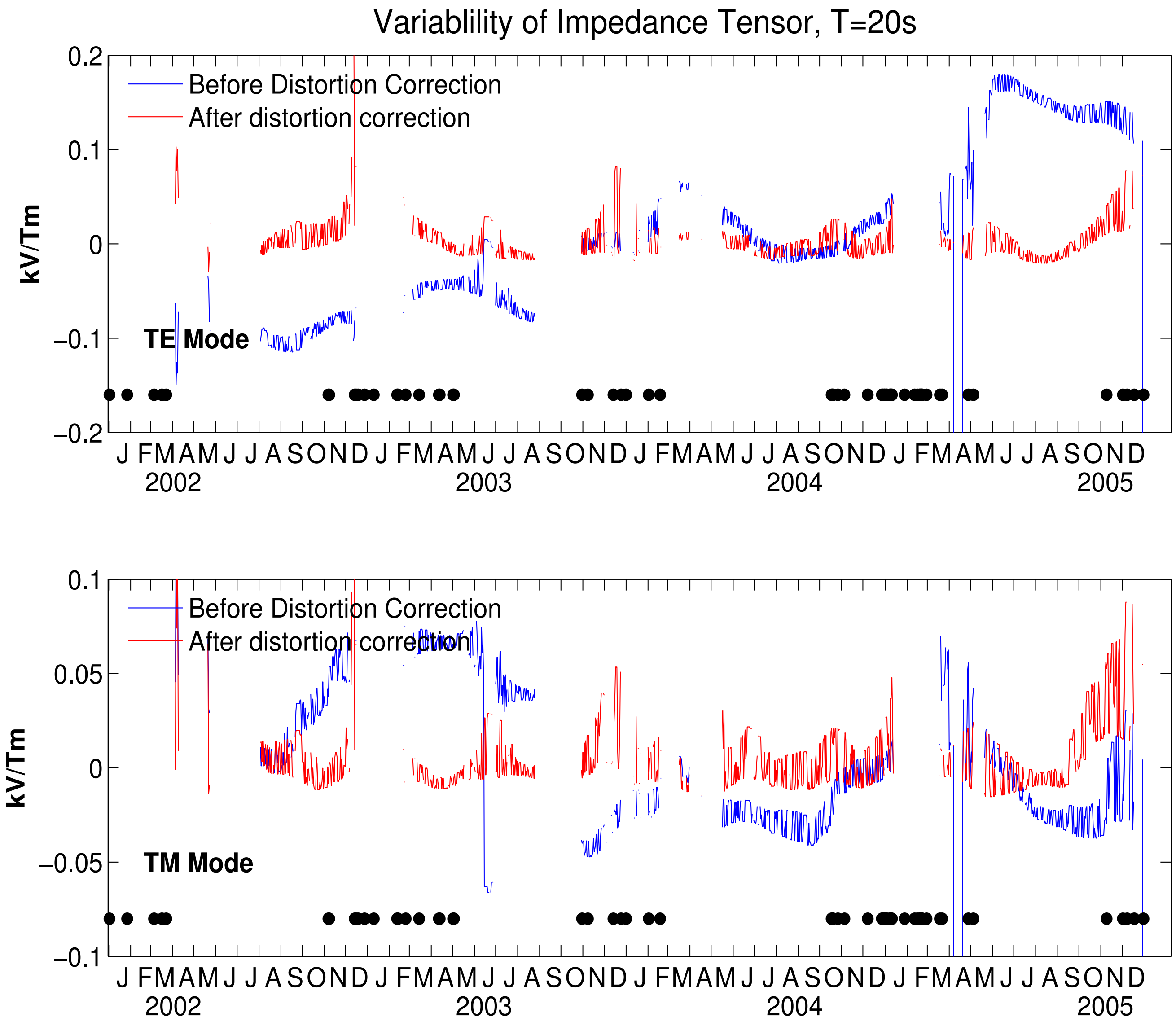


Figure 9.6: Time series of apparent resistivity deviation from median, and distortion-corrected time series deviation from the median. Band centered at 20sec period.

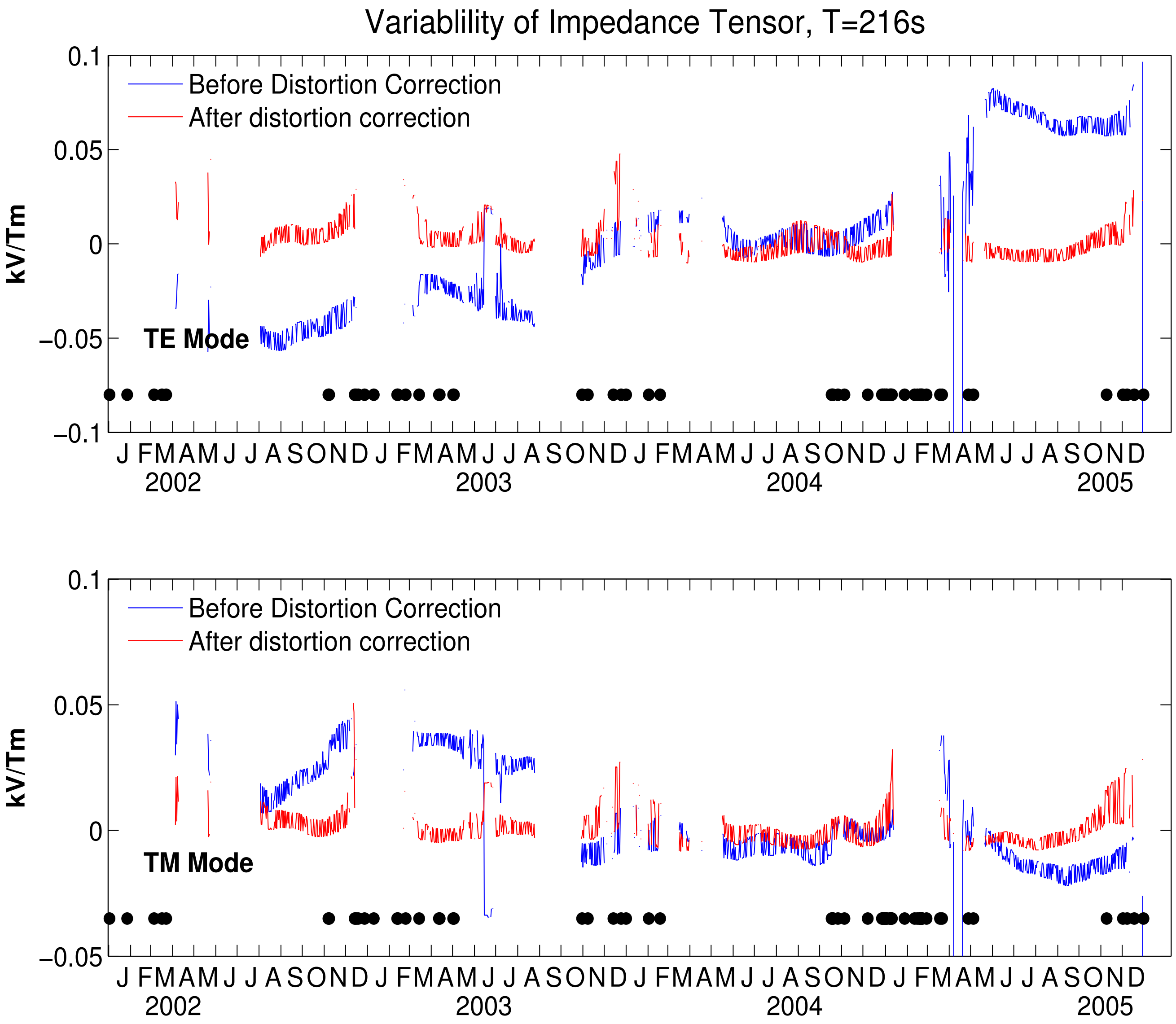


Figure 9.7: Time series of apparent resistivity deviation from median, and distortion-corrected time series deviation from the median. Band centered at 216sec period.

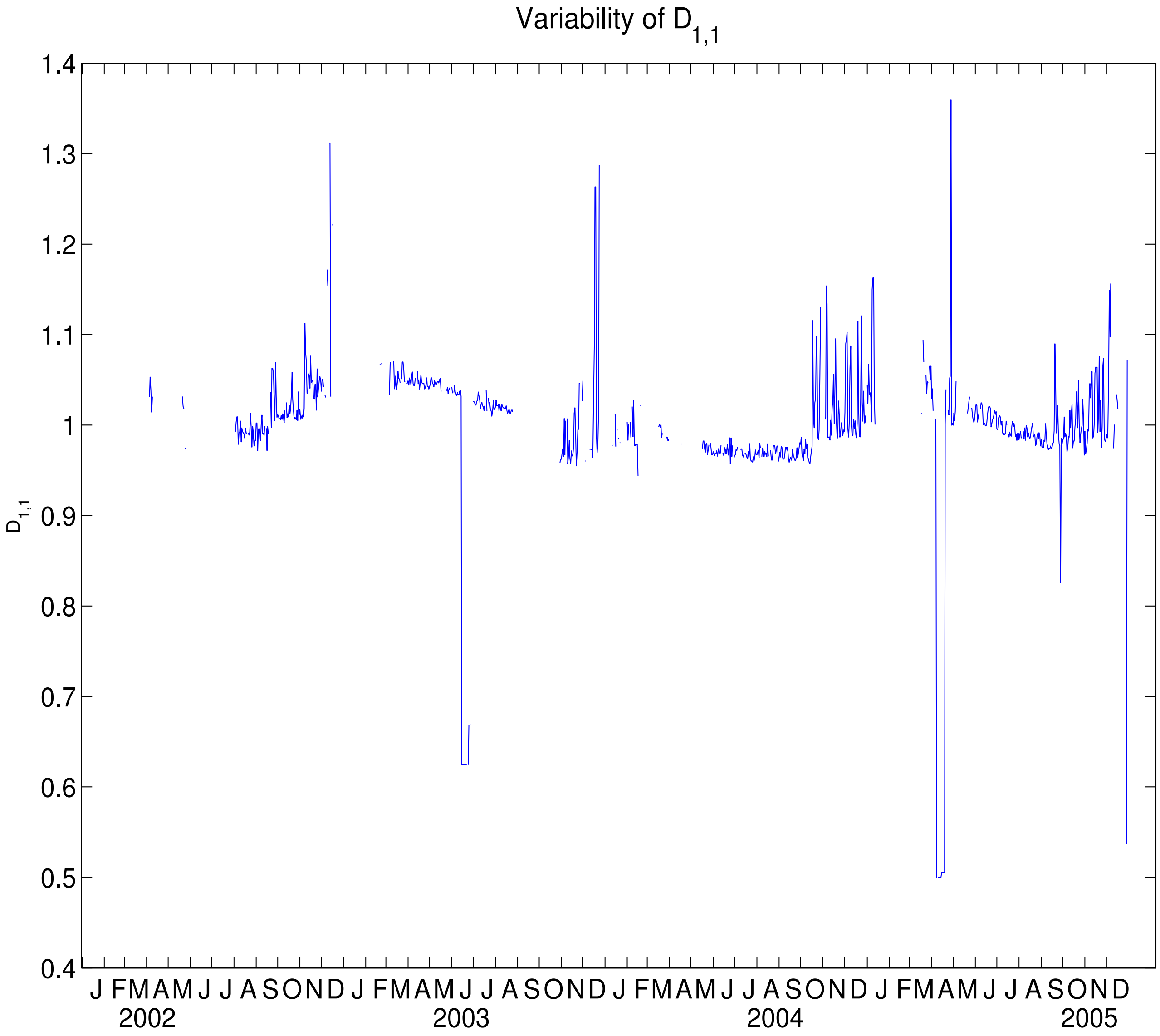


Figure 9.8: Variability of the 1,1 element of the distortion tensor over the 2002-2005 time interval.

Chapter 10

Selections of 40Hz Data

10.1 Overview

Up to this point, most of the analysis was based on 1Hz data, and spanned periods between 2.3 and 1500s. We perform a cursory examination of the 40Hz during a few weeks around the time of the earthquake (days 230-305, 2004). All 40 Hz data were processed using Gary Egbert's RMEV/MMT software available from MTnet. All processed data were found to suffer from a static shift of a factor of 40, which is corrected for before plotting. We perform the same basic analysis on the 40Hz data as on the 1Hz data, except that the data are processed in two hour segments. That is why Figures 10.1-10.14 are pixelated so that there are 12 pixels per day at each of the 57 frequency bins, which span periods of 0.06s

(16Hz) to 40s.

10.2 Signal amplitude

Figures 10.1-10.4 show the signal amplitudes observed by all sensors in two-hour bins. There is clear evidence of broadband diurnal phenomena, especially in the BART band, from 20-40s. The Schumann resonances at $\sim$ 8 and 14 Hz are clearly discernible at the high frequency end of each plot. The energetic band identified in Figure 8.8 at 4s is clearly visible in the y-electrode (NS). In all signal plots, there are occasional bursts of energy in the dead band (1-10s). One of these so-called PC1s occurs almost coincident with the earthquake. We note that the PC1 is present at both sites, which makes it unlikely to be related to the earthquake. Also interesting in Figure 10.2 is the sporadic signal centered around $10^{-0.5}$s, or $\sim$ 3Hz, which wanders in frequency. It is present around 1s and 3s period between days 280 and 290 in the Hollister channels (Figures 10.4 and 10.2).

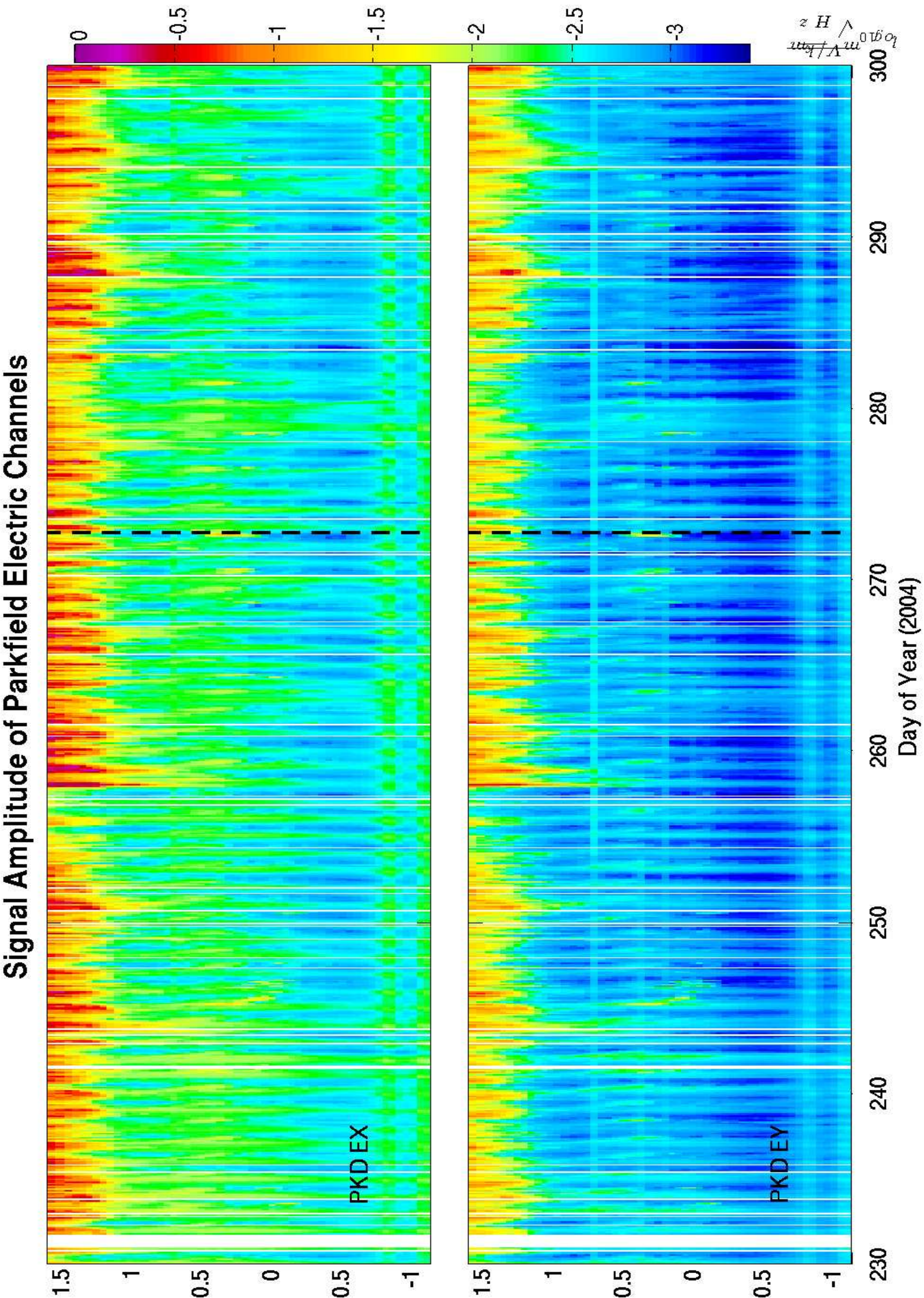


Figure 10.1: Signal amplitude in PKD electric channels at periods from 30s to 0.07s.

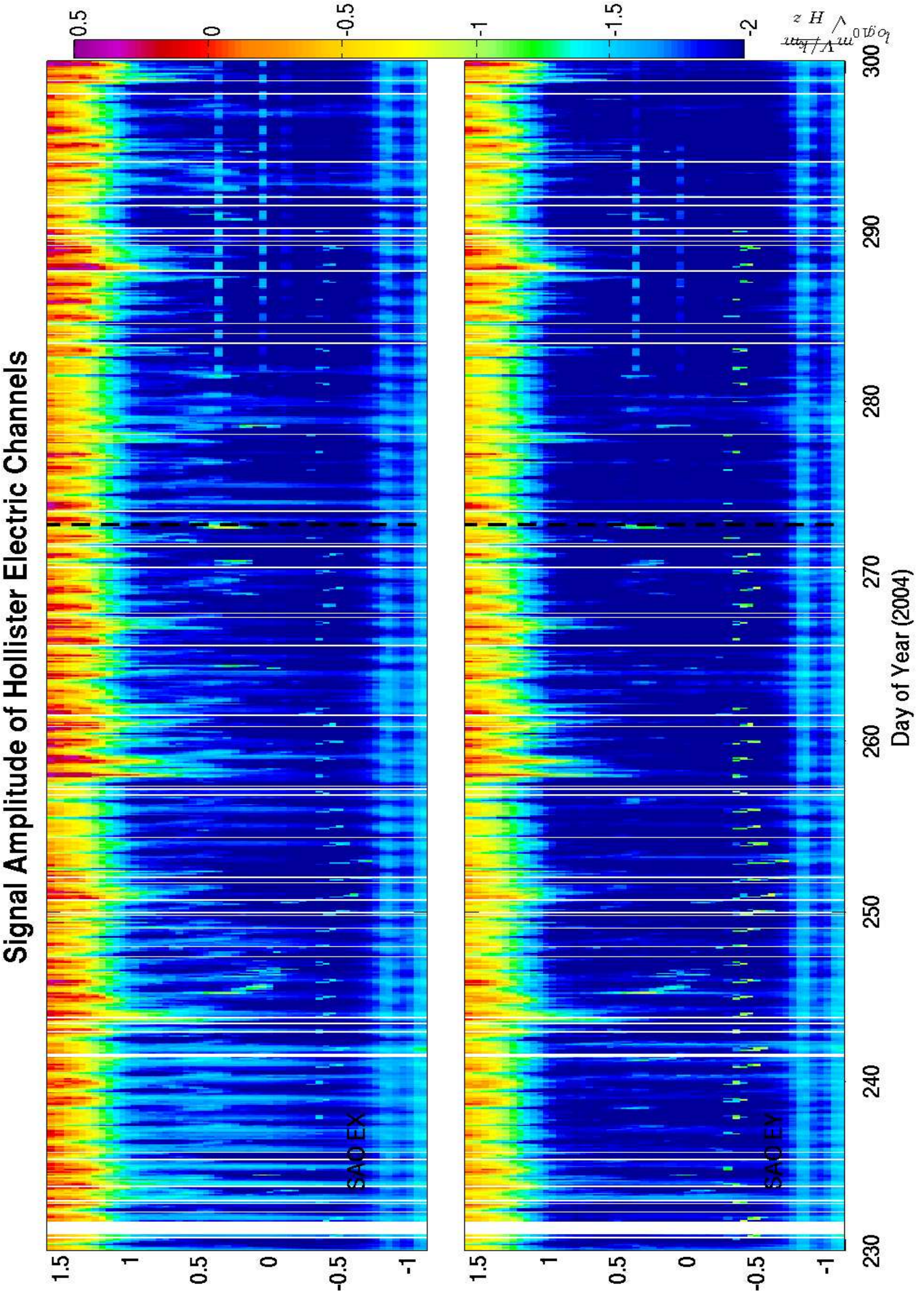


Figure 10.2: Signal amplitude in SAO electric channels in periods from 30s to 0.07s.

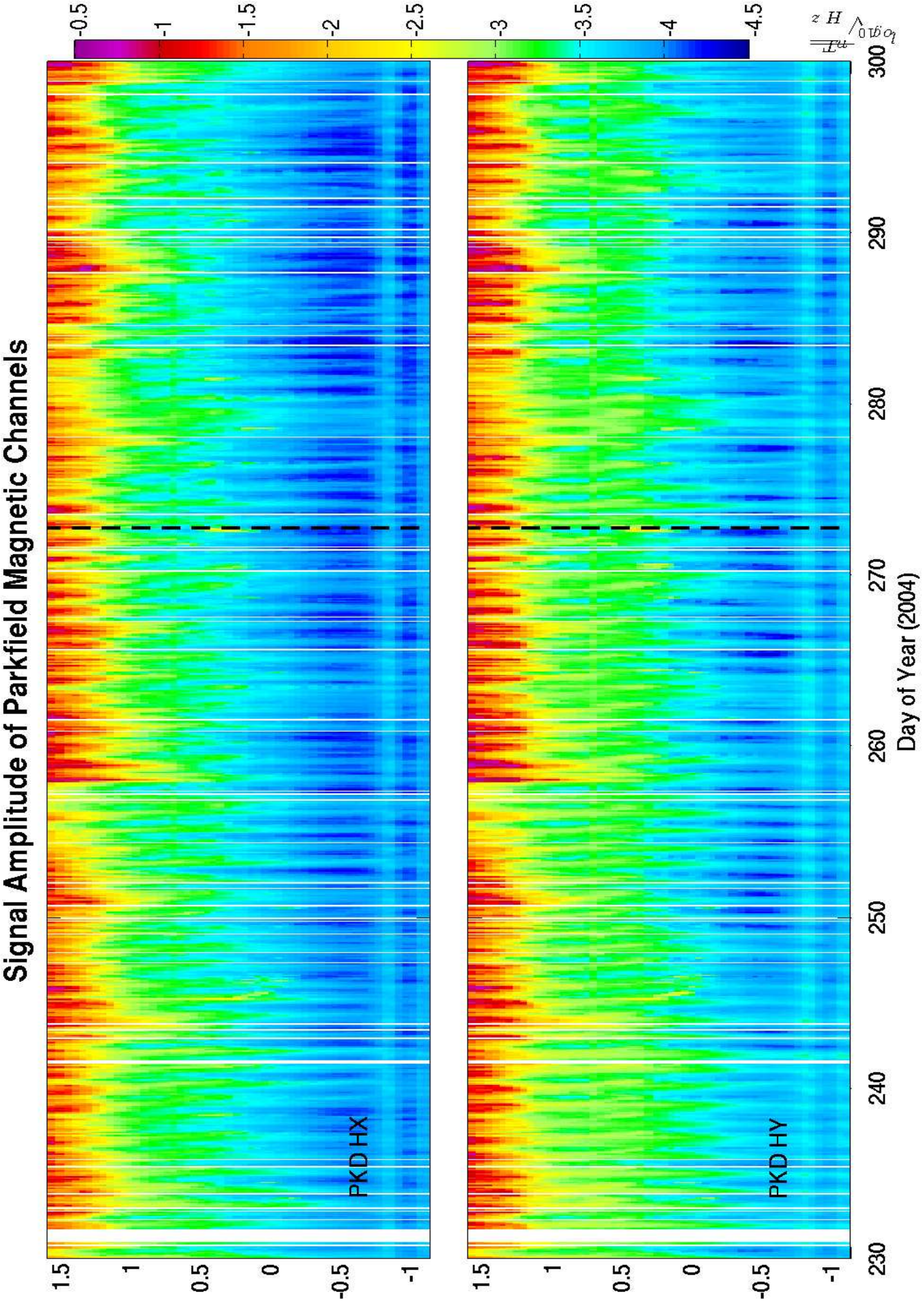


Figure 10.3: Signal amplitude in PKD magnetic channels in periods from 30s to 0.07s.

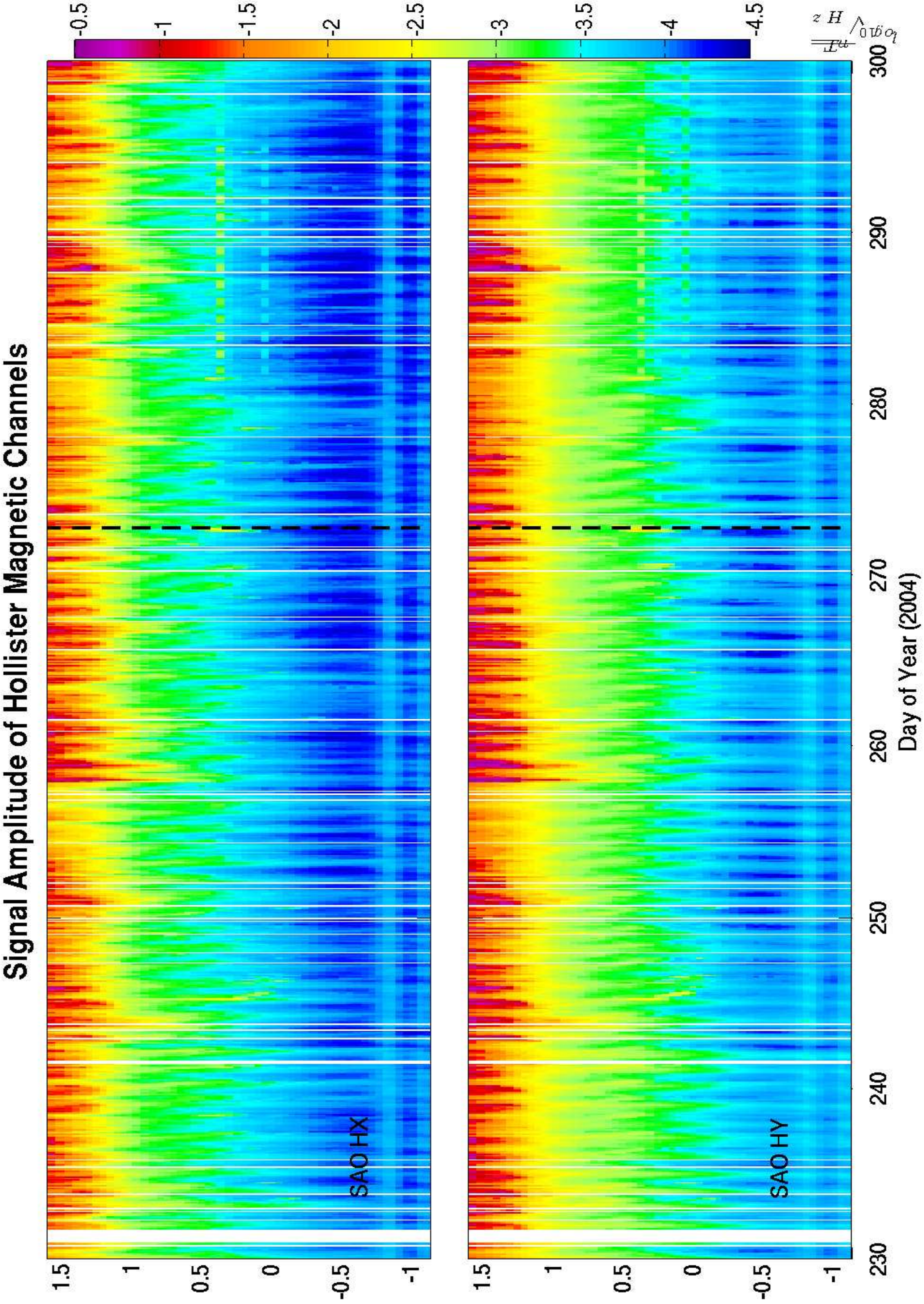


Figure 10.4: Signal amplitude in SAO magnetic channels in periods from 30s to 0.07s.

10.3 Noise amplitude

Noise plots are also calculated over each 2 hour window for the 75 day dataset, exhibiting four remarkable features. First, the noise shows a high degree of stability. Second, the Hy sensor at SAO (Figure 10.8, lower) has significantly higher noise across all bands when compared with the other magnetic sensors (Figure 10.8, upper, and Figure 10.7). Third, the wandering (in central frequency) diurnal pulses can be seen to inhabit the SAO electric sensors exclusively. Fourth, there is a persistent signal at around 10s in SAO Ey which cannot be seen in any other channels. This may be related to the persistent anomaly in the fourth eigenvalue mentioned in description of Figure 7.6. Note also that the noise amplitude of the Parkfield electric channels increases around day 290. This is coincident with the onset of the winter rains at PKD, which appear to affect PKD much more significantly than SAO.

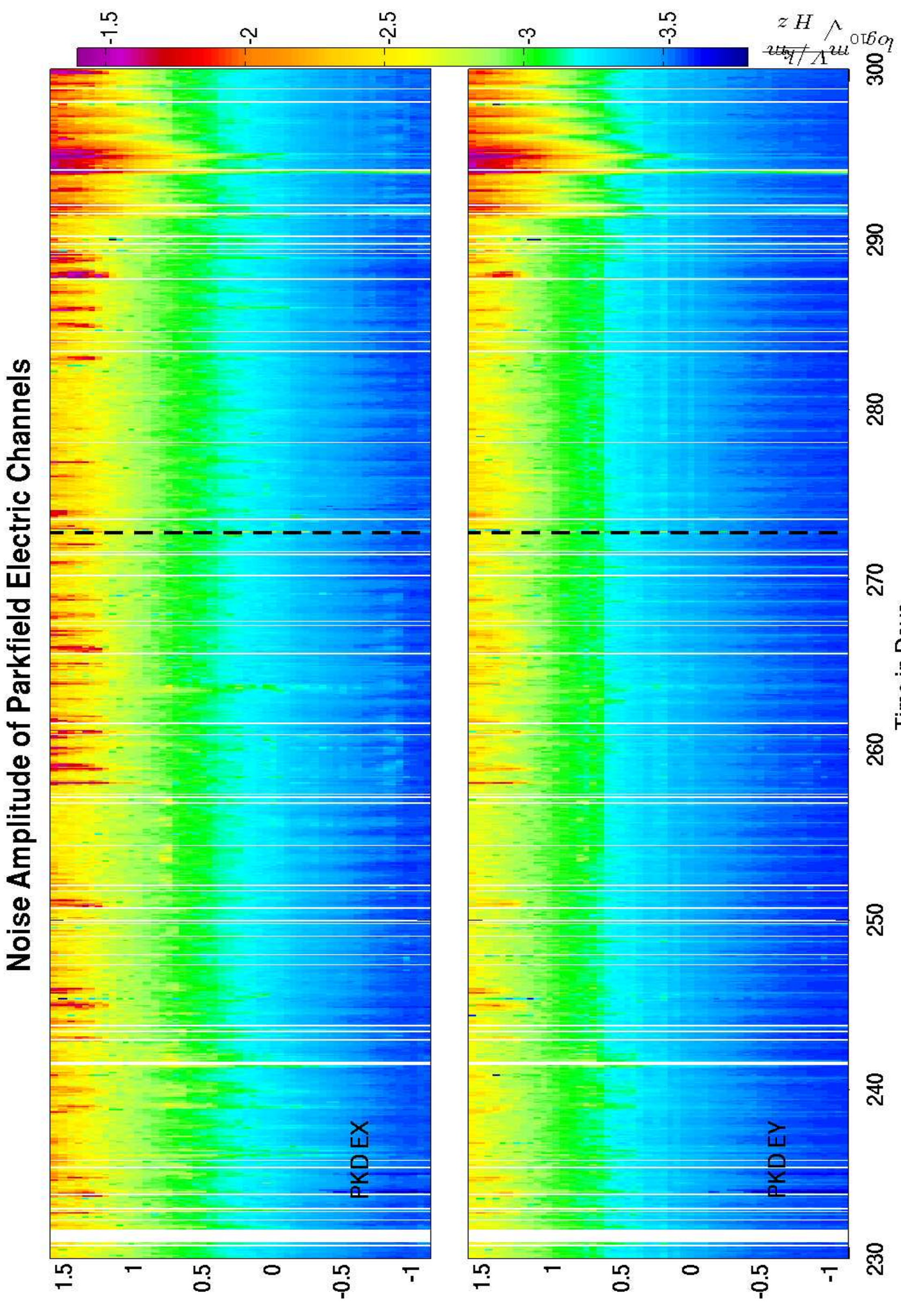


Figure 10.5: Noise amplitude in PKD electric channels in periods from 30s to 0.07s.

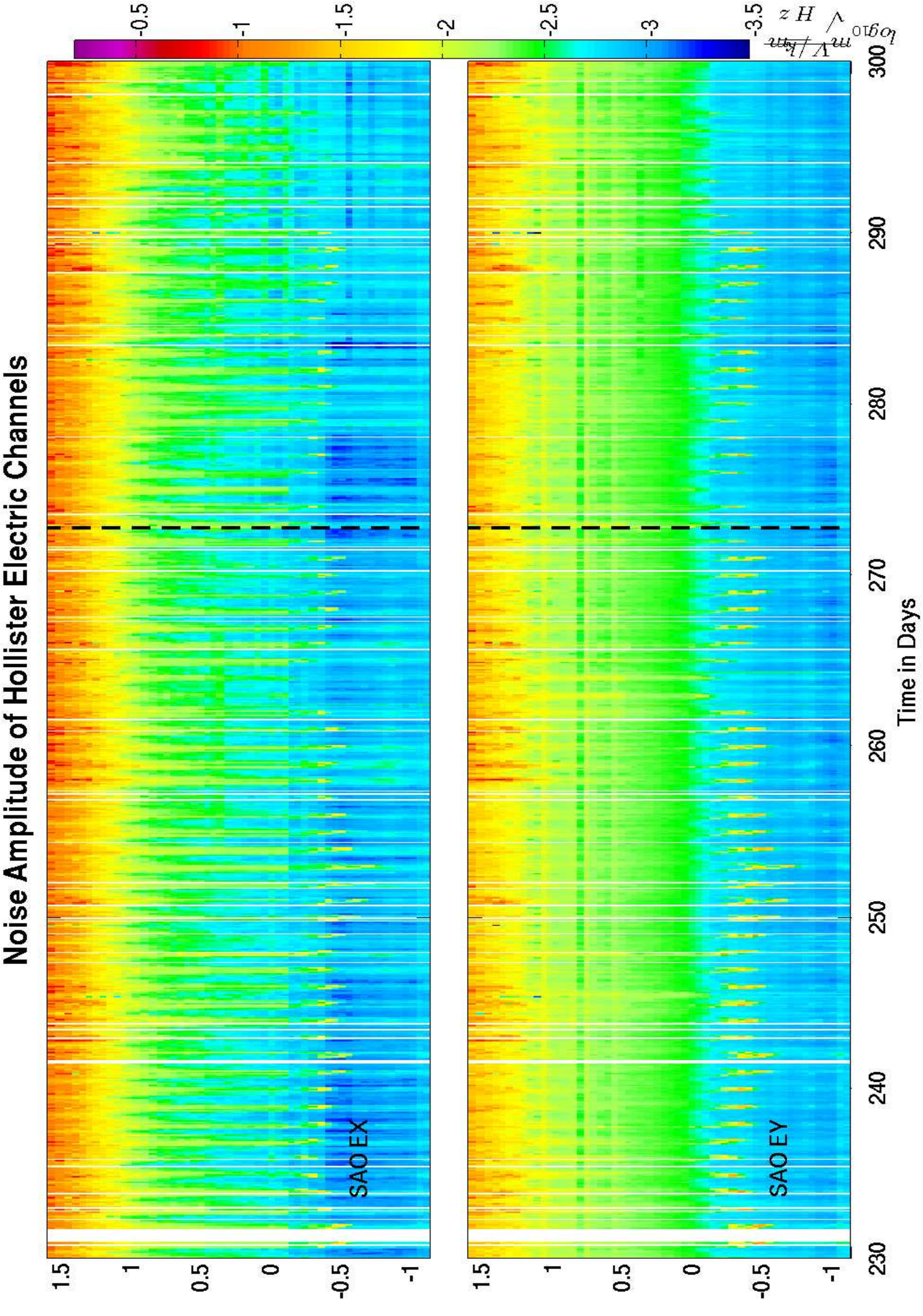


Figure 10.6: Noise amplitude in SAO electric channels in periods from 30s to 0.07s.

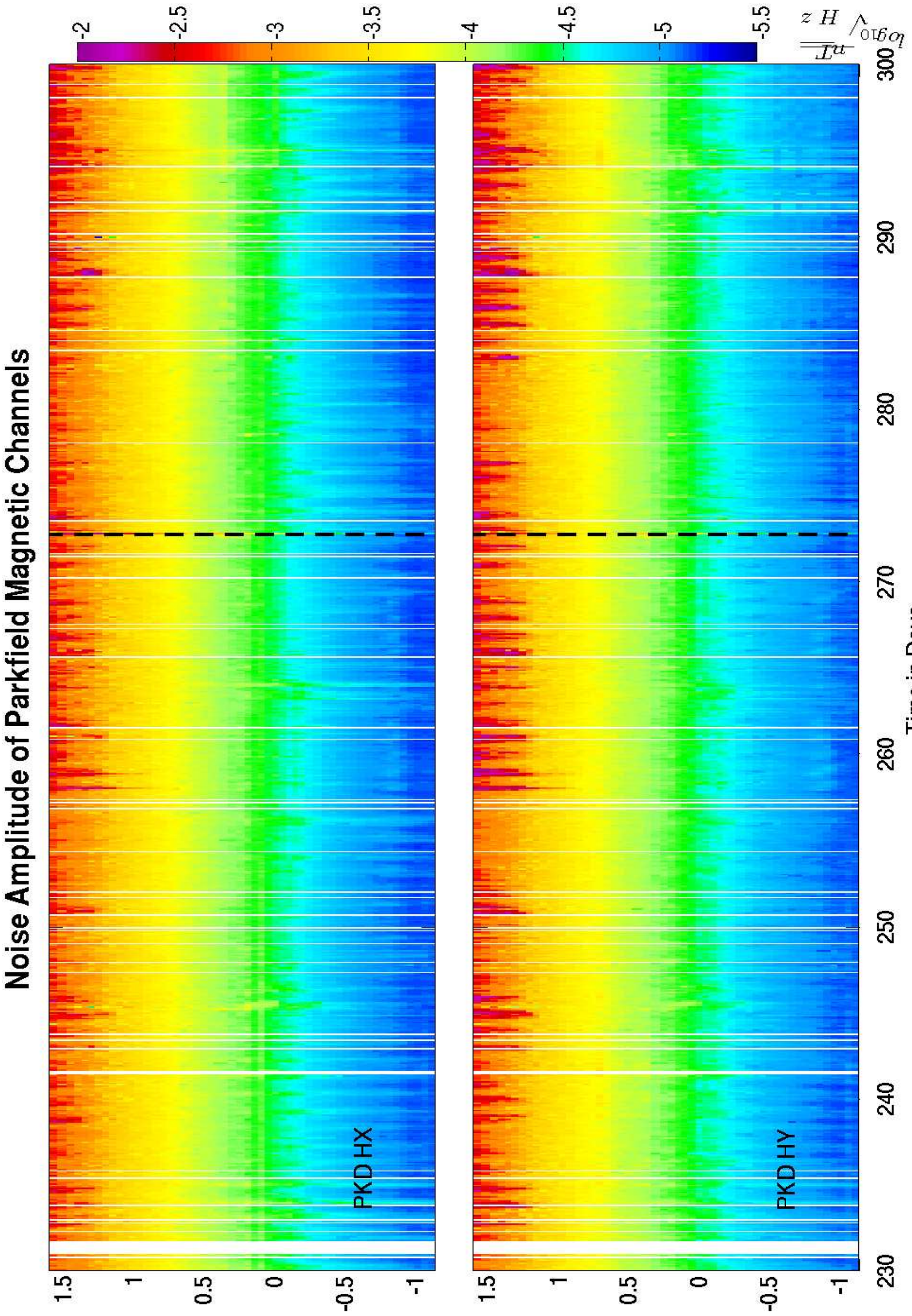


Figure 10.7: Noise amplitude in PKD magnetic channels in periods from 30s to 0.07s.

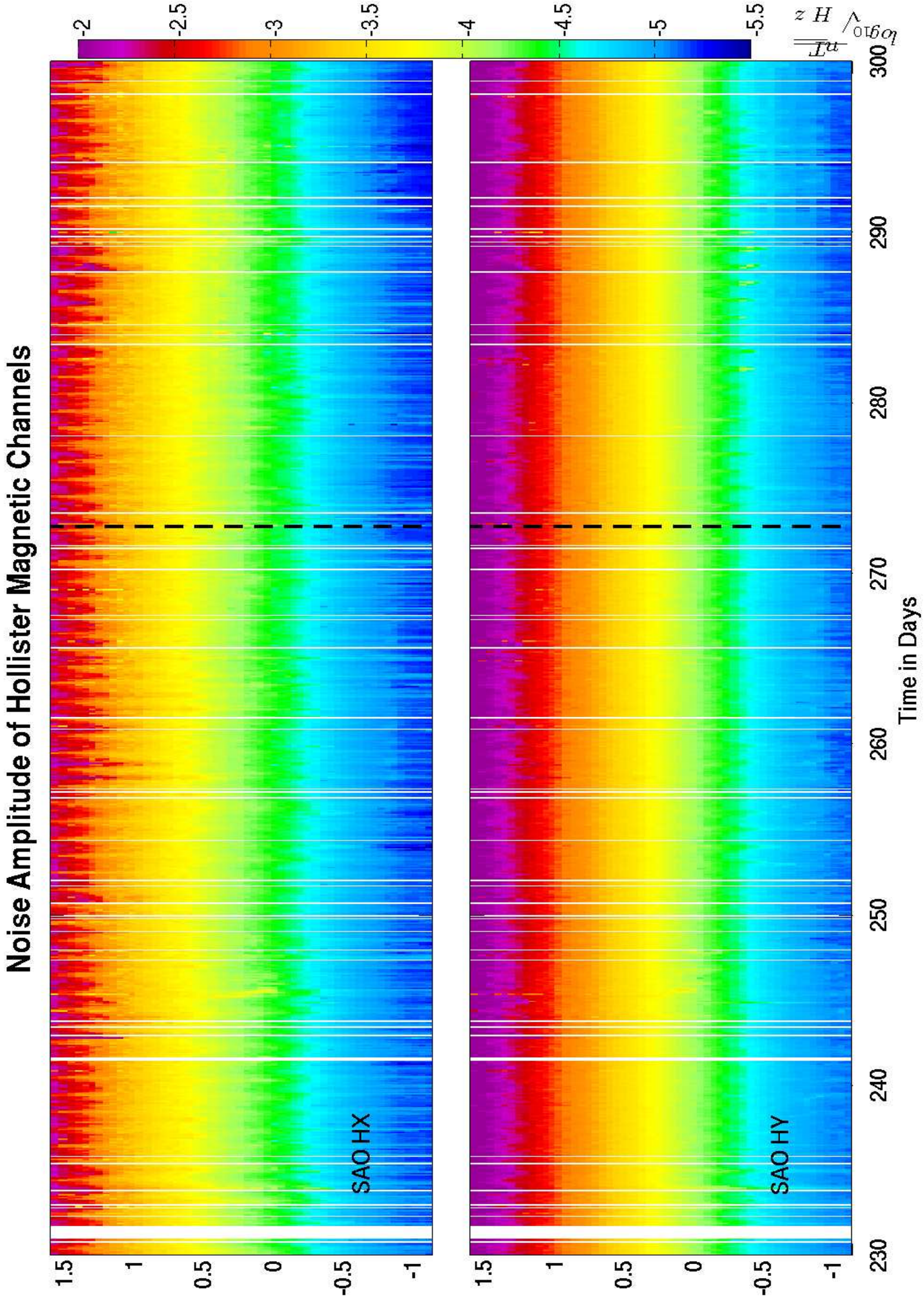


Figure 10.8: Noise amplitude in SAO magnetic channels in periods from 30s to 0.07s.

10.4 SNR

Signal to noise ratio plots of the 40Hz data bear evidence of most of the same phenomena as seen by the 1Hz data. In particular, the highly diurnal character of the signals, especially those associated with the BART is quite apparent. Additionally, Schumann resonances are clearly visible, as are occasional PC1 bursts which are present in all channels. The pulses between days 280 and 300 at SAO at 1s and 3s period are also particularly noticeable. Finally, the persistent signal around 4s period at Parkfield is clearly visible in the y-component electrode SNR.

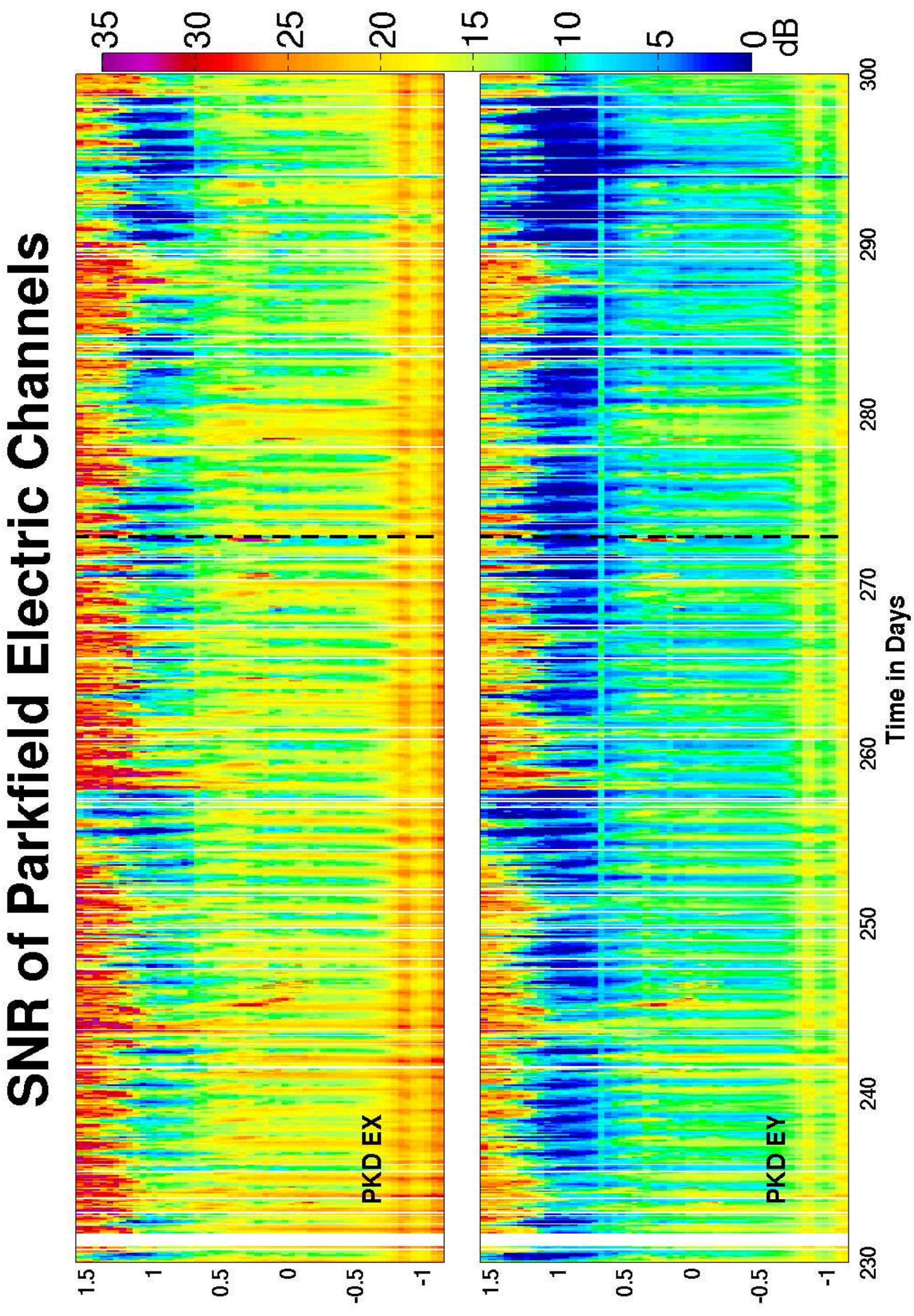


Figure 10.9: Signal to noise ratio in PKD electric channels in periods from 30s to 0.07s.

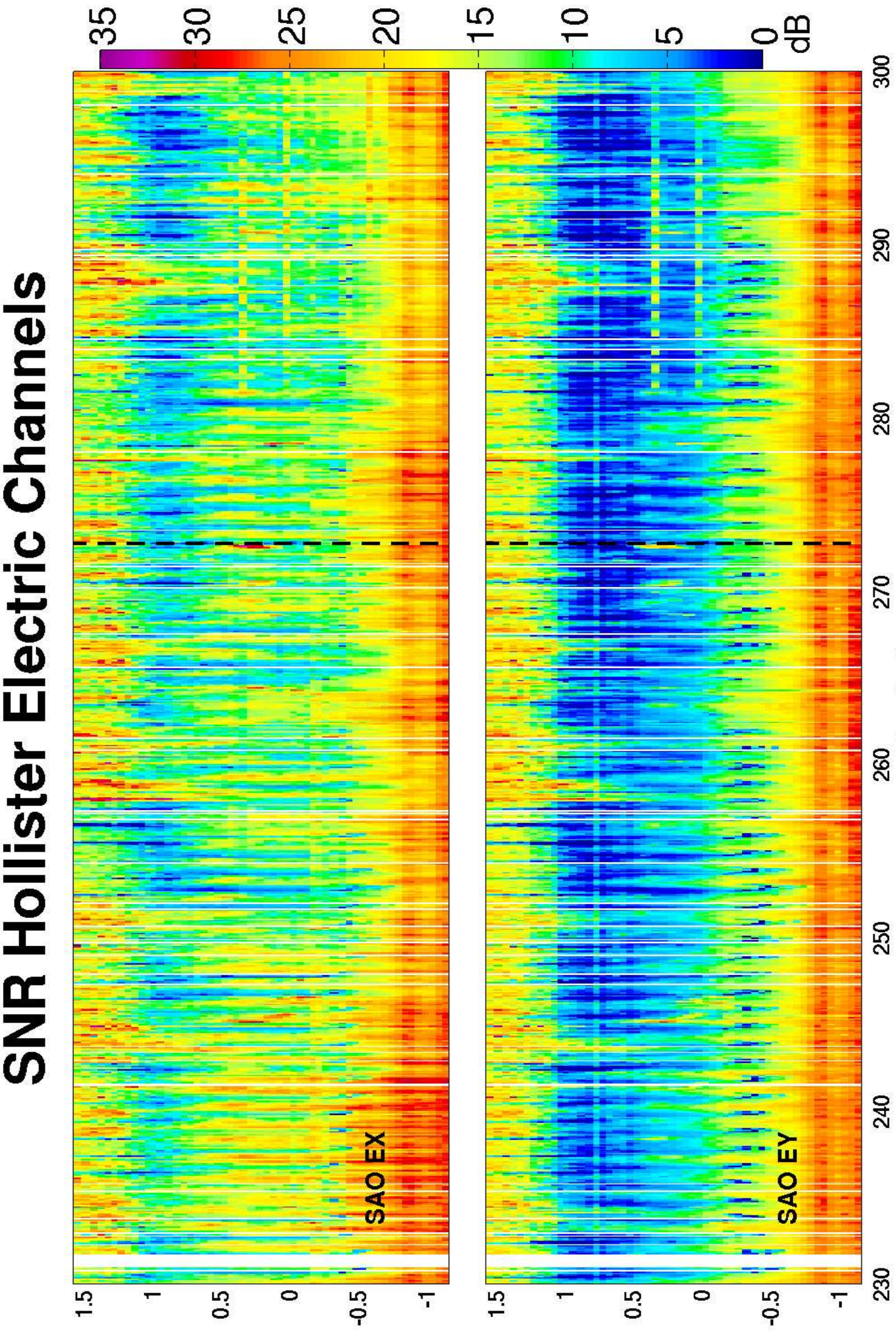


Figure 10.10: Signal to noise ratio in SAO electric channels in periods from 30s to 0.07s.

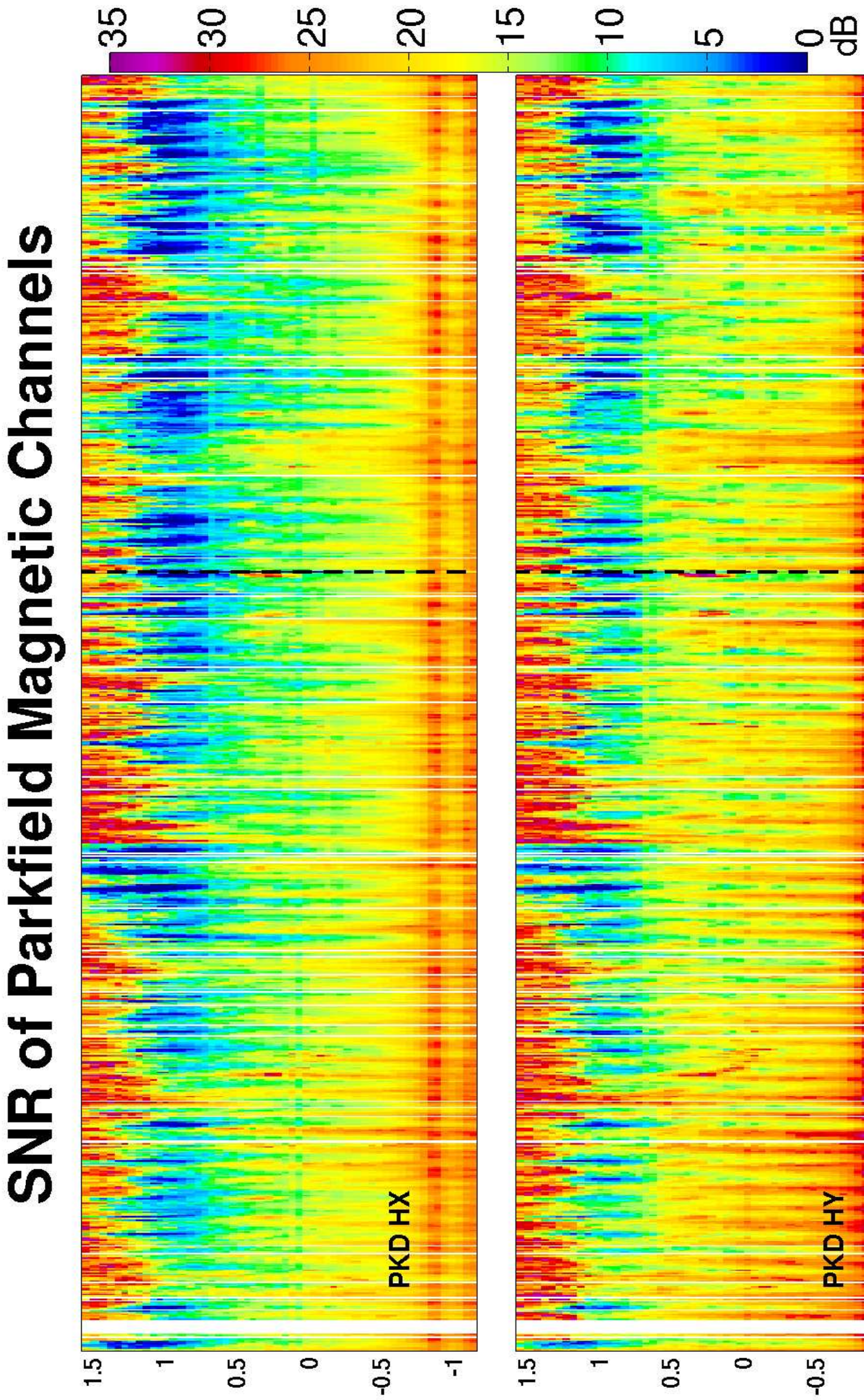


Figure 10.11: Signal to noise ratio in PKD magnetic channels in periods from 30s to 0.07s.

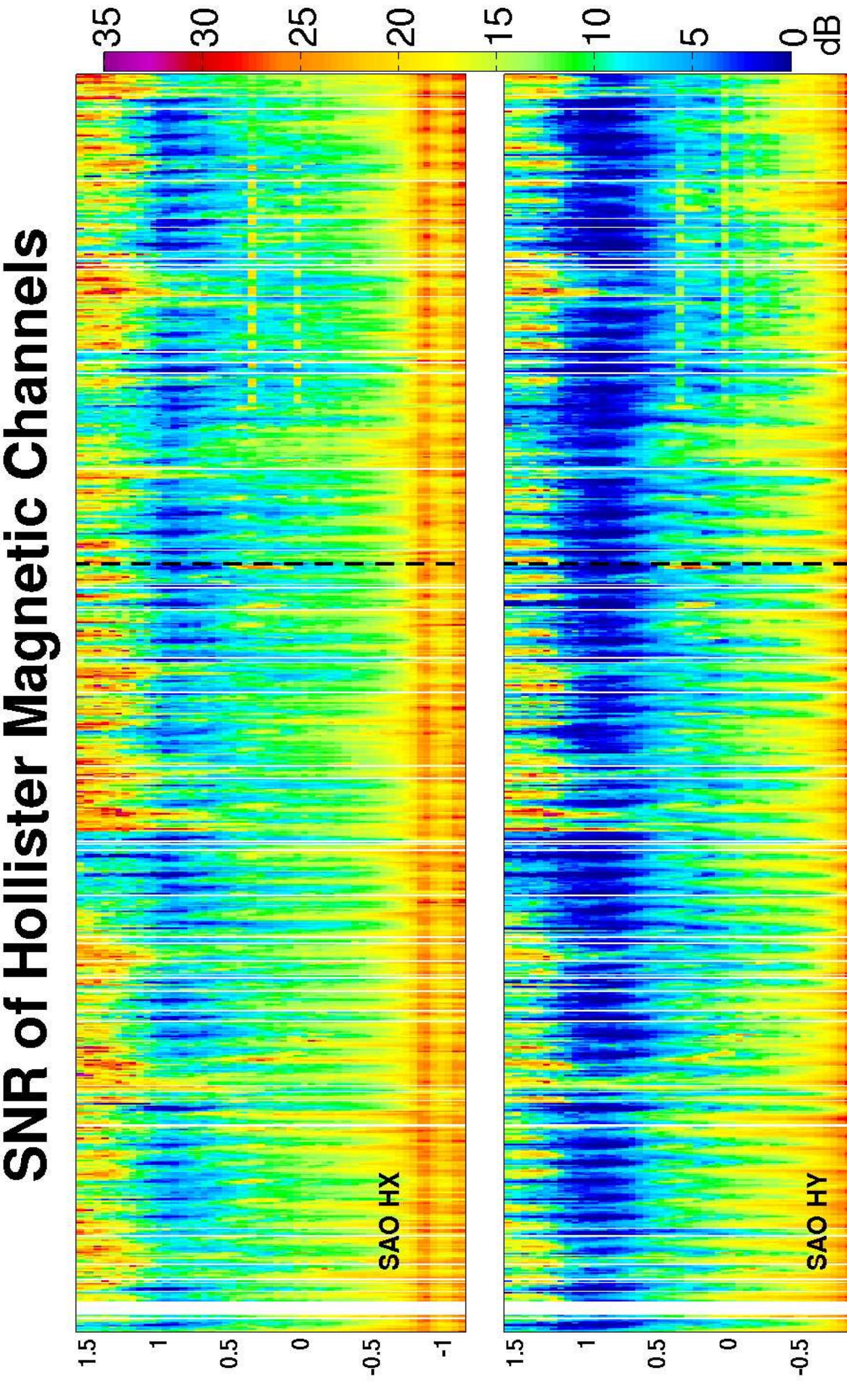


Figure 10.12: Signal to noise ratio in SAO magnetic channels in periods from 30s to 0.07s.

10.5 Eigenvalues

Plots of variation of the dominant four eigenvalues, shown in Figures 10.13 and 10.14 display most of the signals observed in individual channels. Particularly noticeable is the crispness of BART's diurnal period (Figure 10.14), which is due to the 3-4 hour break in train service every night. Note that the Schumann bands are only present in the first two principal components, as is to be expected from a 2-dimensional signal. The PC1s are visible across all modes, just as the major solar storms were in the 1Hz modes. A variety of persistent, narrow band signals can be seen in the third and fourth eigenvalue plots. Most narrow band signals tend to show up in the lower principal components. The dashed lines at 1 and 3s period (observed at SAO) seem to inhabit not only the third and fourth average modes, but also the second mode. These signals stand alone as the only anomalous fields in the second eigenvalue which cannot be straightforwardly, immediately explained by natural variability in the geomagnetic field.

Figure 10.15 focuses on eigenvalue variation during the three weeks around the earthquake. Besides the PC1 on day 272, there are few remarkable features in this plot. We do note, however, that days 263, 270, and 277 are the Sundays on this plot, during which the BART starts service around 3 hours later than on a weekday. The width in time of the blue, low amplitude regions between times of BART activity in the 20-40s bands are decidedly wider on Sundays than on other days.

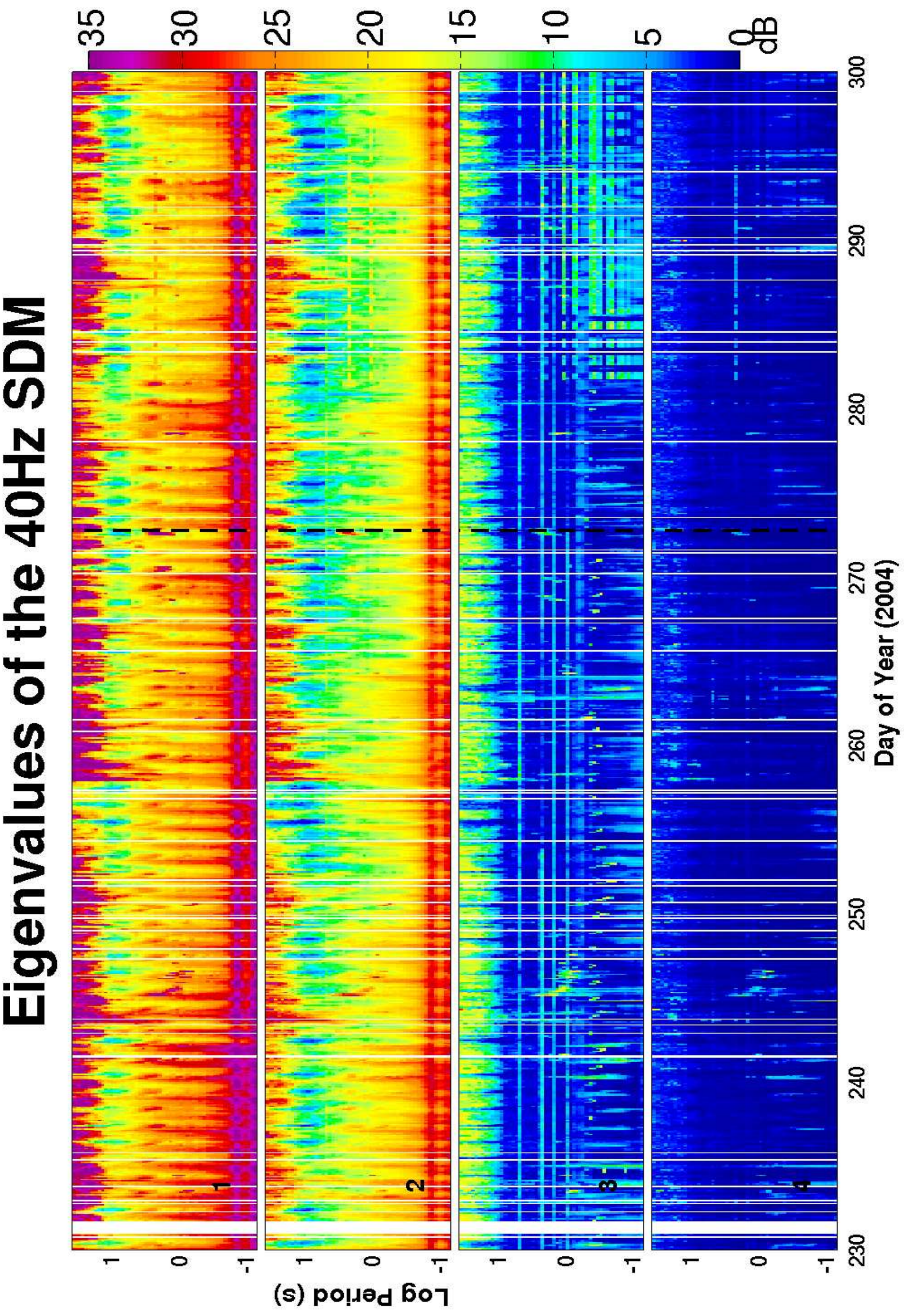


Figure 10.13: Dominant four eigenvalues of the SDM for the 75-day window surrounding the 2004 Parkfield earthquake.

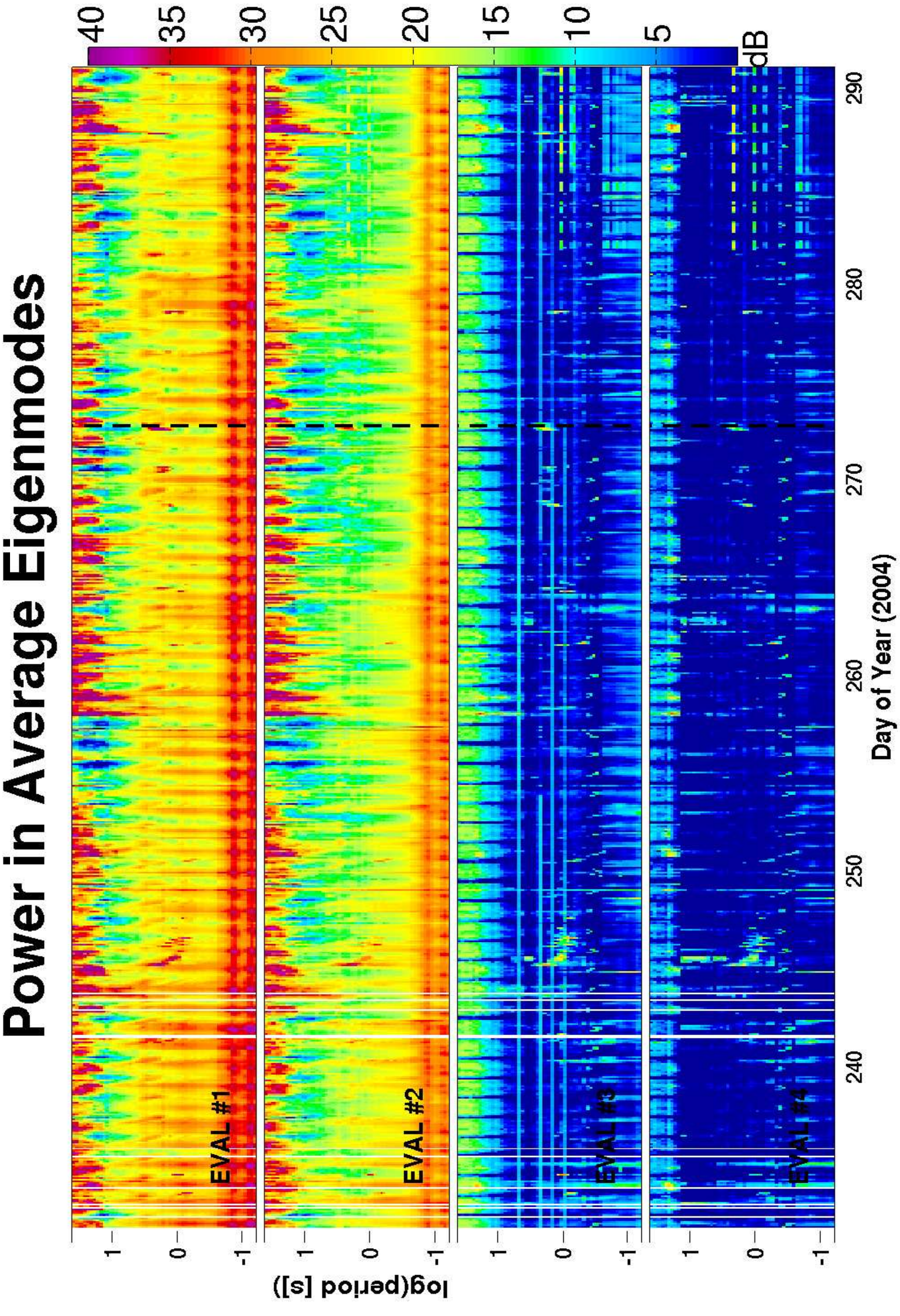


Figure 10.14: Power in the dominant four eigenmodes of the averaged SDM for the 75-day window surrounding the 2004 Parkfield earthquake.

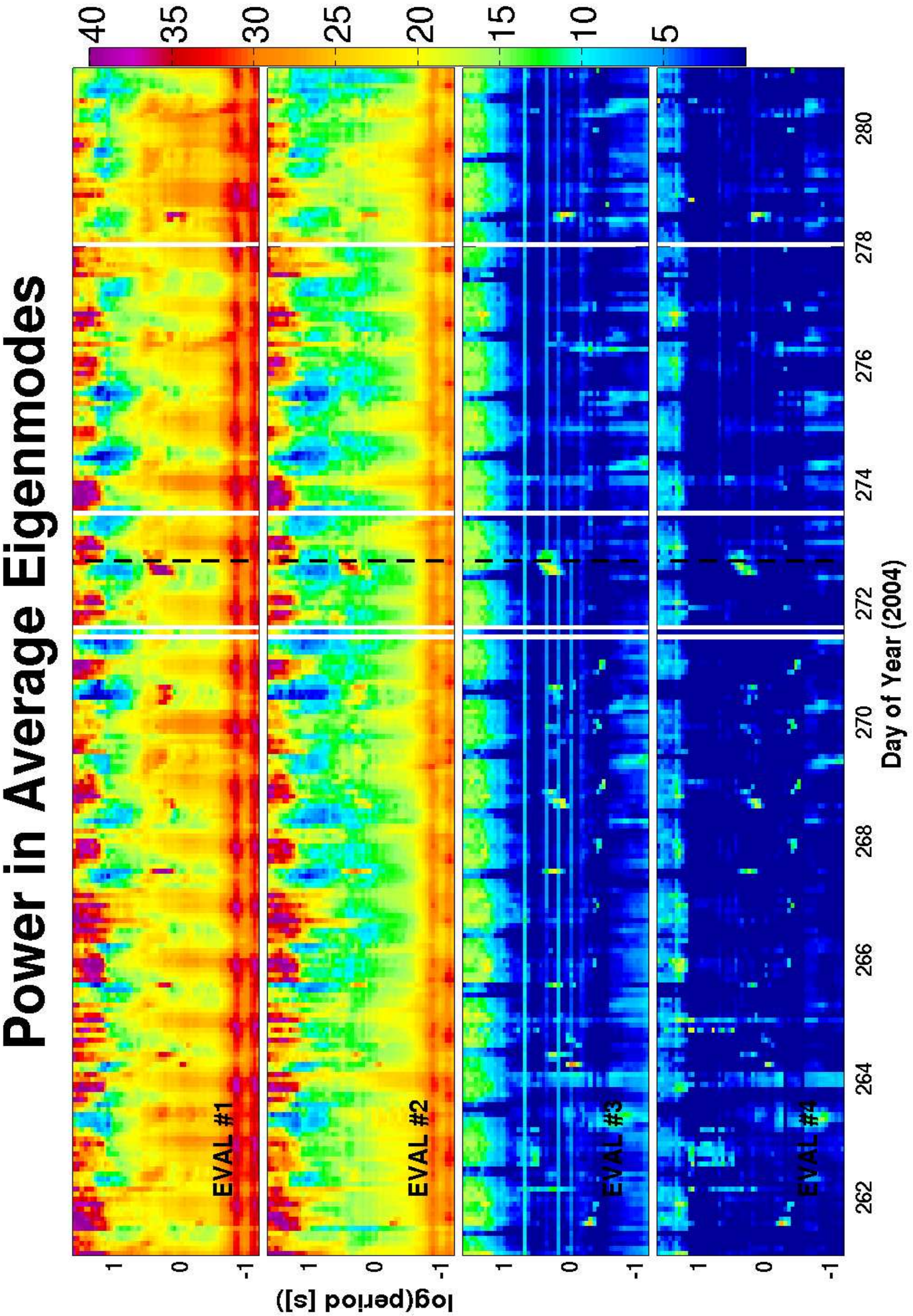


Figure 10.15: Power in the dominant four eigenmodes of the averaged SDM for the 3 week window surrounding the 2004 Parkfield earthquake.

10.5.1 Canonical Coherences

Canonical coherence plots are shown for Parkfield vs. Hollister channel groupings as well as for electric vs. magnetic channels groupings. Plots are shown for the 75 day window, and the 1 week window around the earthquake. A variety of the features mentioned in the previous plots are also seen here. Note that the PC1 activity is clearly present at both sites.

10.5.2 Residuals

Residuals for the 40Hz data are calculated in a slightly more sophisticated manner than they were for the 1Hz data. The transfer function which predicts the PKD data using observations at SAO relates the dominant two principal components at PKD with the dominant two principal components at SAO. This modification is introduced to discourage the predicted time series from migrating single site noise from SAO to PKD. In the case of the plots shown in Figure **??**, this method is probably not appropriate for removing the PC1 activity. Since the PC1 can be clearly seen to be at least 3-dimensional from the eigenvalue plots (Figure 10.15), a two dimensional transfer function cannot possibly yield a residual free of the PC1 signal. Future analysis would benefit from an adaptive filter which calculates the intersite transfer function frequently, and does so with three or more uncorrelated input channels.

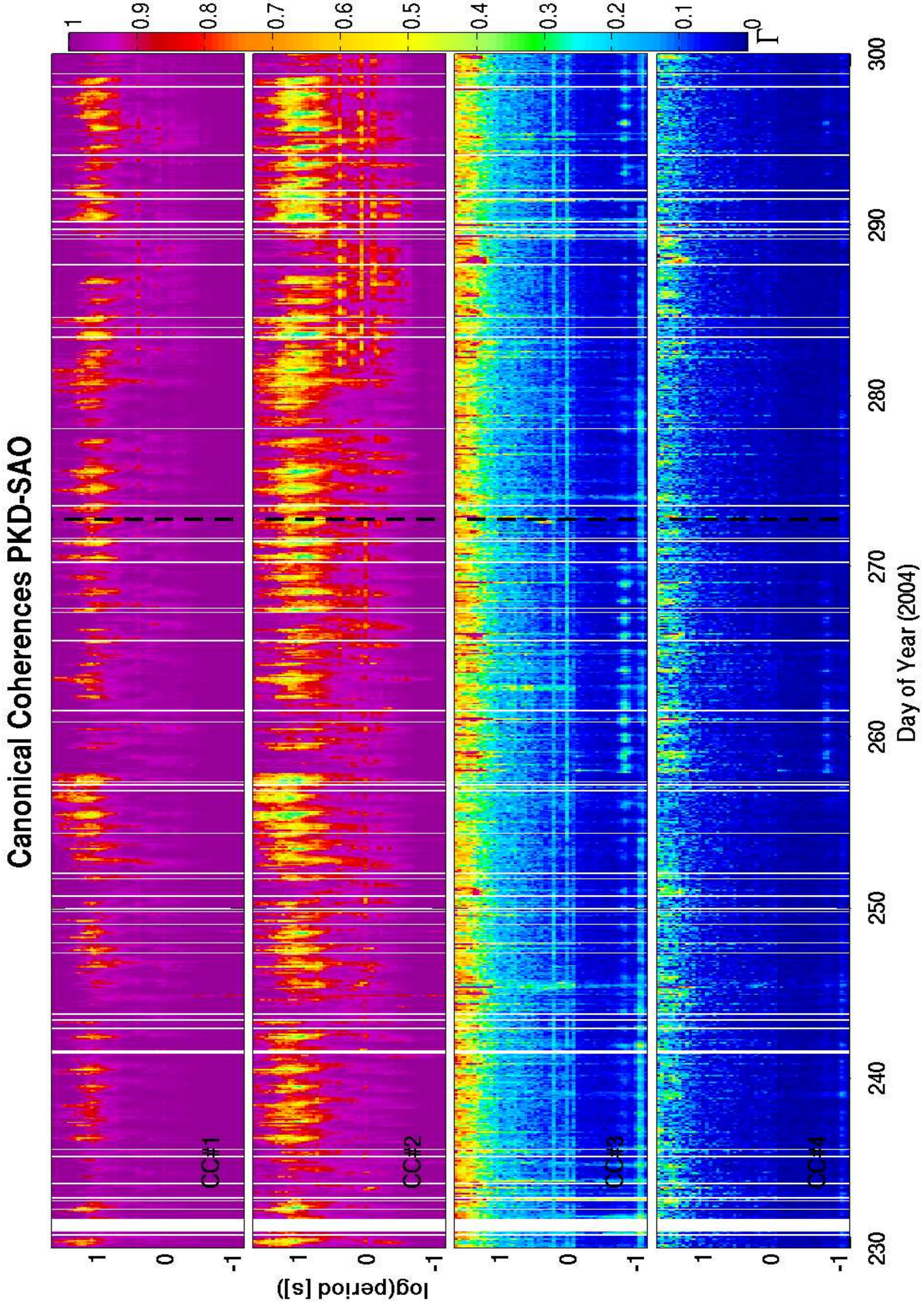


Figure 10.16: Intersite Canonical Coherences at 40Hz for the 75-day window surrounding the 2004 Parkfield earthquake.

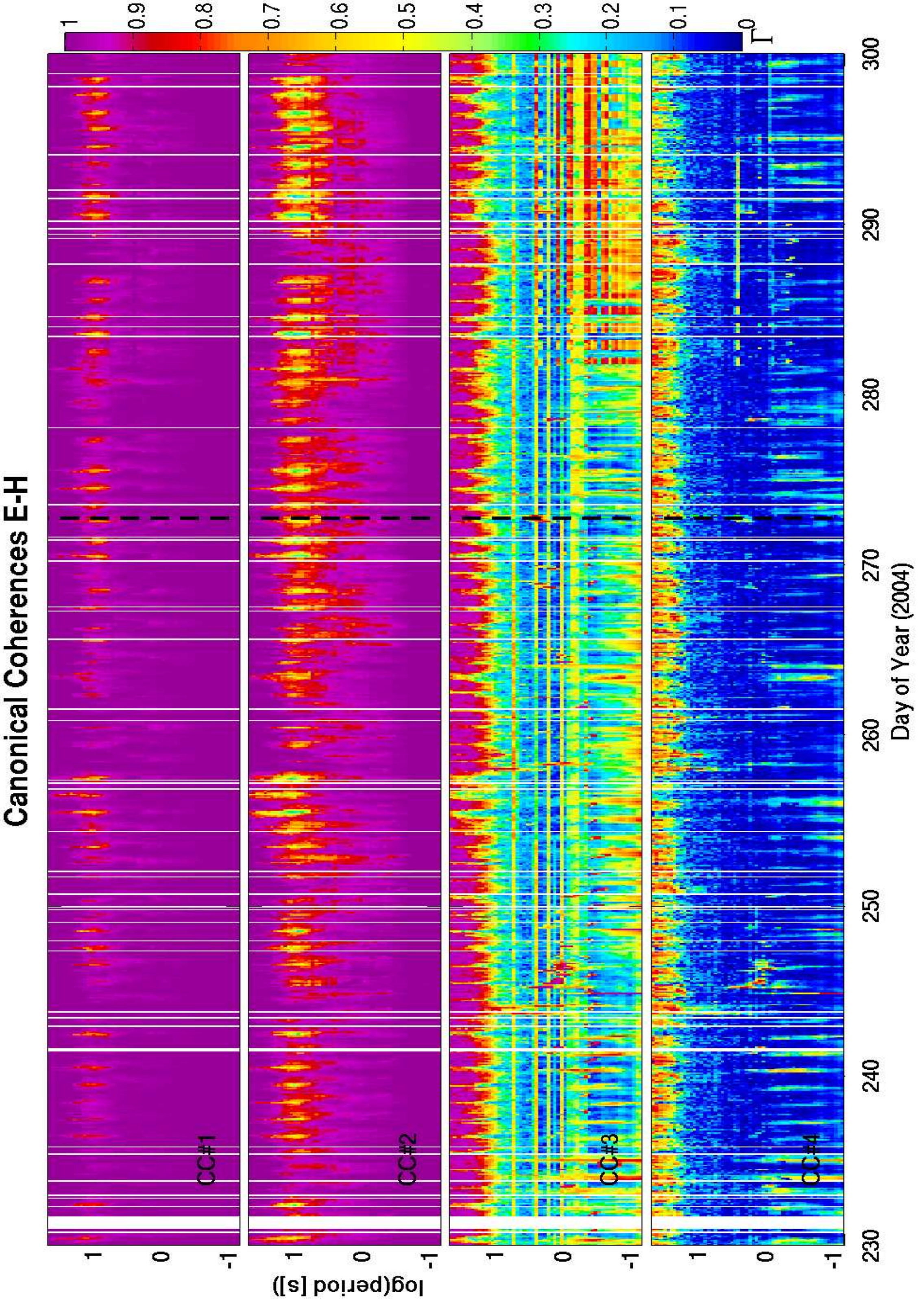


Figure 10.17: Electric/magnetic Canonical Coherences at 40Hz for the 75-day window surrounding the 2004 Parkfield earthquake.

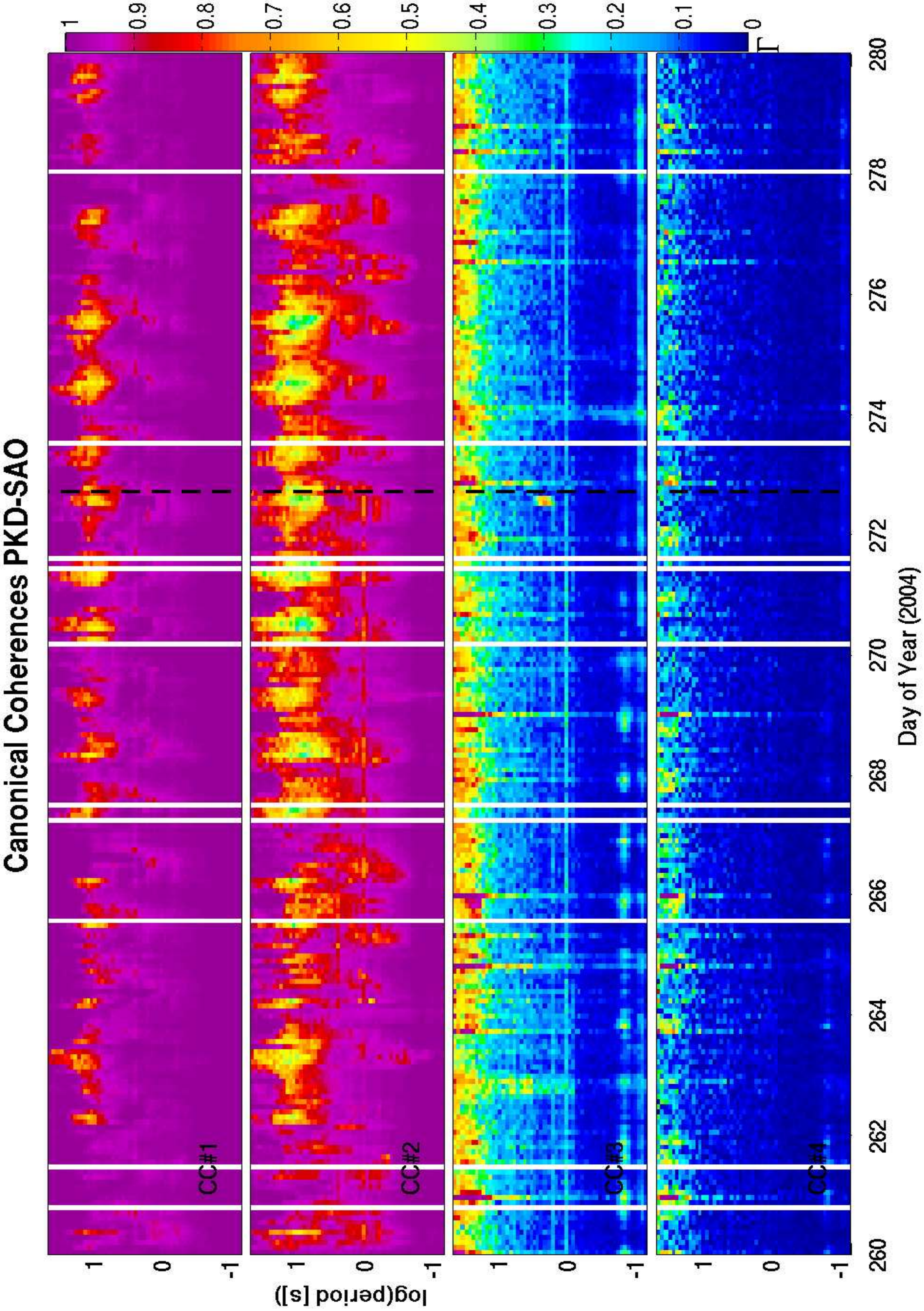


Figure 10.18: Intersite Canonical Coherences at 40Hz for the 3 week window surrounding the 2004 Parkfield earthquake.

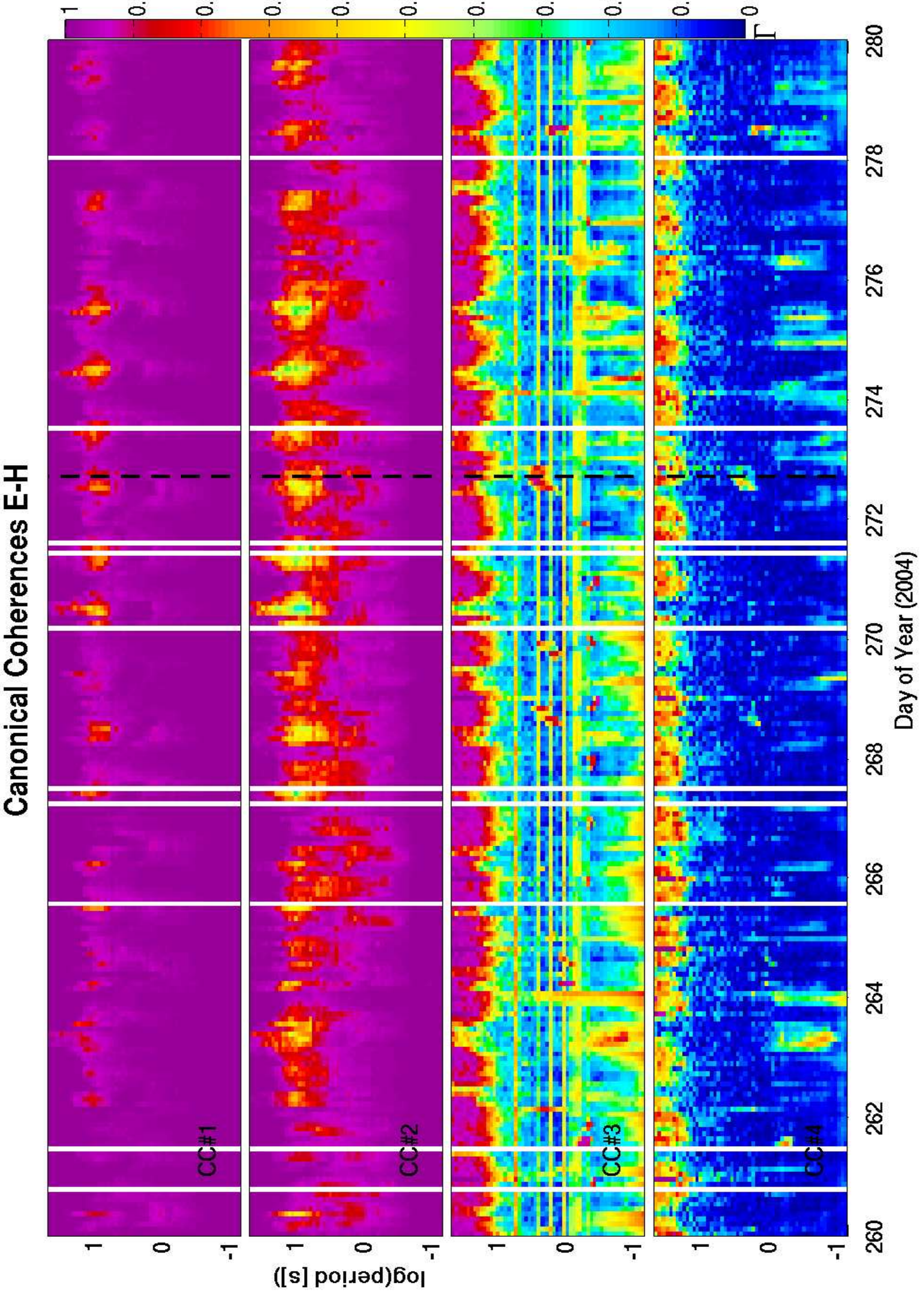


Figure 10.19: Electric/magnetic Canonical Coherences at 40Hz for the 3 week window surrounding the 2004 Parkfield earthquake.

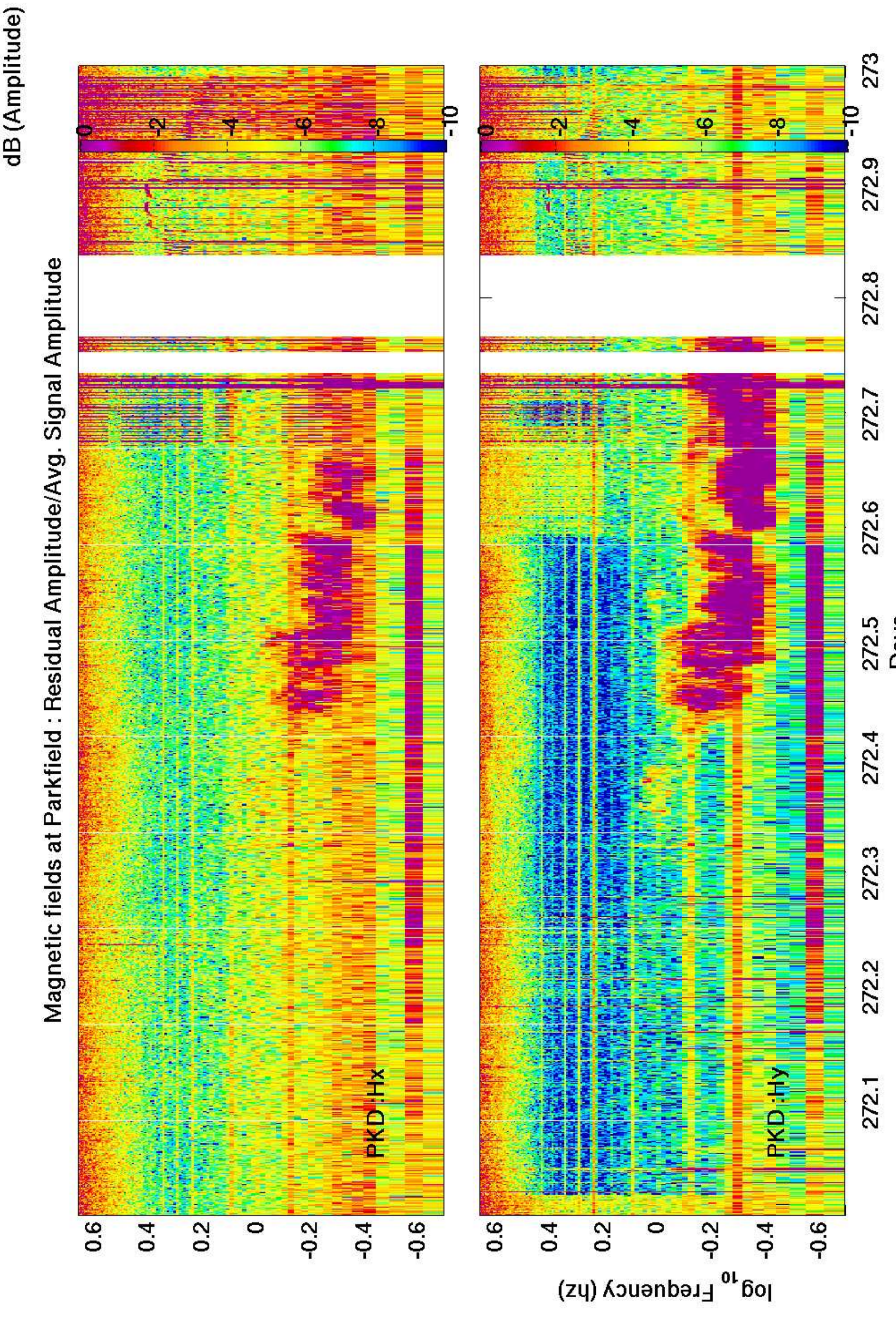


Figure 10.20: Residuals from Parkfield on day 272, 2004.

Chapter 11

Concluding Remarks

11.1 Conclusions

This study was primarily motivated by reports of anomalous electromagnetic activity preceding earthquakes. Data were analyzed from a pair of ultra low frequency electromagnetic observatories, which are instrumented like typical magnetotelluric acquisition sites, modified to monitor data for many years. One of the sites is located 20 km from the epicenter of the September 28, 2004 M6 Parkfield earthquake. The window of time treated (2002-2005) spans four years around the magnitude 6 earthquake. Before analysis, the data were carefully examined for pathological anomalies, such as spikes of non-physical amplitudes, gaps due to power outages or instrument malfunctions, etc. The corrupted data were subsequently either replaced or removed. In order to ensure that the data cleaning process did not remove the very anomalous signals which were later searched for, the temporal distribution of the pathological anomalies was examined to confirm that there no statistically significant variation occurred in the number of anomalies around the time of the earthquake. Once extreme outliers were treated, an examination of array fidelity was undertaken. Some common methods of signal to noise ratio calculation from magnetotelluric data processing literature are used as an index of confidence for data quality. Windows of time reflecting sufficiently high signal to noise ratios were plotted in terms of signal and noise amplitude spectra, and the behaviour of the ULF fields were shown over several years. From these first order field behavior plots it is clear that most of the recorded energy is coherent over the spatial extent of the array. The observed fields are the sum of natural magnetotelluric fields, cultural noise, instrument noise, and possibly other sources.

The ULF fields sampled at 1 Hz were found to be remarkably stable over the period of analysis. Nearly all significant variations in signal amplitude spectra were directly correlated with unusually active periods of global geomagnetic activity. A significant challenge in this study was accounting for DC offsets and gaps in observed data due to instrument swaps and site maintenance. Documentation of site visits and accurate recording of instrument calibrations and calibration changes are crucial to the monitoring effort and must be incorporated in any long term study which engages in ULF monitoring efforts. Long term time series of the Schumann resonance peak amplitude were used as an indicator of DC offsets. These plots showed that variations of several percent field strength were observed across maintenance dates. Such offsets constitute an unacceptable level of noise in a search for subtle variations in field properties. Fortunately, a 163 day window around the earthquake was found to be free of site visits, and associated DC offsets.

The possibility that a local electromagnetic signal near the epicenter may be masked by the coherent magnetotelluric noise was investigated. Various methods of transfer function estimation were employed to subtract the dominant modes of the MT field, in order to look deeper at the 'residual' fields. While these residual fields did show some interesting phenomena, no signals that can be said to be earthquake precursors were found. It is clear, however, that there are some significant signals in the recorded data which cannot be accounted for by simply subtracting the magnetotelluric fields. Most notably is the signature of the BART DC electric train system, which contaminates the data at periods between 10 and 40 seconds, as well as some PC3 activity.

The signals were decomposed into principal components in order to look at the behaviour of the linear combinations of array channels which are maximally uncorrelated. This is done with an eigenvalue decomposition of the array covariance matrix. The fine sensitivity of the array is evidenced by clear identification of a DC electric train which is 125km form the nearest site in the array.

The technique of Canonical Coherences was employed to distinguish anomalous fields which are spatially broad from anomalies which occur at a single site only, and furthermore to distinguish anomalies which are present in both the electric and magnetic fields form those which are present in only one field type. A signal identified by this technique was found to be present in both electric and magnetic fields, and was not present at both sites. The signal in the band around 4s period increased significantly in amplitude before the earthquake and decreased significantly afterward. Only by looking more than a year to either side of the local maxima centered on the earthquake does it become clear that the signal is also present and active when no significant seismicity is present at Parkfield. This demonstrates that long term monitoring is indispensable to any systematic (scientific) search for precursory signal.

Apparent resistivity estimates were generated daily at Parkfield. Most of the variation was observed to be seasonal and frequency independent, suggesting a seasonal distortion effect. Once the data were corrected for distortion, nearly all of the variability in the apparent resistivity was removed.

Results were calculated using data sampled at 1Hz, but a few selected plots of higher frequency data were shown for 75 days around the earthquake. These show a number of interesting signals, some of which are unexplained, but none of which can be associated with the Parkfield earthquake.

Bibliography

Bakun, W. H. and T. V. McEvilly, Recurrence models and Parkfield, California, earthquakes *J. Geophys. Res.*, **89**, 3051-3058, 1984.

Berdichevski, M.N. and V.I. Dmitriev, *Magnetotellurics in the context of the theory of ill-posed problems*, Society of Exploration Geophysicists, ISBN 1-56080-106-9, 2002.

Bernardi, A., Fraser-Smith, A.C. and O.G. Villard, Measurements of BART magnetic fileds with an automatic geomagnetic pulsation index generator, *IEEE Transaction on Electromagnetic Compatibility*, **31**, 1989.

Beard, M.W., Power spectra of geomagnetic fluctuations between 0.02 and 20 Hz, *Masters Thesis, Naval Postgraduate School, Monteray California*, 1981.

Brillinger, D.R., The canonical analysis of stationary time series, in *Multivariate Analysis - II*, ed. P.R. Krishnaiah, Academic Press, New York, 331-350, 1969.

Brillinger, D.R., *Time series: Data analysis and theory*, Holt Rinehart and Winston Inc., ISBN 0-03-076975-2, 1975.

Chave, A.D. and D.J. Thomson, Some comments on magnetotelluric response function estimation, *J. Geophys. Res.*, **94**, 14,215-14,225, 1989.

Christian, H.J., Blakeslee, R.J., Boccippio, D.J., Boeck, W.L., Buechler, D.E., Driscoll, K.T., Goodman, S.J., Hall, J.M., Koshak, W.J., Mach, D.M. and M.F. Stewart, Global frequency and distribution of lightning as observed from space by the Optical Transient Detector, *J. Geophys. Res.*, **108**, doi:10.1029/2002JD002347, 2003.

Corwin, R.F. and H.F. Morrison, Self-potential variations preceding earthquakes in central California, *Geophys. Res. Lett.*, **4**, 171-174, 1977.

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

Egbert, G.D. and J.R. Booker, Multivariate analysis of geomagnetic array data 1. The response space, *J. Geophys. Res.*, **94**, 14227-14247, 1989.

Egbert, G.D., Multivariate analysis of geomagnetic array data 2. Random source models, *J. Geophys. Res.*, **94**, 14249-14265, 1989.

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

Egbert, G.D., Eisel, M., Boyd, S. and H.F. Morrison, DC trains and Pc3s: Source effects in mid-latitude geomagnetic transfer functions, *Geophys. Res. Lett.*, **27**, 25-28, 2000.

Egbert G.D., Processing and interpretation of electromagnetic induction array data, *Surveys in Geophysics*, **23**, 207-249, 2002.

Egbert, G.D., Final technical report to USGS, Award Number 03HQGR0060, 2004.

Eisel, M. and G.D. Egbert, On the stability of magnetotelluric transfer function estimates and the reliability of their variances, *Geophys. J. Int.*, **144**, 65-82, 2002.

Fenoglio, M.A., Fraser-Smith, A.C., Beroza, G.C. and M.J.S. Johnston, Comparison of ultra-low frequency electromagnetic signals with aftershock activity during the 1989 Loma Prieta earthquake sequence, *Bull. Seism. Soc. Am.*, **83**, 347-357, 1993.

Fitterman, D. and T. Madden, Resistivity observations during creep events at Melendy Ranch, California, *J. Geophys. Res.*, **82**, 5401-5408, 1977.

Fraser-Smith, A.C., Bernardi, A., McGill, P.R., Ladd, M.E., Helliwell, R.A. and O.G. Villard, Low-frequency magnetic measurements near the epicenter of the M_s 7.1 Loma Prieta earthquake *Geophys. Res. Lett.*, **17**, 1465-1468, 1990.

Fraser-Smith, A.C., McGill P.R., Helliwell R.A., and O.G. Villard Jr., Ultra-low frequency magnetic field measurements in southern California during the Northridge earthquake of 17 January 1994, *Geophys. Res. Lett.*, **21**, 2195-2198, 1994.

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

Geller*, R.J., Debate on evaluation of the VAN method: Editor's Introduction, *Geophys. Res. Lett.*, **23**, 1291-1293, 1996. * This entire issue is devoted to the VAN method controversy

Geller, R.J., Braginski, A. and W. Campbell, Earthquake precursors or background noise?, *IEEE Spectrum*, April 2006.

Hamada, K., Statistical evaluation of the SES predictions issued in Greece: Alarm and success rates, *Tectonophysics*, **224**, 203-210, 1993.

Hardle, W. and L. Simar, *Applied multivariate statistical analysis*, Springer-Verlag, Berlin, ISBN 978-3-540-72243-4, 1990.

Harris, F.J., On the use of windows for harmonic analysis with the discrete Fourier transform, *Proceedings of the IEEE*, **66**, 51-83, 1978.

Hayakawa, M., Molchanov, O., Tronin, A., Hobara, Y. and T. Kodama T, NASDA's earthquake remote sensing frontier research: Seismo-electromagnetic phenomena in the lithosphere, atmosphere, and ionosphere, *Final Report, The University of Electro-Communications*, Chofu, Tokyo, Japan, 2001.

Heckman, S.J., Williams, E. and B. Boldi, Total global lightning inferred from Schumann resonance measurements, *J. Geophys. Res.*, **103**, 31,775-31,779, 1998.

Hotelling, H., The most predictable criterion, *Journal of Educational Psychology*, **26**, 139-142, 1935.

Huber, P.J., *Robust statistics*, Wiley Series in Probability and Mathematical Statistics, New York, 1981.

Hyvarinen, A., Karhunen, J. and E. Oja, *Independent componenets analysis*, Wiley Interscience, ISBN 0-471-40540-X, 2001.

Jiracek, G.R., Near-surface and topographic distortions in electromagnetic induction, *Surveys in Geophys.*, **11**, 163-203, 1990.

Johnston, M.J.S., Review of magnetic and electric field effects near active faults and volcanoes in the U.S.A., *Phys. Earth planet. Inter.*, **57**, 47-63, 1989.

Johnston, M.J.S., Review of electric and magnetic fields accompanying seismic and volcanic activity, *Surveys in Geophys.*, **18**, 441-476, 1997.

Kagan, Y.Y., VAN earthquake predictions: An attempt at statistical evaluation, *Geophys. Res. Lett.*, **23**, 1315-1318, 1996.

Kappler, K.N., Egbert, G.D. and H.F. Morrison, Long term monitoring of EM signals near Parkfield CA, *AGU Fall Meeting*, 2005.

Kappler, K.N, Cuevas, N.H. and J.W. Rector, Response of induction coil magnetometers to perturbations in orientation, *SEG 76th Annual International Meeting*, 899-903, 2006.

Kappler, K.N. and H.F. Morrison, Observation and analysis of vertical electric fields in the Earth, *Berkeley Seismological Laboratory Annual Report*, 2005-06, Karakelian, D., Beroza G.C., Klemperer S.L. and A.C. Fraser-Smith, Analysis of ultralow-frequency electromagnetic field measurements associated with the 1999 M 7.1 Hector Mine, California, Earthquake Sequence, *Bull. Seism. Soc. Am.*, **92**, 1513-1524, 2002.

Kopytenko, Y.A., Ismaguilov V.S., Molchanov O.A., Kopytenko E.A., Voronov P.M., Hattori K., Hayakawa M. and D.B. Zaitzev, Investigation of ULF magnetic disturbances in Japan during seismic active period, *J. Atmos. Elec.*, **22**, 207-215, 2002.

Lyubushin Jr., A.A., Analysis of canonical coherences in the problems of geophysical monitoring, *Izv. Phys. Solid Earth*, **34**, 52-58, 1998.

Ma, Q., Feng, Z., Song, Z. and W. Zhao, Study on the variation characteristics of the geoelectirc field preceding earthquakes, *Acta Seismologica Sinica*, **17**, 334-343, 2004.

Mazella, A. and H.F. Morrison, Electrical resistivity variations associated with earthquakes on the San Andreas Fault, *Science*, **185**, 855-857, 1974.

Molchanov, O.A., Y.A. Kopytenko, P.M. Voronov, E.A. Kopytenko, T.G. Matiashvili, A.C. Fraser-Smith and A. Bernardi, Results of ULF magnetic field measurements near the epicenters of the Spitak (Ms = 6.9) and Loma Prieta (Ms = 7.1) earthquakes: Comparative analysis, *Geophys. Res. Lett.*, **19**, 1495-1498, 1992.

Morrison, H.F., Corwin, R.F. and R. Fernandez, Earth resistivity, self potential variations and earthquakes: A negative result for M = 4.0, *Geophys. Res. Lett.*, **6**, 139-142, 1979.

Mulgaria, F. and P. Gasperini, Evaluating the statistical validity beyond chance of VAN earthquake precursors, *Geophys. J. Int.*, **111**, 32-44, 1992.

Nadeau, R.M. and T.V. McEvilly, Seimological studies at Parkfield V: Characteristic microearthquake sequences as fault-zone drilling targets, *Bull. Seism. Soc. Am.*, **87**, 1463-1472, 1997.

Park, S.K., Johnston M.J.S, Madden T.R., Morgan F.D., and H.F. Morrison, Electromagnetic precursors to earthquakes in the ULF band: A review of observations and mechanisms, *Rev. Geophys.*, **31**, 117-132, 1993.

Park, S.K., Dalrymple, W. and J.C. Larsen, The 2004 Parkfield earthquake: Test of the electromagnetic precursor hypothesis, *J. Geophys. Res.*, **112**, doi:10.1029/2005JB004196, 2007.

Perrier, F.E., Petiau, G., Clerc, G., Bogorodsky, V., Erkul, E, Jouniaux, L., Lesmes, D., Macnae, J., Meunier, J.M., Morgan, D., Nascimento, D., Oettinger, G., Schwarz, G., Toh, H., Valiant, M., Vozoff, K. and O. Yazici-Cakin, A one-year systematic study of electrodes for long period measurements of the electric field in geophysical environments, *J. Geomag. Geoelectr.*, **49**, 1677-1696, 1997.

Pham, V.N., Boyer, D., Le Mouel, J.L., Chouliars, G. and G.N. Stavrakakis, Electromagnetic signals generated in the solid earth by digital transmission of radio-waves as a plausible source for some so-called 'seismic electric signals', *Phys. Earth planet. Inter.*, **114**, 141-163, 1999.

Press, W. H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T., *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press, ISBN 0-521-38330-7, 1990.

Pulinets, S. and K. Boyarchuk, *Ionospheric Precursors of Earthquakes*, Springer-Verlag, ISBN 3-540-20839-9, 2004.

Simpson, F. and K. Bahr, *Practical magnetotellurics*, Cambridge University Press, ISBN 0-521-81727-7, 2005.

Smith, J.T., Understanding telluric distortion matrices, *Geophys. J. Int.*, **122**, 219-226, 1995.

Tullis, T.E., Deep slip rates on the San Andreas Fault, *Science*, **285**, 671-672, 1999.

Unsworth, M., Egbert, G. and J. Booker, High-resolution electromagnetic imaging of the San Andreas fault in central California, *J. Geophys. Res.*, **104**, 1131-1150, 1999.

Uyeda, S., Nagao T., Orihara Y., Yamaguchi T. and I. Takahashi, Geoelectric potential changes: Possible precursors to earthquakes in Japan, *Proceedings of the National Academy of Sciences*, **97**, 4561-4566, 2000.

Varotsos, P. and K. Alexopoulos, Physical properties of the variations of the electric field of the Earth preceding earthquakes, I, *Tectonophysics*, **110**, 73-98, 1984.

Varotsos, P. and M. Lazaridou, Latest aspects of earthquake prediction in Greece based on seismoelectric signals, *Tectonophysics*, **188**, 321-347, 1991.

Varotsos, P., Alexopolous, K. and M. Lazaridou, Latest aspects of earthquake prediction in Greece based on seismoelectric signals II, *Tectonophysics*, **224**, 1-37, 1993.

Varotsos, P., Eftaxias, K., Vallianatos, F. and M. Lazaridou, Basic principles for evaluating an earthquake prediction method, *Geophys. Res. Lett.*, **23**, 1295-1298, 1996.

Wyss, M., Inaccuracies in seismicity and magnitude data used by Varotsos and coworkers, *Geophys. Res. Lett.*, **23**, 1299-1302, 1996.

Wyss, M. and A. Allman, Probability of chance correlations of earthquakes with predictions in areas of heterogeneous seismicity rate: the VAN case, *Geophys. Res. Lett.*, **23**, 1307-1310, 1996.

Zhao, Y. and F. Qian, F., Geoelectric precursors to earthquakes in China, *Proceedings of the Lake Arrowhead Workshop*, 1992.

Appendix A

A short tutorial on eigenvectors and eigenvalues of the SDM.

The relationship between the vector valued time series of Fourier coefficients at some frequency ω, $\mathbf{X}_\omega$(t) (in this case an 8xT complex array) and the eigenvectors and eigenvalues of the scaled SDM $\mathbf{S}_\omega \approx (1/T)\mathbf{X}\mathbf{X}^*$ is not intuitively obvious. This appendix is intended to build some intuition with these quantities using two main examples. The first example emphasizes the interpretation of a square NxN matrix as a mapping from one N-dimensional vector space to another N-dimensional vector space. The other example is a graphic description of how the eigenvectors and eigenvalues of a covariance matrix relate to the distribution of points which generated that covariance matrix.

A.1 Examples of arrays as operators

There are several contexts in which a matrix can be viewed. In this discussion an mxn matrix A is thought of a transformation., or a mapping from V1 to V2, where V1 has dimension n, and V2 has dimension m. Formally, we say A $\in$ $M_{m,n}(\mathbb{F})$, where $\mathbb{F}$ is some field, usually the real numbers $\mathbb{R}$ or the complex numbers $\mathbb{C}$. For instance, the SDMs of Chapter 4 can be viewed as a mapping. In that case m=n=8, and the field is $\mathbb{C}$. Square arrays like these can be thought of as automorphisms, which is to say they their domain and range are the same spaces, and the mapping is bijective, i.e., it is one to one, and onto. All this means is that every point in $\mathbb{C}^8$ is the image of some point in $\mathbb{C}^8$ under the transformation **S**, according to the standard rules of matrix multiplication.

Visualizing the behaviour of $\mathbb{C}^8$ under **S** is not easy. However, we can build intuition by looking at lower dimension subspaces which we can visualize. For the sake of clarity, let us consider a 2x2 real matrix **A**, which in our formal notation maps the real plane onto the real plane, or A:$\mathbb{R}^2$->$\mathbb{R}^2$ where A $\in$ $M_{2,2}(\mathbb{R})$.

Figure $A.1$ shows how the 2x2 operator A affects the unit square. In this figure, the points marked by x symbols transform under A to the points marked by the circle symbols (their images under **A**).. The lines linking the x-symbols also transform to the corresponding coloured lines linking the o-symbols. Moreover, a second square of the same size adjoining the one shown along some face of a colour *c* would transform to a parallelogram with the same orientation as the one shown, which would share with the shown parallelogram the face of colour c. The area of the unit square is by definition 1, and the determinant of the operator A corresponds to the area of the transformed square. Thus, in Figure A.1, the parallelogram has area 1.55. Figure A.1 shows a few properties of the covariance matrix A that are not proved here, but that can be found in any elementary text on linear algebra.

The eigenvalues, (solutions of A's characteristic polynomial) are real;

The eigenvectors are orthogonal.

The transformation A maps straight lines to straight lines, and moreover maps parallel lines to parallel lines.

A point x in the domain of **A** has image **A**x. By looking at x, denoted by a dot on the plane, and looking at the elements of the matrix A, it is not obvious where the point Ax will lie. The eigenvectors of the matrix A (indicated by magenta lines) however, form an orthogonal basis for $\mathbb{R}^2$. Since an eigenvector **e** of **A** has the property that $\mathbf{Ae}=\lambda\mathbf{e}$ for some λ, it seems that a more canonical representation of the data would be to use the coordinate system shown by the magenta lines. This can be done by rotating the page in front of you until the magenta lines lie in the horizontal and vertical directions. Then, a point x can be specified by its position relative to these axes (you may want to extend the lines out with a marker to make them the same size) and $\mathrm{Ax}=\mathrm{A}[\mathrm{x}_1, \mathrm{x}_2]=[\lambda_1 x_1, \lambda_2 x_2]$ is simply determined by multiplying the first element of x by the eigenvalue of the the eigenvector which is horizontal, and the second element of x by the eigenvalue corresponding to the vertical eigenvector of A. Of course, performing this change of coordinate systems to a collection of data is accomplished by multiplying the data vectors by a rotation matrix. This matrix U is found by decomposing A=UDU* where U is unitary, and D is the matrix of eigenvalues. The canonical property of the eigenbasis (shown in magenta in Figures A.2 and A.3) becomes more clear. Also, one sees that the condition numbers of the matrix examples are directly related to the eccentricity of the parallelogram images of the squares. This shows at least that eigenvectors and eigenvalues (eigen from German for inherent) can be used to gain insight into the dominant direction in which the Matrix A is changing the underlying vector space.

A.2 The eigenvectors of a simple distribution

A spectral density matrix (SDM) is a form of covariance matrix and has several useful properties, a few of which are listed here: 1. It is square, i.e. m=n

2. The Eigenvalues are all real and positive semidefinite

3. The eigenvectors are orthogonal

Consider a two dimensional distribution of observations. A sample of 1500 observations of a 2D variable is plotted in Figure A.3. By taking each observation $\mathbf{x}_i$=[$\mathrm{x}_{1,i}x_{2,i}$] and

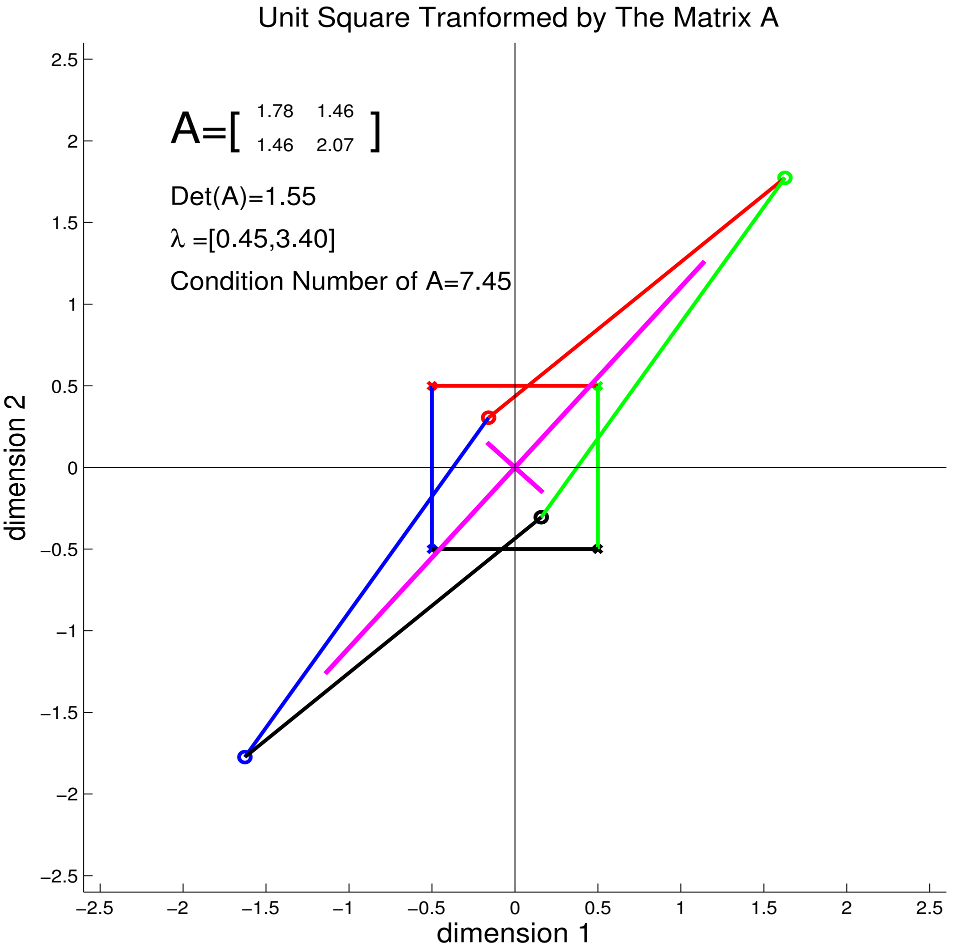


Figure A.1: The effect of the 2x2 operator A on the unit square

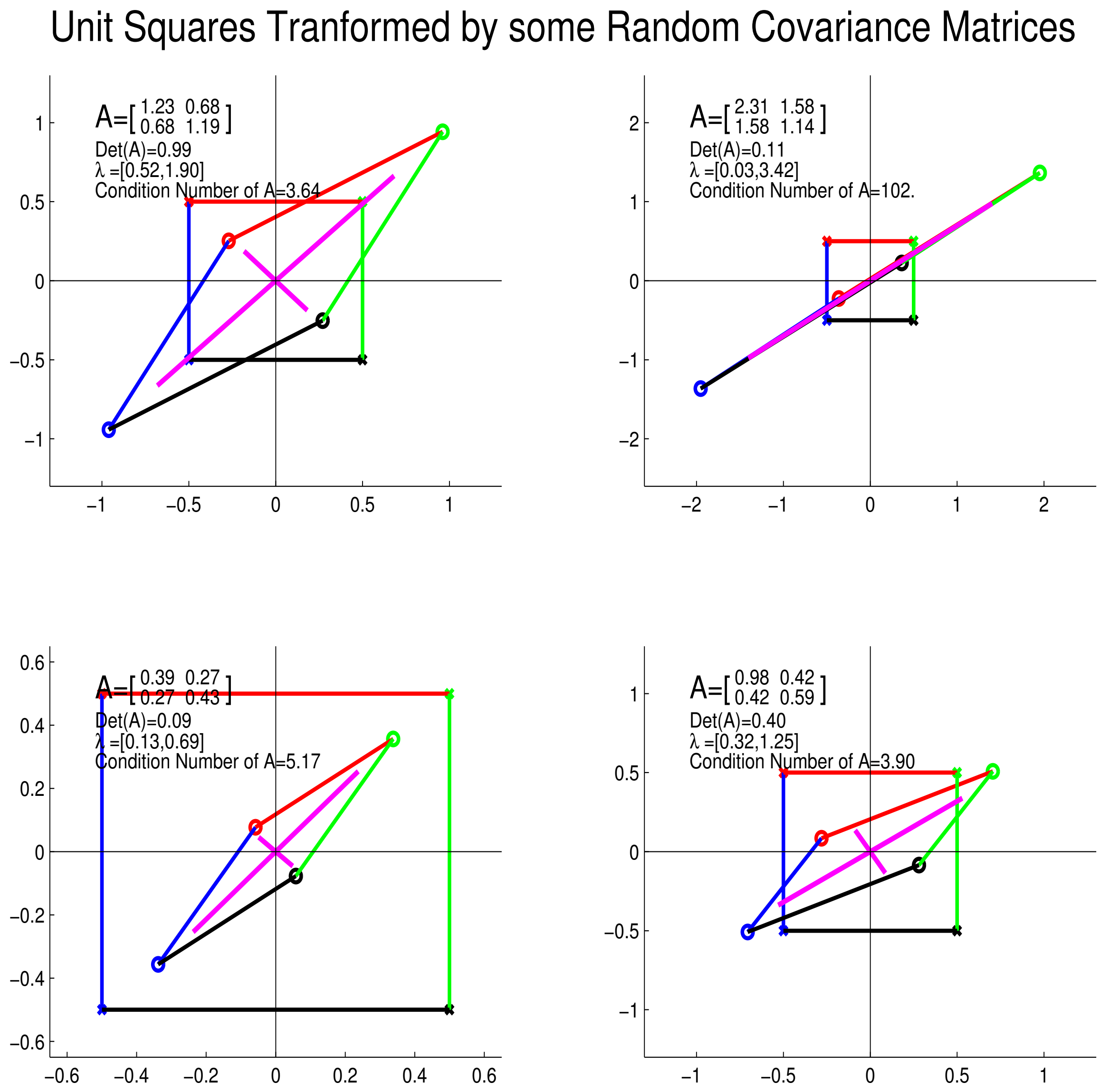


Figure A.2: The effect of some randomly generated covariance matrices on the unit square

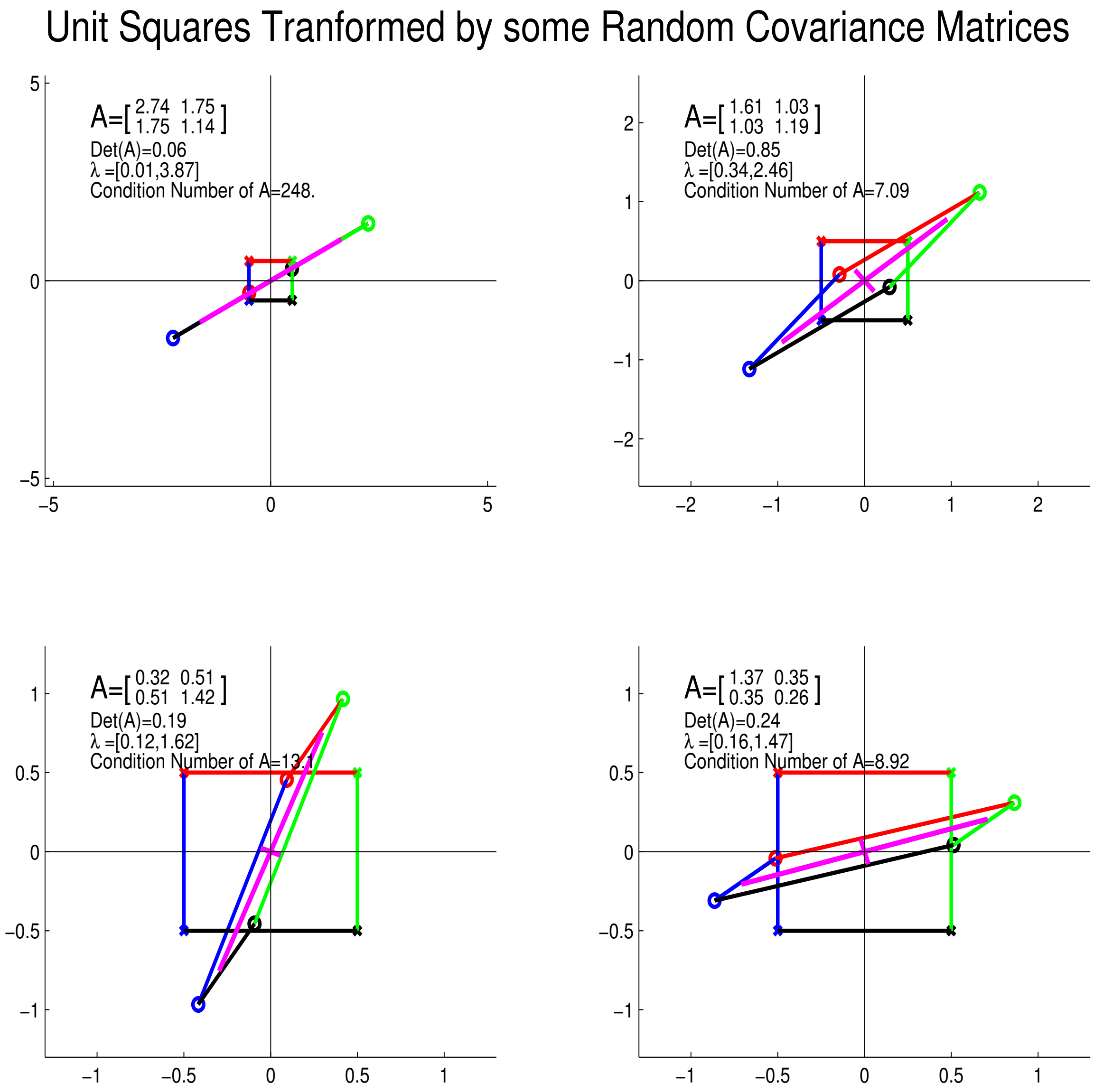


Figure A.3: The effect of some randomly generated covariance matrices on the unit square calculating its outer product matrix $\mathbf{xx}^*$, then averaging over all of these matrices, we obtain an SDM. When the eigenvectors of the SDM are plotted (shown by black lines), one can glean that the eigenvectors of an SDM carry information about the directions in which the data are most broadly distributed, or tightly clustered.

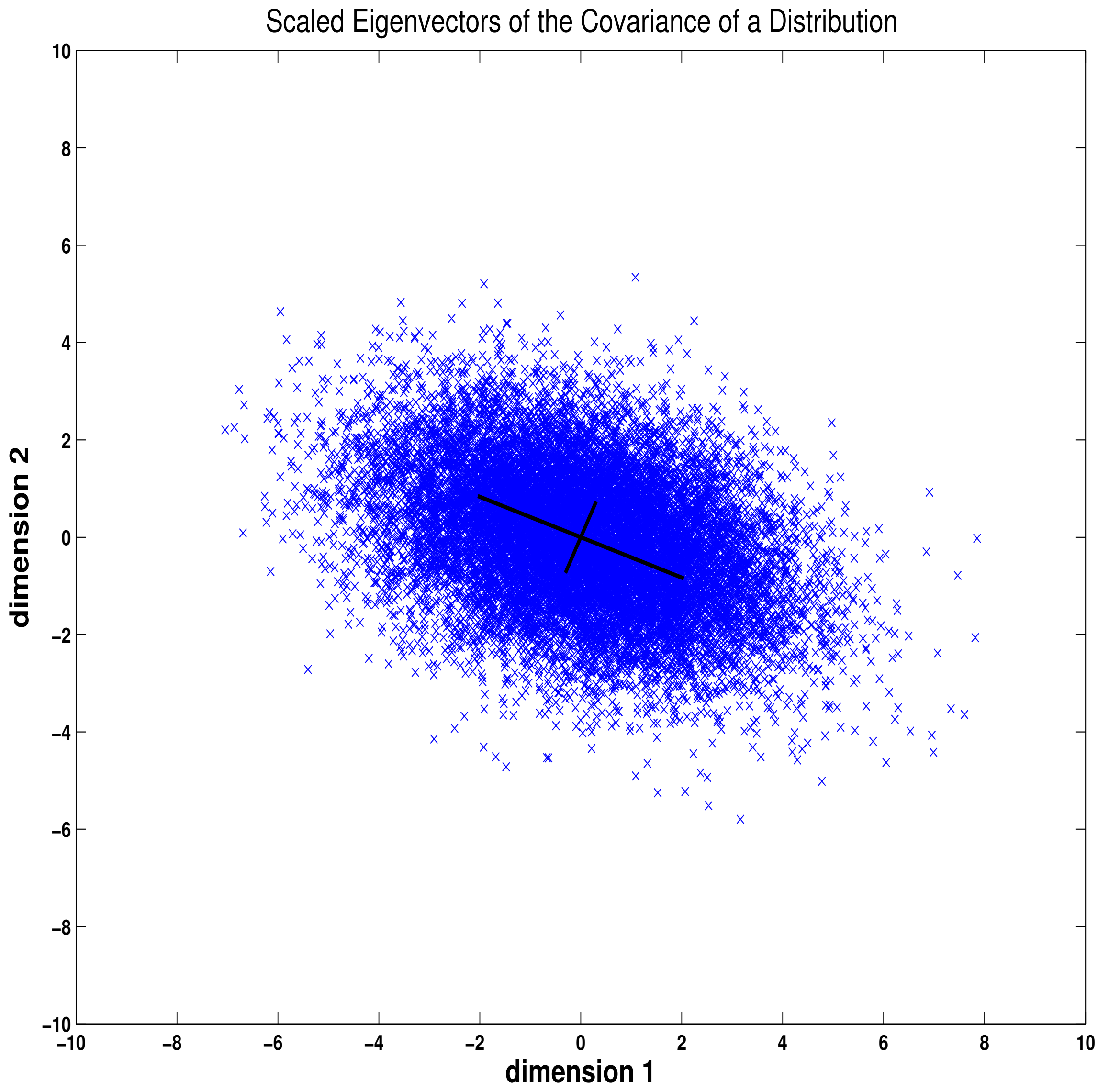


Figure A.4: The eigenvectors for the SDM of the distribution given by the blue dots.

Appendix B

Recommendations for future work

Following are some point form observations which relate to data in the PKD-SAO dataset.

Vertical Magnetic field measurements were available at both sites during much of the time window considered. The vertical coil at SAO was damaged during almost all of 2004, including during the 163 day window where much of the data processing shown in this document was focused. Lacking a reference vertical coil, we chose to omit the vertical measurements from this analysis. Analysis of the vertical field data may show other phenomena not observed in the horizontal magnetometers, and also could be used to calculate ‘Tipper’ transfer functions [*Simpson and Bahr* 2005]. We also had a vertical electrode operating at Parkfield during the second half of 2004. Again, due to the lack the of a duplicate measurement, the data from this channel were omitted from this study. A brief overview of the vertical electric field measurements can be found in *Kappler and Morrison* (2006). Both the electric and magnetic measurements could be used to identify the polarities of transient events, particularly the anomalous field at PKD at period $\approx$4s. A study of the polarizations of certain observed signals would be helpful in understanding their nature. It seems worthwhile to ask: is the polarization of the anomalous field always the same, since quanttifying the stability of any polarization might indicate whether the source is moving in space or not. The vertical field measurements already mentioned in this section could prove valuable in such a polarization study.

In regards to signal processing, there is room for an exploratory data analysis project using the relatively recent technique of Independent Components Analysis [*Hyvarnen et. al.* 2001]. This technique is typically applied to datasets after treating them with PCA as in chapter 7. The method seeks to decouple a multivariate time series into linear combinations that are maximally statistically independent. It would be interesting to compare this technique against PCA in terms of its ability to extract particular noise source signatures, for example the BART signal. Additionally, a signal processing method, termed redundancy analysis works similar to CC analysis, but extracts the maximally correlated linear combinations of array channels under the constraint that each dimension must explain the greatest portion of the variance of the data in the channel groupings. Further information of the application of redundancy analysis can be found in [*Hardle and Simar* 2007].

A more in-depth study of the modes of the MT field could also be carried out. If indeed it is true that the first two modes of the SDM constitute almost all of the natural field, it should then follow that projecting the data onto the the 2 dimensional space spanned by the first two dominant modes as defined in Chapter 7 should yield apparent resistivity estimates in line with those obtained using the raw data. The precise relationship between the modes identified in the CC method and those of the PC method could be undertaken. Given the success of the technique of residual calculation by using a transfer function on only to two dominant modes as used in chapter 10, one is encouraged to attempt a similar residual calculation method based on the intersite canonical variates.

The co-seismic signals plotted in Figure 7.3 should be analyzed. Specifically, the magnetic coseismic signals are very likely caused by motion of the coils in the Earth's DC magnetic field *Kappler* (2006), though no detailed model this supported by real data has been published. More interesting is the source of the electric field coseismic signals. These cannot be explained, to first order, by the motion of the wire in the Earth's DC magnetic field. Instead, the source of these signals could be very local electrode noise. There may also be streaming potential signals of the kind discussed by [*Nourbehecht*, 1963] and [*Onsager*, 1931]. A study of coseismic electrode effects is necessary in order to pursue field research in electrical monitoring during hydro-fracture [*Cuevas*)(personal communication)].

It is absolutely essential in any future continuation of this monitoring, or proposed augmentation of the array, that sites be maintained regularly, and that detailed logs be kept which document changes in site configuration and instrumentation. It is clear that the health of the system needs to be monitored in real time, by qualified and available technical support staff so that these gaps can be reduced, should another long term monitoring project be undertaken. Without these sorts of records we are left to speculate as to the causes of the signals recorded in, for example, Figures 4.25-4.33.