An Automated Local/Regional Seismic Event Detection and Location System Using Waveform Correlation

Submitted in Partial Fulfillment of

the Requirements for the Degree of

Doctor of Philosophy

Mitchell M. Withers

New Mexico Tech

Socorro, New Mexico

April, 1997

Abstract

New Mexico Tech operates a short-period digital seismic network consisting of 9 vertical-component seismometers deployed in an approximate 60 km- aperture centered on the central Rio Grande rift near Socorro, New Mexico. Historically, the principal mission of this network was to monitor local seismic activity associated with an inflating magma body residing in the mid-crust at a depth of approximately 19 km, and to monitor larger threshold ($m_d > 1.3$) seismic activity for the state of New Mexico. We report on the development of an automated Local Waveform Correlation Event Detection System (*LWCEDS*) and its application to this network. *LWCEDS* is an adaptation of global waveform correlation system *WCEDS* which is a matched filtering algorithm for global Comprehensive Test Ban Treaty (CTBT) monitoring applications developed at Sandia National Laboratories and New Mexico Tech. Although the current CTBT monitoring system is based on teleseismic phase detection, effort is being placed on research to "highlight" specific areas of the globe for which local and regional seismic networks could be employed. An automated regional location system will also assist state and local emergency management officials by providing magnitude and hypocenter information within a few minutes of the occurrence of a hazardous earthquake. The interests of the general scientific community, in an era of reduced research budgets, can be served by providing automatic bulletins which can considerably reduce the time investment required by seismic analysts.

With the *LWCEDS* algorithm, processed waveforms are correlated with theoretical travel-time envelopes and a grid search is performed to identify the points in space and time which yield the highest correlations. High correlation indicates that an event has occurred and that a good approximation to the correct origin time and hypocenter has been determined; explicit phase identification is not required. To avoid the large computational expense of a complete correlation for each grid point, we reformulate the problem into a single matrix multiplication and matrix assessment for each time step. *LWCEDS* has been successfully tested on a suite of local (epicenters within the $\sim$ aperture of the network) and regional seismic events (epicenters outside the network and within $\sim$ 510 km) selected to span the range of expected event quality. Preliminary results show epicentral errors less than 3 km for local events and less than 50 km for regional events within New Mexico or Eastern Arizona. Similar results were obtained during an online experiment conducted to generate daily bulletins for the period from December 3, 1996 through January 7, 1997. Data from 156 triggers were processed and stored including 33 teleseisms, 102 regional events (includes explosions), and 21 local earthquakes. Results from this expanded test set are encouraging but reveal the need for a method to mask or flag various electronic and telemetry spikes.

Acknowledgments

Special thanks go to my advisors Rick Aster and Chris Young for outstanding scientific and professional guidance. Fellow PhD students Bob Balch and Kuo-wan Lin provided invaluable advice regarding the seismicity of Socorro and New Mexico and great Bar-B-Q's and Bocce. Judy Beiriger, Julian Trujillo, and Mark Harris at Sandia National Laboratories provided excellent programming advice and helped prevent programming in C from causing a nervous breakdown. Professors John Schlue and Allan Sanford proved valuable additions to my advisory committee. And of course gratitude is reserved for the CTBT R&D group at Sandia National Laboratories for paying for all of this. Indeed collaboration between Sandia and Tech has been mutually beneficial and it is difficult to discern where the ideas from one end and the other begin. Thanks go to Rebecca Knight at the Center for Earthquake Research and Information at the University of Memphis for helping to transform my disjointed thoughts into an organized first draft. Finally, my entire education would not have been possible without the moral and financial support of the Withers family in Pennsylvania.

Table of Contents

List of Figures

List of Tables

1. Introduction

Automation can significantly improve the efficiency of seismic network data processing. Consider, earthquakes and mining blasts greater than $m_b \sim 3$ occur on a global scale at the rate of approximately $50,000$ per year (e.g. Gutenberg and Richter, 1954). Within the Socorro Seismic Anomaly (Sanford *et al.*, 1995)—which accounts for about 36% of New Mexico seismicity but covers only approximately 2% of the area of the state—199 earthquakes greater than $m_d \sim 2.0$ were recorded for the 34 year period from 1962 through 1995 (Lin et el., 1996). This seismicity may be related to a mid-crustal magma sill which has a "strong spatial coherence" with the seismic anomaly (Balch, 1997). The digitized New Mexico Tech Seismic Network (NMTSN) triggers approximately 1800 times per year on local events, mine and quarry activity, earthquakes from New Mexico, west Texas, and eastern Arizona, and teleseismic events. Reliance on manually locating these events requires an enormous investment in human resources—during a period of modest funding for seismic networks—thus providing motivation to automate the process. The Comprehensive Test Ban Treaty signed by President Clinton on September 24, 1996 has also stimulated great interest in automating the event location process for the treaty verification community. Current global detection thresholds for the CTBT sanctioned International Monitoring System are capable of detecting a well coupled 1 kiloton blast ($m_b \sim 4$, see van der Vink et al., 1996). Increasing the station coverage and analysis resources by combining global and regional seismic systems (which is consistent with recent efforts of the U.S. monitoring community to closely monitor, or "highlight," certain areas of the globe) can reduce the detection threshold in many areas of the world to a decoupled 1 kiloton blast ($m_b \sim 2$). Further, because explosions and earthquakes may not occur during normal working hours, and because it may be critical to respond as quickly as possible after an event is detected (e.g. disaster response for a large earthquake or site inspection for a suspicious explosion) it is desirable to develop an automated location system which can provide rapid warning of potentially interesting events. After an alert has been sounded, the automatic location can be refined by seismic analysts to provide more detailed and accurate information. In many cases the human resources do not currently exist to carefully locate the many smaller earthquakes, making an automated (although approximate) location useful for updating a comprehensive unreviewed catalog of seismicity.

There are existing automated detection and location systems that provide timely seismic bulletins, but they have shortcomings. Two of the most widely used systems,*Earthworm* (Johnson, 1995) and *Global Association* (*GA*; Le Bras et al., 1994b), rely on explicit phase identification and complex logical association techniques (which often use only P and S, or just P) which must be reformulated for different velocity models. In collaboration with the CTBT Research and Development group at Sandia National Laboratories, Albuquerque, New Mexico, we have developed an automated location algorithm based on work by Shearer (1994) which does not rely on explicit phase identification, incorporates model adjustments much more efficiently than *Earthworm* or *GA*, and has the capability to use much more of the information available in the waveforms. The global version of this system, under development at Sandia Labs, is the Waveform Correlation Event Detection System (*WCEDS*) (Young et al., 1996). The local version, being developed at New Mexico Tech, is the Local Waveform Correlation Event Detection System (*LWCEDS*). Although program development for the global and local correlation systems has been performed separately, they use similar algorithms and mutual collaboration has been extensive. The primary difference between the two is that *WCEDS* is designed to operate on continuous waveform data while *LWCEDS* operates on finite length time series from triggered systems. With the anticipated 50 to 60 station International Monitoring System, a global network is likely to observe overlapping phases from multiple events—particularly local events superimposed on teleseismic arrivals. Thus the global correlation is performed iteratively from the largest to smallest events in a given time segment. Indeed, most global location algorithms have subroutines that group arrivals into sets from the same event (i.e. associate arrivals to events) to be subsequently used in the location (e.g. Le Bras et al., 1994a). *LWCEDS* exploits the fact that the local network is less likely to record overlapping phases from multiple events and thus considerable savings in processing can be gained by operating on triggered time series and assuming records from a given trigger contain a single event. The largest event in the triggered data stream will be located but other smaller events will be missed. A proven system to generate triggered data streams is already in operation at the NMTSN (Lee and Dodge, 1992 and Skov, 1994). By automatically producing a seismic bulletin, the intent is not to replace seismic analysts, but rather to provide preliminary, unreviewed locations which can subsequently be refined using traditional location methods (e.g. manually revised picks and generalized least squares location). The *WCEDS*/*LWCEDS* routines provide more efficient use of human resources in two ways. First, for many events the *LWCEDS* location may be adequate. Second, even for events that must be more precisely located it is generally far less demanding to relocate an event than it is to start from scratch. A further benefit is that bulletins and alerts of preliminary automatic locations can be generated in near real time.

Currently the most widely used automatic earthquake location systems are *GA* and *Earthworm* and these are reviewed in the following sections. If successful, the *LWCEDS/WCEDS* algorithm may be deployed to augment these systems and could develop into the basis for the next generation of automated locators.

1.1 Global Association

GA (Le Bras et al., 1994b) is the global system used by the Group of Scientific Experts Third Technical Test (GSETT-3; e.g. Console, 1994) designed to test monitoring scenarios for the CTBT and was germinated from the Intelligent Monitoring System (IMS; Bache, 1990). *GA* is used by the prototype International Data Center (IDC) which has the mission of monitoring compliance with the CTBT. It is a grid search scheme consisting of three principal modules: *GAcons*, *StaPro* and *GAassoc*. *GAcons* is the routine that constructs the global grid and is used only when a new grid is needed. *StaPro* performs the initial processing and phase identification for each station. *GAassoc* combines the *StaPro* information from all stations, associates the phase arrivals with a particular event, and resolves conflicts between phases used for multiple event locations. *GA* must identify arrivals, and uses a complex logic flow to locate global events, as described in detail by Le Bras et al. (1994a, 1994b) and Beal et al. (1995). Each module is summarized in greater detail below.

GAcons constructs an approximately equally spaced grid over the surface of the earth. Grid points are also constructed for user-defined depths in areas where deep seismicity is known to occur. For each grid cell, information is included which defines a list of arrivals at each station. For each of the arrivals, estimates are made and stored for travel time, slowness, azimuth and probability of detection over a range of earthquake magnitude.

For each detected arrival *StaPro* performs phase identification, initial association, and writes the results to a Center for Seismic Studies (CSS) database subsequently used by *GAassoc* (see Section 2 and Appendix II of this document for a discussion of the CSS version 3.0 schema). Arrivals are automatically "picked" and classified into four wave types: teleseismic, regional P, regional S, or noise. Classification of the arrivals is based on a variety of tests which may be station specific and include amplitude threshold, rectilinearity of particle motion, dominant period, frequency-wavenumber analysis, and the ratio of horizontal to vertical power. An initial attempt to group the phases based on predicted arrivals is made, and single-station locations and magnitude estimates are used to "filter" local and regional events. Single station locations are possible because only arrays and 3-component broadband instruments are used in the IDC primary network. The results are then written to a modified CSS database.

GAassoc reads the single station results of *StaPro* from the database, combines information from the entire network to associate phases with specific events and grid locations, and resolves conflicts caused by arrivals that have been associated with multiple events by the single station processing. The first arrival is identified as the "driver" arrival for each grid point. These arrivals are only allowed at a limited number of stations as defined by *GAcons* and must meet specified slowness parameters. A candidate location is found based on the "driver" arrival. Later arrivals at all stations are compared with predictions in a search for groups that are related to the "driver." Weighted preliminary event confirmation criteria which include checking residuals, azimuths and slowness information are then applied. Redundancy checks, conflicting arrivals, and various other tests are performed to arrive at a final solution which must pass more restrictive confirmation criteria than was applied in *StaPro*.

A test of *GA* on a synthetic data set which included a large swarm on the first day and normal seismic activity on four subsequent days revealed mixed results. Of the events in the data set, 82% were detected and 42% were located within 50 km of the known epicenter. Approximately 40% of the events in the automatically generated bulletin were false alarms (e.g. detections based on noise) and are likely due to an unrealistically low detection threshold. 18% of the events in the synthetic data set were missed by *GA*. Most of these missed events are due to errors resulting from complex logic related to phase identification and association (Le Bras et al. 1994b).

LWCEDS avoids many of the *added* event and false alarm pitfalls by employing *Xdetect* (Lee and Dodge, 1992) to generate detections and triggered time series, and by avoiding explicit phase identification and requisite association logic.

1.2 Earthworm

Earthworm (Johnson, 1995) was originally developed by the U.S. Geological Survey's Joint Seismic Program and is deployed at numerous institutions for automated local and regional network processing. It is a collection of processing modules that send and receive data packets across a common communication connection. This connection can be an ethernet line or a block of shared computer memory. *Earthworm* is intended to provide near real time seismic data processing and automated bulletin generation with alert capabilities. Significant modifications, enhancements and additions have been performed at various institutions. In various forms, *Earthworm* is currently used at most major regional seismic networks in the U.S. Indeed, so many disparate *Earthworm* modules have been developed that an *Earthworm* Community Center was approved by the Council of the National Seismic System at the 1997 annual meeting in Memphis. One goal of the Community is to establish a standard *Earthworm* release package. Part of this package will be the P-Picker and associator modules. The P-Picker is based on the work of Rex Allen (1978) which produces a processed data stream for the i^{th} time step, E_i, by summing a weighted square of the raw data, x_i, and a weighted square of the first derivative of the data:

$$E_i = a{x_i}^2 + b\frac{(x_i - x_{i-1})^2}{\Delta t} \qquad\qquad i = 1 \ldots N$$

where Δt is the length of a time step (e.g. sample rate), N is the number of data points, and a and b are user defined parameters. Amplitude and minimum duration tests are applied to E, and routines are performed to determine accurate p-picks.

The output of the Allen picker is sent to the associator module using the common communication connection. The associator module (Johnson, 1994) back-projects for all arrivals within a given time window over a four dimensional grid (time and three spatial dimensions). For each p-pick a suite of candidate locations are back-projected through the velocity model. The grid point where the epicenter projections overlap for a majority of picks is used as the automatic location. The actual implementation is a matrix solution of N picks for M hypocenters that minimizes the residuals. A weighting scheme is applied to favor solutions with many picks to prevent large events that are observed at many stations from being distributed into multiple smaller events with fewer picks each (Carl Johnson, pers. comm. 1997).

While *Earthworm* and *GA* are the most widely used systems, most automatic location algorithms have three common elements. First, the arrivals of various seismic waves must be timed and identified. Second, these arrivals must be associated with a single earthquake or separated into groups that are related to multiple events. And third, hypocenters must be estimated based on the associated arrivals. It is often difficult, and requires complex logic, to identify arrivals and associate them with specific events. Thus we are motivated to develop an automated location system which avoids the complex logic that is necessarily contained in other automated location systems.

2. Statement of Purpose

Both *GA* and *Earthworm* must first generate arrival times for various phases (detections). *LWCEDS* does not use explicit phase picking but rather correlates the entire processed waveform with a predicted processed waveform. We are thus able to avoid the complexity and potential inflexibility inherent in phase detection-based systems. For *LWCEDS*, all that is needed to modify the velocity structure and the expected pattern and assortment of phase arrivals is a new model rather than a new logic set. The goal of this research is to develop a local implementation of the waveform correlation event detection system which can automatically locate seismic events in a variety of seismogenic situations. Events within the Socorro Seismic Anomaly and at regional distances in New Mexico were used as a test bed.

3. Methods

3.1 Basic Algorithm

LWCEDS is based on a correlation and grid search algorithm. Consider the hypothetical arrangement of stations in Figure 3.1, where three stations (S1, S2, and S3) have recorded an event from epicenter E1. For illustrative purposes, the epicenter is hypothesized to be at either grid point G1 or G2.

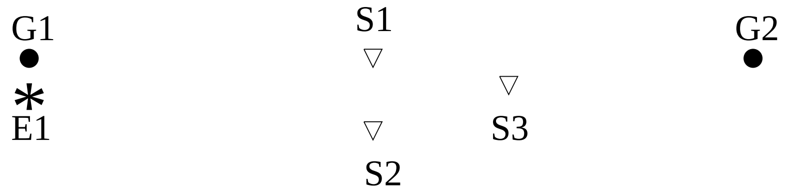
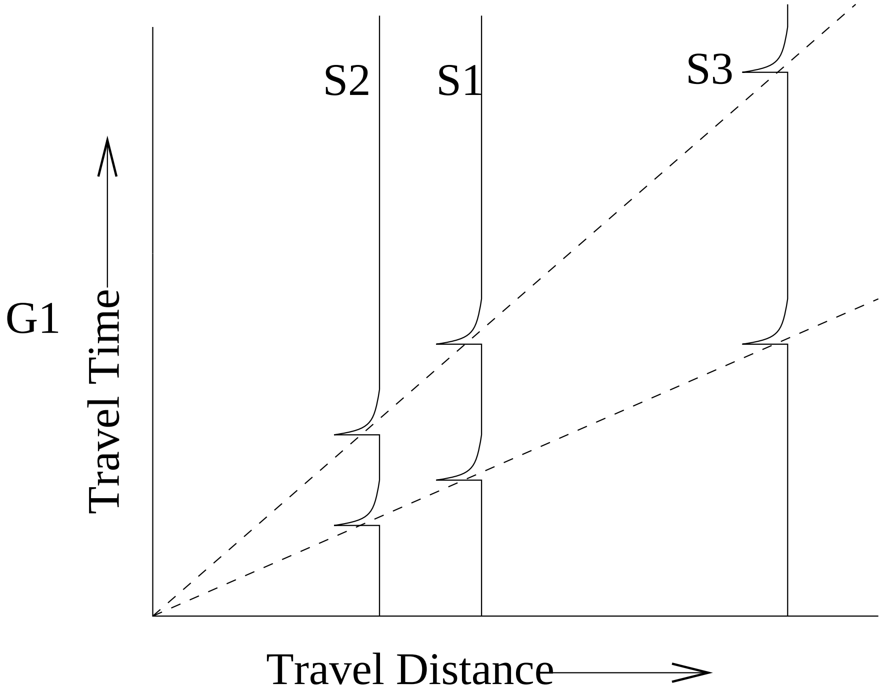


Figure 3.1 Hypothetical Station Arrangement illustrating different travel distances. For stations S1, S2, and S3, and an earthquake at E1, candidate location G1 will provide a good approximation between the observed and theoretical arrivals. Candidate location G2 will provide a poor, but non-zero correlation between predicted and observed arrivals.

A plot of the theoretical data from each station at the appropriate distance for a given grid point, allows a visual estimate of the correlation value for the grid point. Point G1 (top of Figure 3.2) is a good approximation for the epicenter, but G2 is not (bottom of Figure 3.2). Note that the correlation value for G2 is non-zero because the second observed arrival at S2 aligns with the first theoretical arrival at that distance. This correlation value is much smaller than the value for G1, because all observed arrivals align well with the theoretical arrivals there. The same theoretical and observed arrivals are used for both cases; only the hypothesized travel distance is changed.



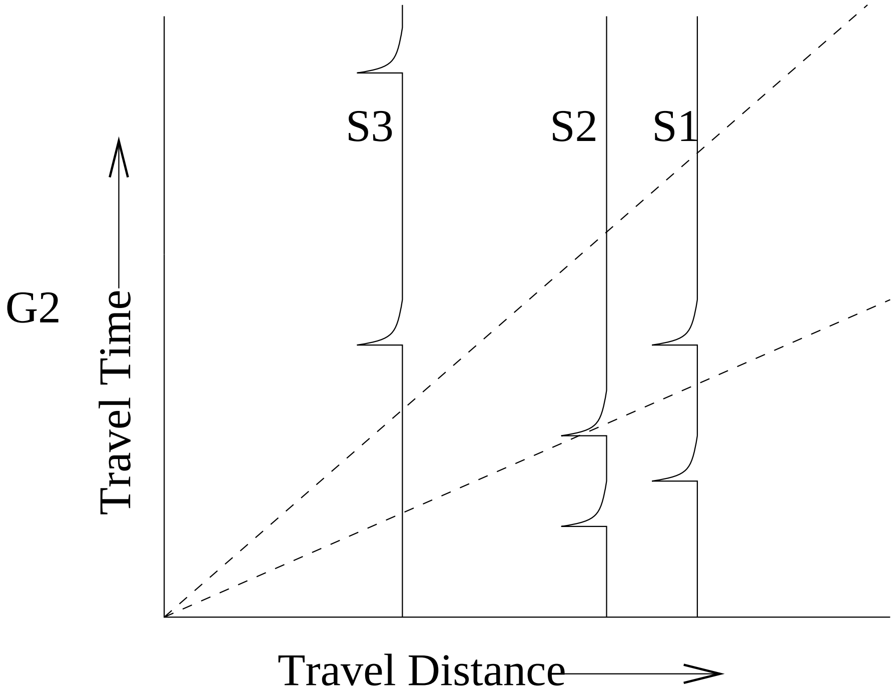


Figure 3.2 Hypothetical Correlation showing theoretical and observed arrival alignment for grid points G1 (top) and G2 (bottom) from Figure 3.1. Even though point G2 is a poor approximation to the correct hypocenter the observed second arrival at station S2 aligns with the theoretical first arrival yielding a small but non-zero correlation.

After a comprehensive grid is constructed for a given area of interest (e.g. the Socorro Seismic Anomaly), a given event is located by calculating the correlation value for every grid point and then searching the grid for the maximum correlation. Each grid point will have an associated correlation value, and the maximum value indicates the grid point that is the best estimate of the true epicenter.

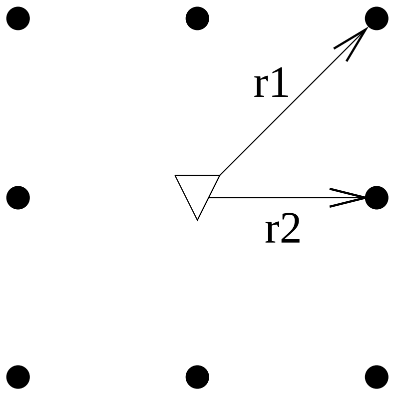


Figure 3.3 Hypothetical Grid illustrating equidistant grid points from a given station which, assuming azimuthal symmetry in the velocity model, are redundant correlations.

Performing a correlation for each grid point is time consuming and contains many potentially redundant calculations. Consider the grid points surrounding a station as shown in Figure 3.3. There are four points shown with distance r1 from the station, and another four points with distance r2. If azimuthal symmetry in the velocity model is assumed, then the correlation value for any station will be identical for a number of grid points. In Figure 3.3, the four points with distance r1 will all yield the same correlation value for this station. Thus a complete correlation for each grid point is not necessary. Instead, the correlation value for each station at each distance is calculated. This process can be performed in a single matrix multiplication (Young et al., 1996) as illustrated in Figure 3.4. M is a matrix containing theoretical data. D is a matrix containing observed data which has been processed to enhance arrivals, and C is the correlation matrix which contains a correlation value for each station at every distance.

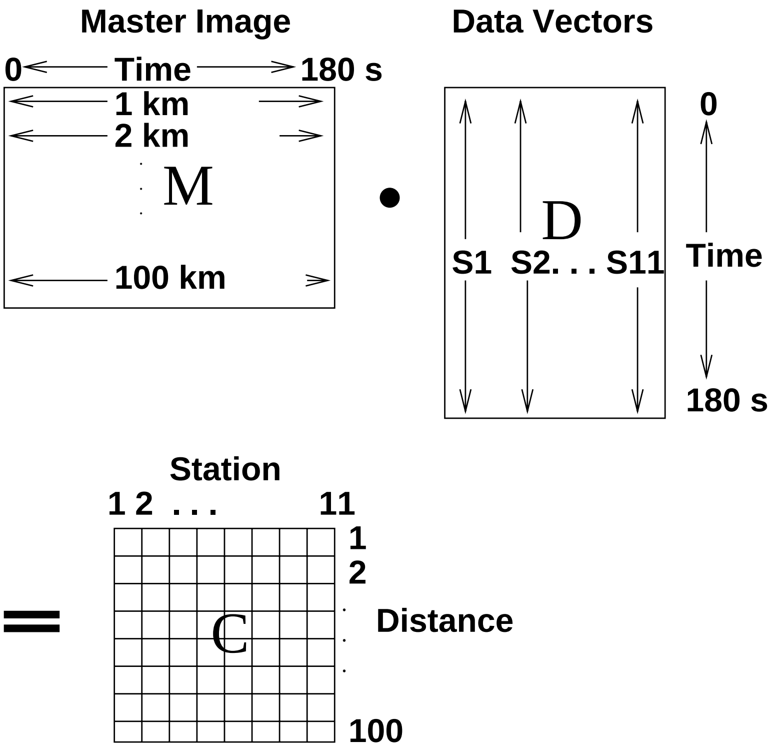


Figure 3.4 Basic Algorithm schematically illustrating the correlation of data from each station with every distance. Preprocessed data are contained in D, the model matrix is M, and the matrix multiplication yields the correlation matrix C. C contains correlation values for data from every station with model data from every distance.

3.2 Model Matrix

If there are data for a sufficient number of known hypocenters, the model matrix can be constructed by stacking observed processed data, yielding an empirical stack similar to that of Shearer et al. (1991). For *LWCEDS* however, M was constructed for several depths using theoretical travel-time curves based on the Hartse (1991) *curvfnd* algorithm. Examples of the curves are shown in Figure 3.5. These travel-time curves are found by ray tracing using the crustal model of Hartse (1992) which is three layers over a half space with a magma sill at 19 km (Figure 3.6). While depth to the Moho varies considerably throughout the state (from about 30 km to more than 50 km), this model was also used at regional scales for simplicity (the magma sill was only used at local scales).

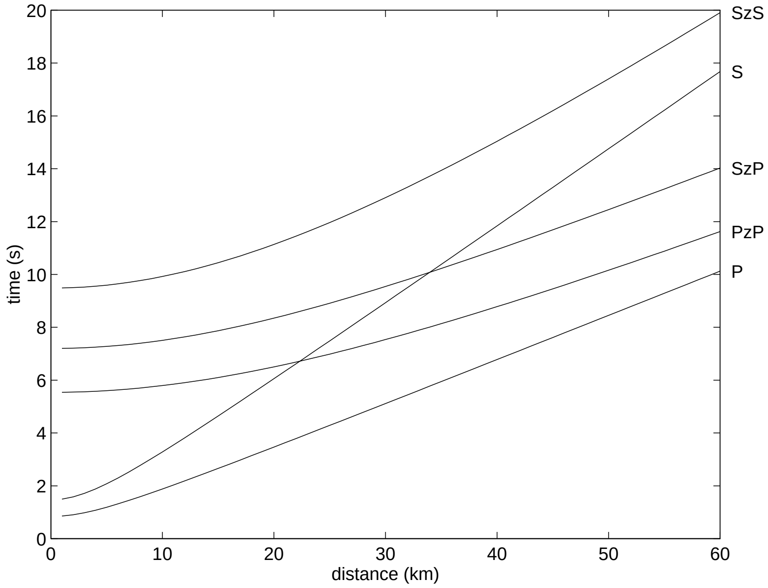


Figure 3.5 Travel Time Curves for the Socorro Seismic Anomaly from an event at 5 km depth. The curves are calculated using the Hartse (1991) raytracing algorithm and the Hartse (1992) velocity model. Magma sill reflections are indicated with a z in the arrival identifier, e.g. SzS.

The minimum lateral extent of the magma body covers more than 3400 km^2 (Balch et al., 1997) and is shown in Figure 3.7. Minimum distances for refracted phases are governed by the critical distance, and maximum distances for reflected phases are governed by the lateral extent of the reflector which is used to estimate the most distant reflection point from any station.

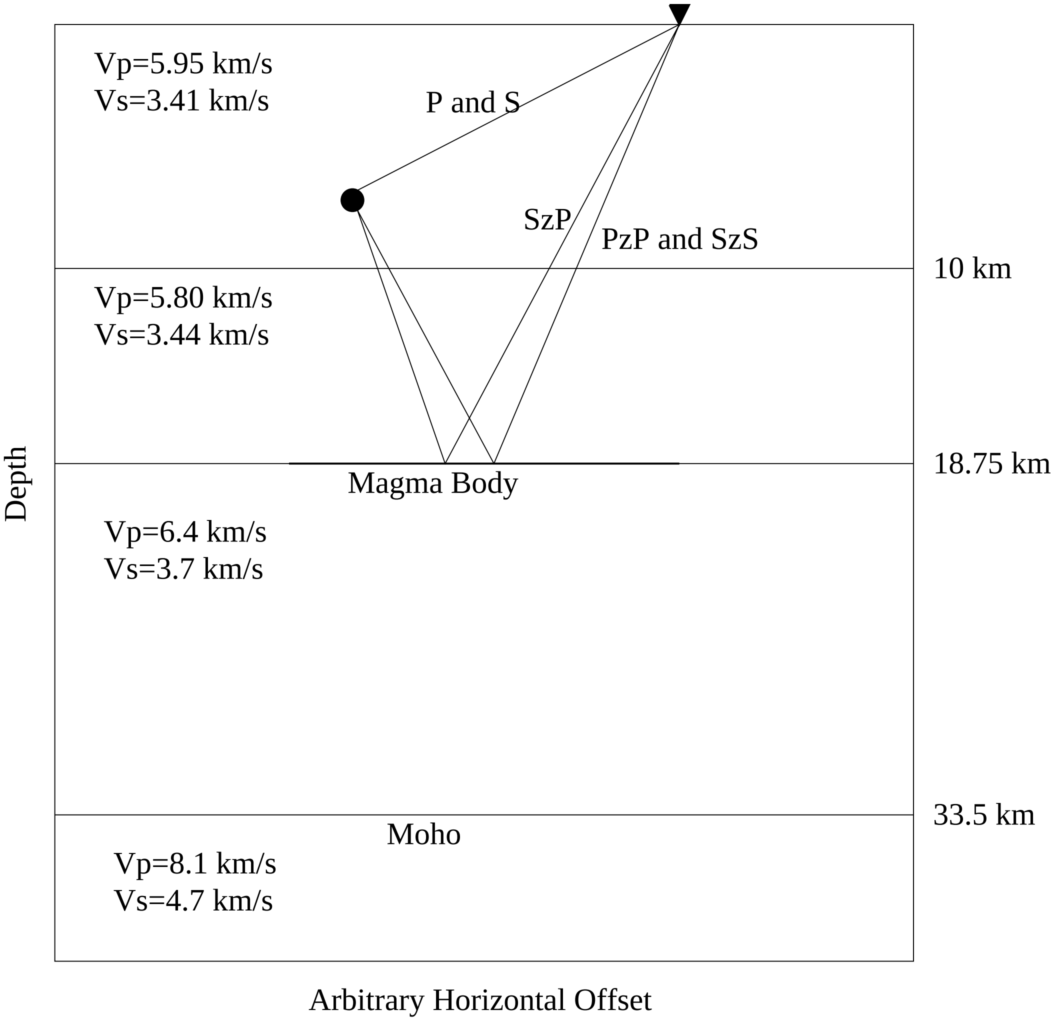


Figure 3.6 Velocity Model schematically illustrating the model developed by Hartse (1992) appropriate for the Socorro area of New Mexico and illustrating hypothetical travel paths for magma sill reflections (after Balch, 1997).

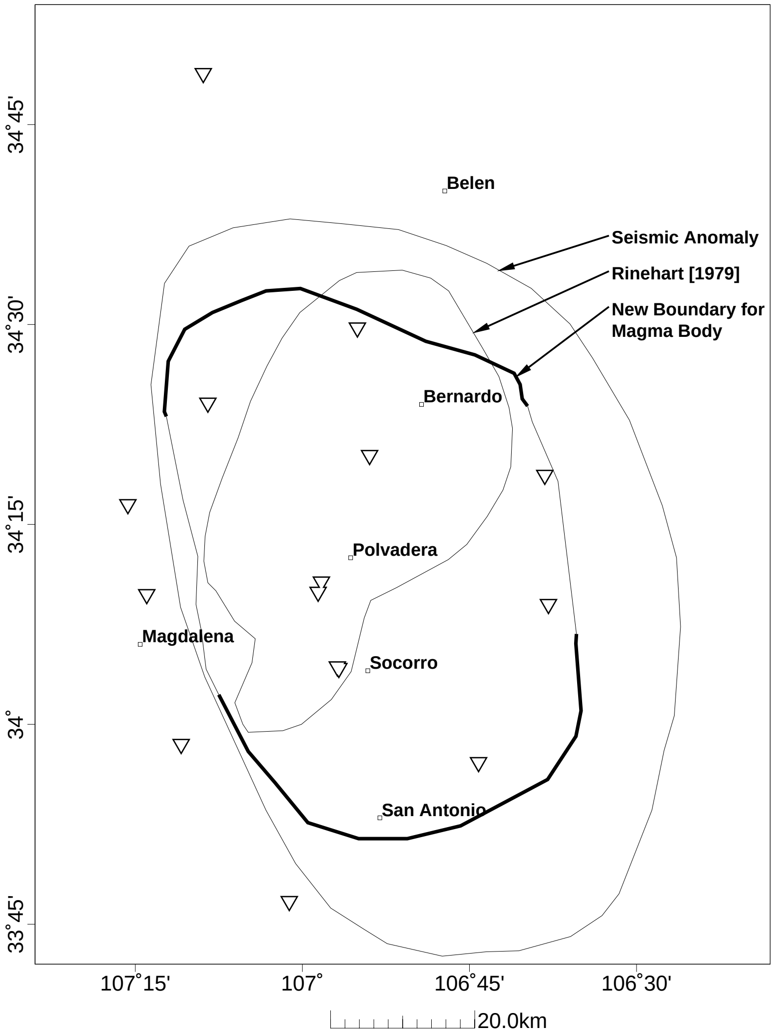


Figure 3.7 Magma Sill Map showing the geographic extent of the magma sill (from Balch, 1997).

+

To generate the model matrix, M, the travel-time curves are transformed into a set of time series by applying an envelope function to the arrival times for each distance bin. *LWCEDS* currently uses a rectangular envelope with a phase dependent minimum width and a maximum width of 10 s. The width of the envelope is determined by the grid spacing of potential epicenters, by the apparent velocity of the travel time curve and by the discretization of the model matrix. Because the model matrix may be empirically derived rather than generated from a theoretical model, M is also referred to as the Master Image.

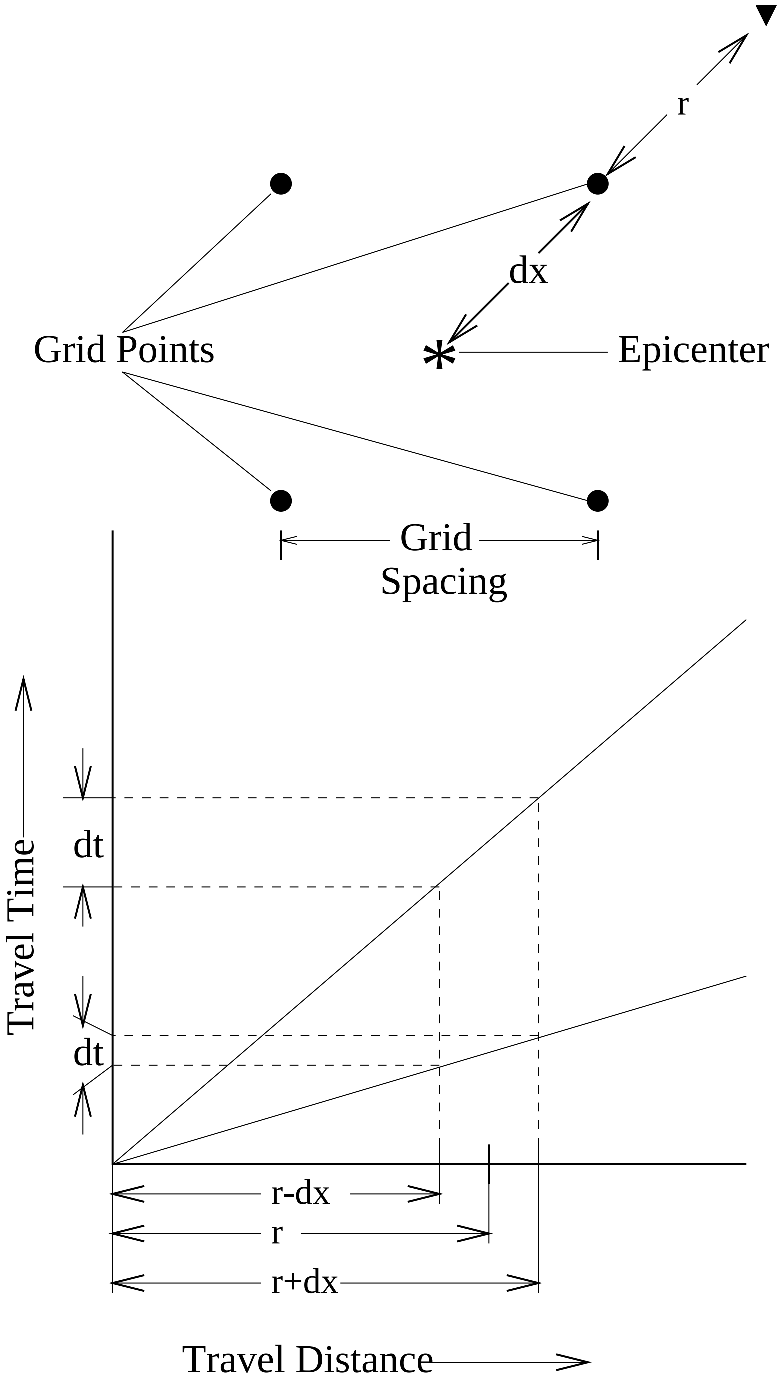


Figure 3.8 Phase Width illustration. For the grid spacing illustrated at top, the distance between the station and grid point is r, and the distance to the "true" epicenter will be r±dx. The maximum error, dx, for a square grid is related to the grid spacing by $dx = \frac{(grid\ spacing)}{\sqrt{2}}$. This error translates into a maximum envelope width, dt, for a given phase based on the slope (apparent velocity) of the travel time curve for that phase (bottom).

By using a discrete grid of potential epicenters, we introduce error of dx (Figure 3.8), the maximum of which for a square grid is related to the grid spacing by $dx = \frac{(grid\ \ spacing)}{\sqrt{2}}$. The error, dx, from imposing a discrete grid will map into time, dt, by the slope of the travel time curve. Ignoring all other factors, dt is the desired width of the arrival envelope. Both dt and dx may also depend on the time and distance discretization of the Master Image which sets minimum values for the time and distance error. For reflected phases, the slope of the travel time curve approaches zero as travel distance approaches zero which could produce an envelope of zero width—ignoring the minima set by the time bins in the Master Image. Thus, to allow flexibility in the selection of Grid and Master Image discretization parameters, we impose a 0.1 s minimum envelope width for p-arrivals (*LWCEDS* currently decimates data to 10 sps) and 0.2 s for s and reflected arrivals. The final distance error is selected from the larger of the error associated with the grid spacing and the distance discretization in the Master Image. The final envelope width is determined by the larger of the apparent velocity slope, the width of the time bins in the Master Image, and a parameter used to define an absolute minimum width for each phase.

The height (or weight) of the arrival envelopes depend on phase and width. The first observed arrival is compressional and usually has the largest signal-to-noise ratio (SNR). This behavior is mimicked in the model. The largest SNR peak is aligned with the first arrival, by giving the predicted p-arrival the largest weight. S-arrivals are given intermediate weights and other phases (e.g. PzP) are given lower weights. The current values of the weights used for LWCEDS (determined by trial and error) are $P = 1.4,\ \ S = 1.2,\ \ SzS = 1.0,\ \ PzP = 1.0,\ \ SzP = 1.0$. The PzS phase is not currently used since it has yet to be observed.

The weights are further adjusted by the envelope width. Consider the occurrence of a long period arrival with a slow apparent velocity (e.g. L_g). The width of the theoretical envelope is wide due to the slow apparent velocity, and the observed processed data envelope is also wide because of the long period coda of this hypothetical arrival. Because the correlation value is found using a vector dot product, and because this hypothetical arrival contains more points than other shorter period, faster apparent velocity arrivals, it has relatively more non-zero points in the summation. This effectively gives greater weight to later long period arrivals which generally provide less accurate information for location purposes. To mitigate this effect the weight is normalized by the width. For example, if the theoretical P-envelope is 10 points wide with a weight of 1.4, then the actual height of the envelope is 0.14. This prevents unintentionally giving too large a weight to slow phases that have wide theoretical envelopes. The model matrices used in this study for a 5 km source depth are shown at local and regional scales in the top and bottom of Figure 3.9, respectively. $\mathrm{S_n}$ is not currently used in the regional model because it is usually overwhelmed by compressional body wave coda.

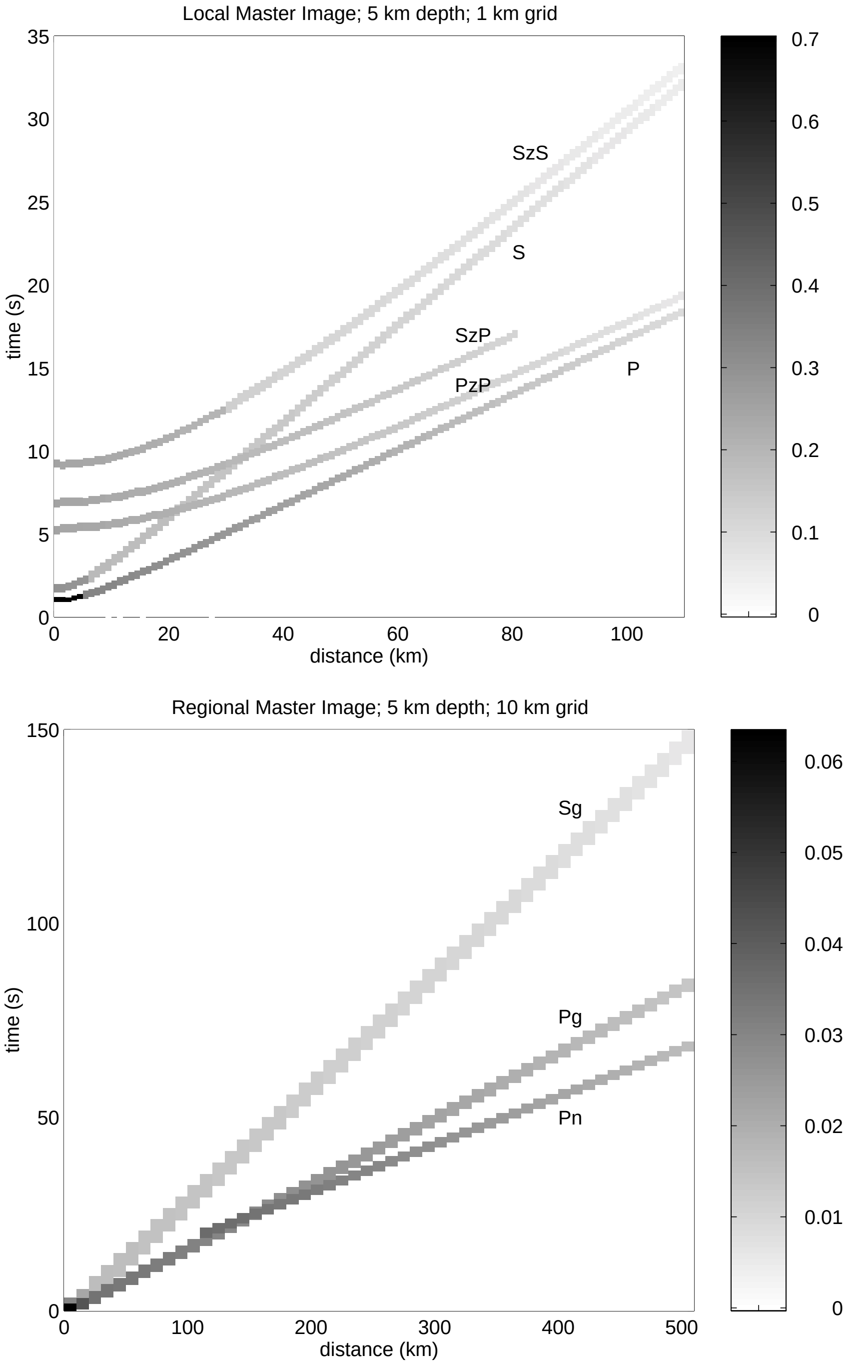


Figure 3.9 Master Images used in this study. The local Master Image (top) is calculated from travel time curves for an event at 5 km depth. Phase weight is given by the color bar. The regional Master Image (bottom) is calculated from travel time curves for an event at 5 km depth. The weights are normalized by the respective arrival envelope widths to prevent later, longer period arrivals from dominating the correlation. A travel time and distance dependent plane with negative slope is applied to mimic signal attenuation.

Note further in Figure 3.9 that the envelope weight decreases linearly with time and distance. After the phase weights, and widths have been determined as described above a preliminary, time and distance dependent Master Image, $m(t, d)$, is generated. The final Master Image, $\mathrm{M}(t, d)$ is produced by applying, a time (t) and distance (d) dependent plane to the preliminary image:

$$\mathrm{M}(t, d) = m(t, d) * (a + bd + ct),$$

where a, b, and c are user defined parameters (currently $a = 1.0$, $b = -0.6$, and $c = -0.4$). Particularly for clipped data where only one arrival is observed on all data streams, there can be cases where equal correlation values occur for more than one grid point and the preferred location may be governed by relatively small peaks in the background noise. By applying this travel time and travel distance dependent plane to the weights, arrivals with smaller travel distances and times are favored. The plane weighting is intended to crudely mimic the travel distance dependent attenuation of signal amplitude.

3.3 Data Processing

LWCEDS presently uses data recorded by 9 short-period (1 Hz) seismographs, which are FM telemetered to NMT where the data are digitized at a nominal 100 samples per second on an IBM 386 clone. *Xdetect* (Lee, 1992) is used to trigger and record time series on the PC which are transferred daily to the Geophysics/Hydrology network of Sun workstations (Skov, 1994). *LWCEDS* operates on these triggered data streams and locates no more than a single event per trigger; only the hypocenter for the event with the highest correlation value above a user defined detection threshold is stored as opposed to finding all events contained within a given triggered data stream. The locations of the local stations are shown in Figure 3.10. All stations are short period (1 Hz) single-component verticals except for station LEM which was upgraded to 3-component on July 2, 1996 to facilitate recording S-arrivals.

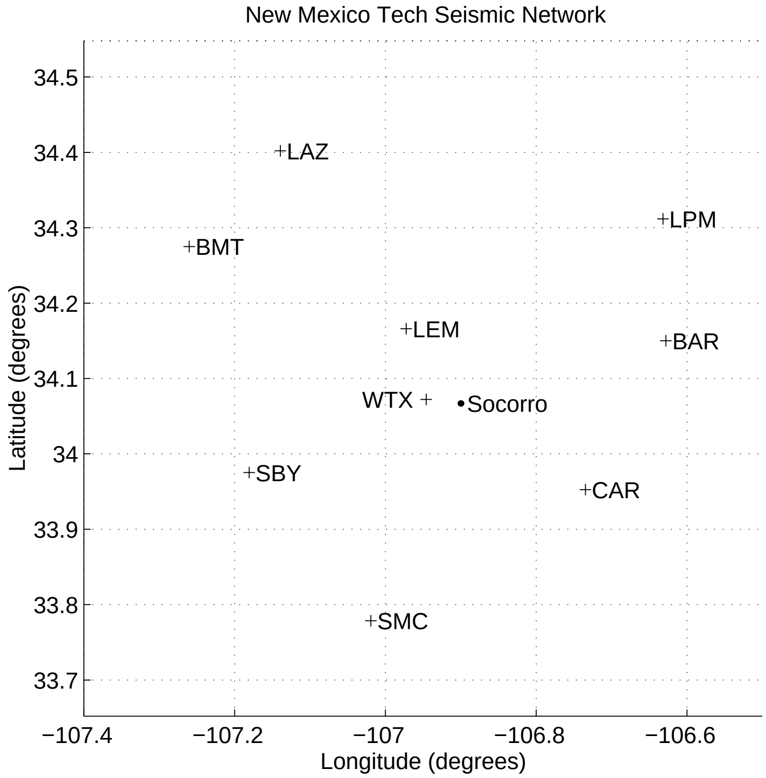


Figure 3.10 Station Locations in the local network used in this study.

After converting the data from the PC-based data-logger output to a more convenient format (University of Alaska, Fairbanks version of *ah*) they are filtered via a time domain convolution with a 100 point, 3 pole, 1 to 35 Hz bandpass butterworth filter to remove microseismic and aliased 60 Hz noise. The data are then processed to provide an envelope which has smooth, low amplitude response to background noise and seismic coda and which peaks at a maximal number of phase arrivals. The preprocessing is performed to optimize the correlation with the Master Image. Even if an empirical Master Image were used, preprocessing would be necessary prior to stacking the image to prevent convolution with a model that would contain a prohibitive level of crustal detail (i.e. matching the seismograms ”wiggle for wiggle”). An unprocessed model would necessarily include detailed information regarding the effects on seismic waves of azimuth, distance, anisotropy, source, site and many other effects. A traditional method for preprocessing is computation of the ratio of average energy in a short time window to that in long time window (STA/LTA). The response of the STA/LTA is dependent on the lengths of the windows and because it has been traditionally used as a detector, the window length of the STA/LTA is usually tuned to respond best to the expected first arrival. Because the frequency content of phase arrivals can vary considerably and because *LWCEDS* incorporates all available arrivals, an STA/LTA algorithm with adaptively varying window lengths was developed. Figure 3.11 schematically illustrates this operation. As a fast estimator of the dominant frequency content, the zero crossings in the demeaned data are used to establish the window length. A reasonable compromise between sensitivity and noise reduction is obtained by requiring 5 zero-crossings in the short-term window (2.5 ”cycles”). The long-term window is then set to 5 times the length of the short-term window. This strategy was applied to the non-recursive STA/LTA (boxcar impulse response), operating on the energy (square of the data; 3-component data are summed in a Pythagorean sense):

$$STA_i = STA_{i-1} + \frac{1}{Nsta}(x_i^2 - x_{i-Nsta}^2) \qquad \mathrm{non-recursive}$$

where STA_i is the short-term average energy at time step i, x_i is the data point at i, and $Nsta$ is the number of points in the short-term window.

Tong (1995) developed an adaptive STA/LTA which selects zero-crossings by requiring that the first derivative be greater than some threshold value. While this may be a more robust method and warrants further investigation, Tong's method incorporates a low pass filter, sections the data, and has constant window length within each section. Taking the derivative of the data applies a low pass filter which removes signal information contained in data recorded at local and regional distances. Further, adjusting the window length on a point by point basis is preferable to prevent discontinuities in the processed data stream and to allow more rapid updating of the window length. Tong also uses the recursive implementation of the STA/LTA, adjusting the effective window length by changing the decay constant. It is common to use the recursive STA/LTA to avoid keeping a long data vector in memory. This is more efficient and allows a decaying exponential, which will recover more quickly from large energy transients, rather than a rectangular impulse response. The recursive scheme also yields smaller "shadow zones" which are created when a large transient passes the STA and continues to dominate the output by causing a large LTA. The characteristic decay time, T, is the time required for the impulse response to decay to $\frac{1}{e}$ of its original value (Evans and Allen, 1983) and is embedded in the decay constant, C:

$$STA_i = Cx_i^2 + (1-C)STA_{i-1} \qquad \text{recursive}$$

$$C = 1 - e^{S/T}$$

where S is the sample rate, T is the characteristic decay time, and x_i is the data point at time step i (it is common to use $C = \frac{1}{Nsta}$ and $C = \frac{1}{Nlta}$, for the short and long-term decay constants).

Implementing the recursive STA/LTA in an adaptive window environment will produce discontinuities in the output and in some pathological cases could be mathematically imprecise. Consider, the recursive algorithm requires keeping only the previous average and one data point in memory and it represents an infinite summation:

$$\begin{aligned} STA_i &= Cx_i^2 + (1-C)STA_{i-1} \\ &= Cx_i^2 + (1-C)x_{i-1}^2 + (1-C)^2 x_{i-2}^2 + \dots \end{aligned}$$

where C represents the decay constant and x_i is the i^{th} data point. To change C requires resumming the entire data stream. One might be tempted to apply the new decay constant beginning at the i^{th} point but this presents the possibility of generating a very poor estimate. This is particularly true for the LTA in an expanding window (Figure 3.11) which can cause the LTA to be unjustifiably small because the appropriate new decay constant should be less than the old, shorter window, constant (but won't be without resumming all of the data). Thus we have opted to sacrifice memory and the exponential decay by using the non-recursive STA/LTA. Because the windows are usually short, there is very little penalty in not using the exponential decay and processing time is not significantly affected. For large data sets, memory use could become critical. A more comprehensive analysis of select trigger algorithms has been submitted for publication and is included in Appendix IV.

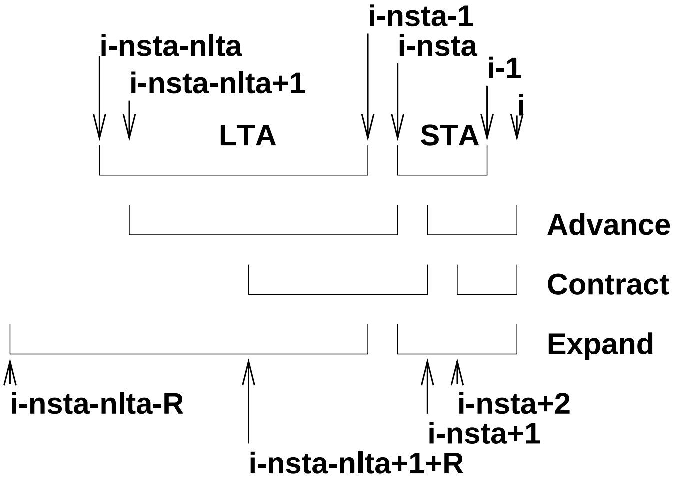


Figure 3.11 STA/LTA Index schematic showing the window placement for a single step forward to the i^{th} data point, a contracting window, and an expanding window, respectively from top to bottom. The indices are what are used in the C code. The variable, nsta, is the number of points in the short-term window and is selected such that 5 zero-crossings from the bandpass filtered data are contained in this window. The number of points in the long-term window is nlta=R*nsta, where R is currently set to 5.

Low-frequency signals (e.g. surface waves) may produce extremely broad peaks in the output of the adaptive STA/LTA, and this is not desirable for the *LWCEDS* correlation because of the large residuals associated with the travel times of these waves and because of the potentially large correlation value resulting from the broad peaks. This effect is mitigated by capping the STA window length at 0.5 s. Within the limitation imposed by the cap, STA windows may expand within the seismic coda if they encounter longer period arrivals. When the LTA expands backwards into a previous large arrival (e.g., into P while operating on PzP) the previous arrival may significantly perturb the long-term average and thus adversely bias the desired output peak corresponding to the later arrival (i.e. create a "shadow zone"). To control this effect, the additional data to be added into the back-expanding long-term window is examined. If the power in the new long-term window segment is greater than ten times that of a similar-length average piece of the previous long-term window (indicative of a large energy transient within the data to be added), then the STA/LTA calculation is performed using an appropriate fraction of the previous LTA energy instead of the data itself (not unlike the recursive realization where $C = \frac{1}{Nlta}$).

An STA/LTA maximum of 50 is set to prevent any single phase from completely dominating the correlation value (e.g. prevents small SzS contributions from being overwhelmed by P). Next, the data are decimated (currently from 100 sps to 10 sps) using a running mean window to reduce the processing time required for the correlation, which is currently the most computationally expensive part of *LWCEDS*. Finally, any STA/LTA values less than a given threshold (currently 2) are reset to zero to help prevent background noise from contributing to the correlation (which may become important for low SNR signals).

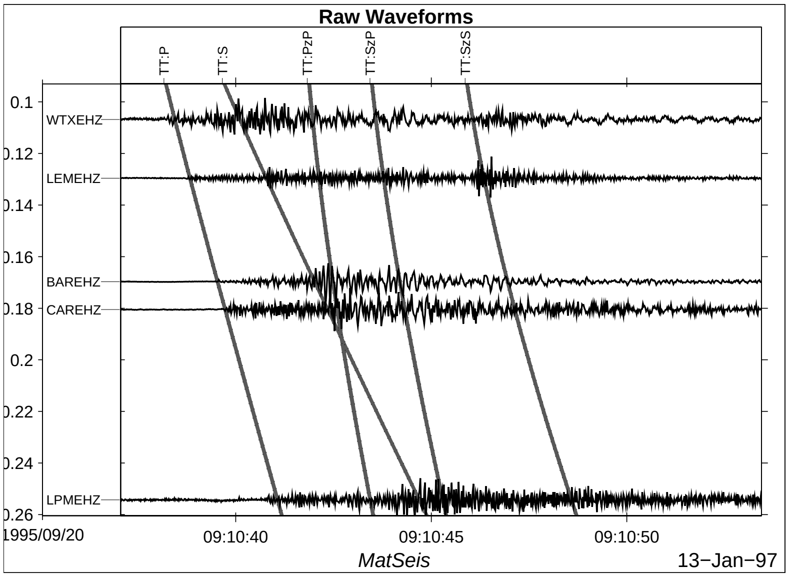


Figure 3.12 Example Raw Waveforms. A gather of unprocessed data are shown for an $m_d \sim 0.84$ earthquake near Polvadera on September 20, 1995. The waveforms are aligned vertically to reflect distance (in degrees) from the epicenter (closest at top). Local travel time curves are overlain for reference.

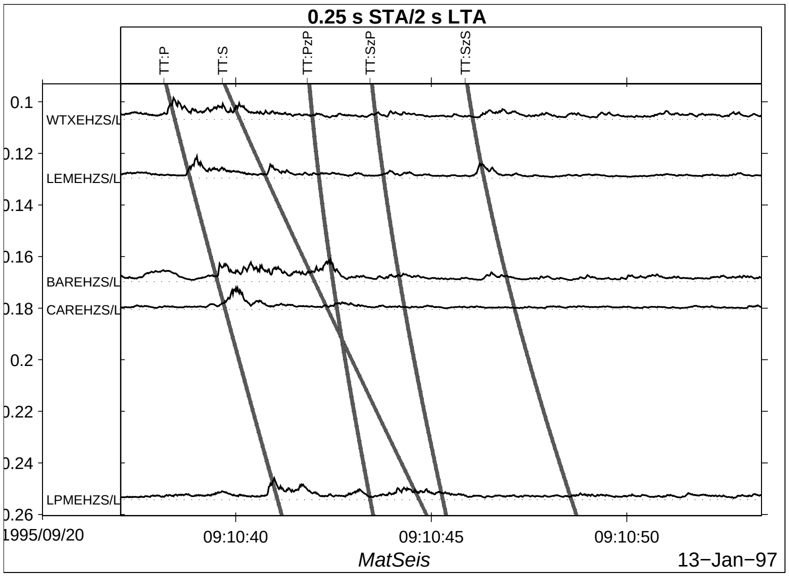


Figure 3.13 Example Processed Waveforms. A gather of waveforms processed with the recursive STA/LTA using a static 0.25 s short window and a 2 s long window.

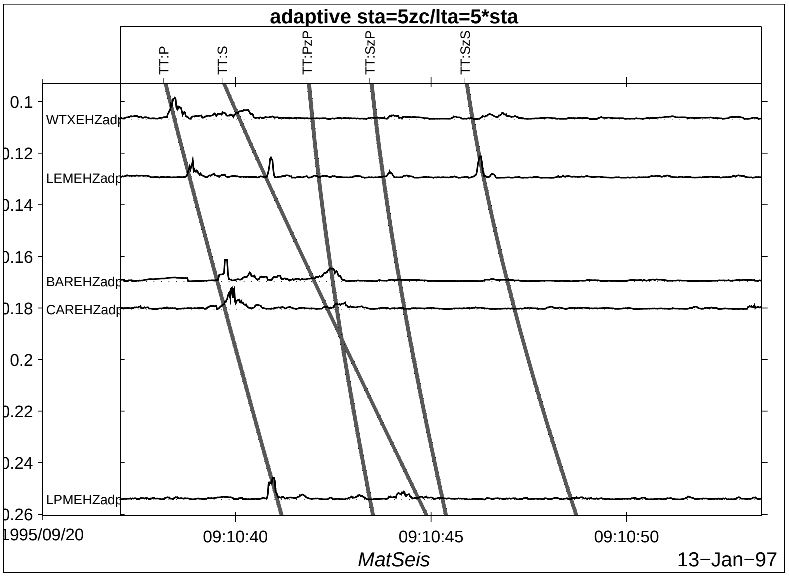


Figure 3.14 Example Adaptive Processed Waveforms. A gather of waveforms has been processed with the adaptive STA/LTA using 5 zero-crossing per short window and a long window 5 times the length of the short. Because the window lengths of this algorithm need not be "tuned" to a given arrival it can encompass energy transients over a broad frequency range.

Figure 3.12 illustrates example waveforms from an $m_d \sim 0.84$ earthquake near Polvadera on September 20, 1995. The waveforms are aligned vertically to reflect distance (in degrees) from the epicenter (closest station at top). Local travel time curves are overlain for reference. The profile in Figure 3.13 shows waveforms processed with the recursive STA/LTA using a static 0.25 s short window and a 2 s long window. For comparison, the data in Figure 3.14 has been processed with the adaptive STA/LTA using 5 zero-crossings per short window and a long window 5 times the length of the short. It can be seen that the first arrivals that have been preprocessed with the adaptive STA/LTA have much narrower peaks than the non-adaptive, yet still become broad where two later overlapping phases arrive nearly simultaneously. The adaptive STA/LTA need not be tuned to a given frequency band and is able to enhance both short and long period arrivals with a single algorithm. The resulting processed waveform is a truer representation of the data envelope than static window length algorithms are able to provide. Thus the adaptive STA/LTA is more suitable for correlation, and the algorithm requires only very small increases in memory use and processing time over the recursive STA/LTA. An additional data stream (zero-crossings) for each station is incorporated, but these could be determined ”on the fly” if computer memory use became critical.

3.4 Correlation Matrix

While the data matrix is generated for each trigger interval, the Master Image matrix is generated only once. These matrices are loaded into memory and the matrix multiplication is performed for the entire time series using the Numerical Recipes in C (1992) subroutine, *correl*. For each time step, the dot product of a Master Image vector for a given distance and a data vector must be computed to find one element of C (Figure 3.4). This element of C represents how well the data vector correlates with the model at that distance. For a given Master Image-data vector pair, the multiple dot products for successively increasing time lags are in fact the crosscorrelation, which can be performed more efficiently in the frequency domain with a single multiplication for each frequency bin (as opposed to a complete dot product for each time step).

After the Correlation matrix has been computed, the grid search is performed. For each grid point a contribution from each station is summed to determine the correlation value at that grid point. Each station (column in C) is a fixed horizontal distance from a given grid point (row in C) and thus for a given grid point, a value from the C matrix for each station is summed. The appropriate value included in the summation is governed by the discrete Master Image distance which is nearest to the horizontal distance between station and grid point. Indexing into the C matrix for the proper distance between each station and grid point is computationally expensive, so a map (the Cmap) providing these indices is precomputed. Hypothetical summation paths are illustrated in Figure 3.15 for grid point G1 (dashed line) and G2 (solid line) from Figure 3.1. For hypothetical epicenter G1, the appropriate values to be added are C(s1,d3), C(s2,d2), and C(s3,d6). For point G2 the elements of C to be summed are C(s1,d5), C(s2,d3), and C(s3,d2). Here S1-3 represent the station number and d1-6 represent the distance bins in the Master Image. A single value for each station is summed and the result is normalized by the number of stations to compute the final correlation value. The sum is normalized by the number of stations to provide event declaration thresholds that are independent of the number of stations in operation. After the summations using the Cmap are performed over the given time interval, specified depths and for all grid points, the maximum sum is determined. If this maximum sum is above a detection threshold, then an event is declared and the appropriate grid point, origin time, and depth are stored.

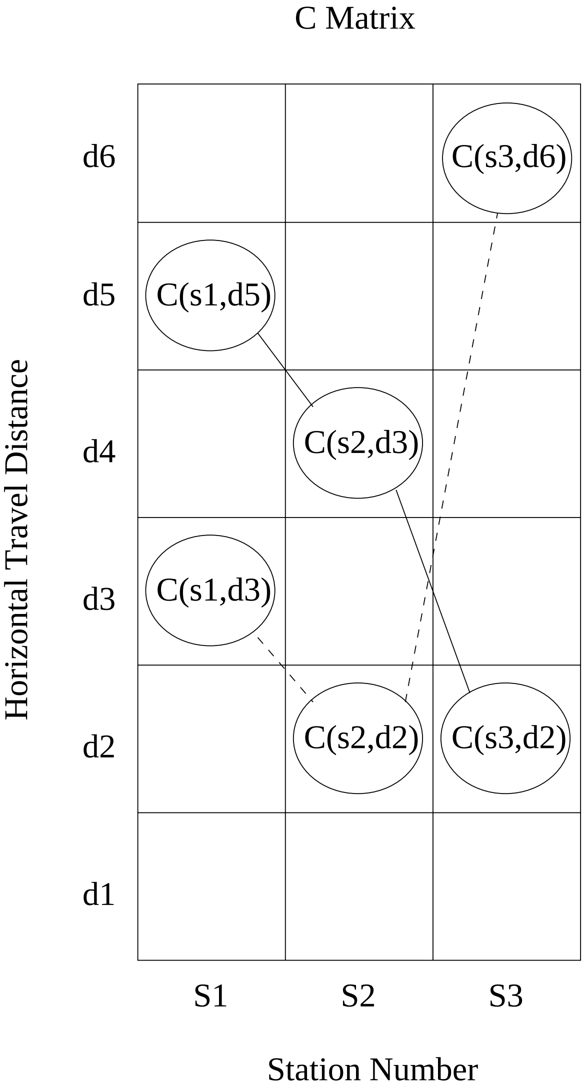


Figure 3.15 Example Correlation Matrix schematically illustrating summation paths for grid points G1 (dashed) and G2 (solid) from Figure 3.1. For a given grid point, a single correlation value for each station (s1-3) is added to the Csum. The value to be added depends on the distance (d1-6) of a given station from the grid point.

3.5 Grid

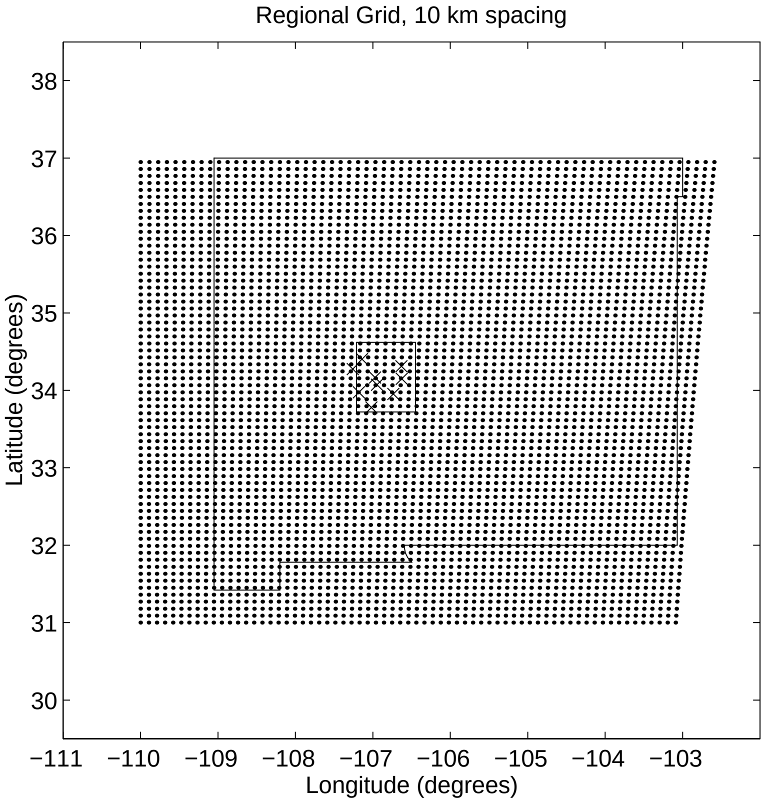


Figure 3.16 Regional Grid using a 10 km grid spacing. The method of using the western most meridian, and southern most parallel as fixed lines of reference, along with requiring a square equally spaced grid, becomes apparent if we plot meridians as straight lines. This is why the right edge curves toward the east. The box in the center of the grid is the nominal outline of the local grid and the crosses are the locations of the receivers used in this study.

Two grids are employed: a coarse grid covering an area slightly larger than the state of New Mexico (Figure 3.16) and a fine grid approximately encompassing the area of the Socorro Seismic Anomaly. The coarse grid is searched first. If the maximum correlation value on this grid exceeds the detection threshold and has an associated grid point within the local area, a fine grid search covering a smaller, local area is performed at greater resolution. If the coarse location is outside the fine grid, the event is declared a regional event, and the coarse location is used as the final hypocenter. The coarse search currently assumes event depths of 5 km and uses a grid spacing of 10 km. The fine search currently uses depths of 3, 6, and 9 kilometers (the seismogenic zone in the Socorro Seismic Anomaly is roughly 3 to 10 kilometers; Balch et al., 1997) and a grid spacing of 1 km.

A recursive procedure for calculating the grids is employed using the Andoyer-Lambert approximation to first order flattening as described by Thomas (1965). The four corners of the grid are specified in degrees and the grid spacing is specified in kilometers. First, a point is determined that is a given distance (specified by the grid spacing) due north of the southwest corner. Next, the point directly north of the previously calculated point is computed, continuing until the maximum latitude has been reached. In this way lines of latitude are determined that are all separated by the grid spacing and lie on the western most meridian of the grid. For each line of latitude, the grid is filled by determining the coordinates for the point which is the given distance due east of the previous point. This continues for each line of latitude until the grid is completely filled in. This procedure provides nodes which form squares of equal area, but is not effective for grids covering the entire globe since the eastern edge of the grid is designed to keep the entire grid rectangular in a cartesian sense, rather than follow any particular meridian.

3.6 Program Flow

LWCEDS can be divided into two distinct sets of programs: Static programs (top of Figure 3.17) and Location programs (bottom of Figure 3.17). All programs are called from the command line and must have a parameter file as a command line argument. In addition to the parameter file, individual parameters may be overridden from the command line. There are approximately seventy available parameters which are explained in Appendix I. Changing some parameters (such as grid size or included phases) will require generating new static files, while others (such as decimation rate, or detection threshold) will not. In any case, the parameter file allows the program to be flexible without recompiling and, since most memory is allocated dynamically, parameter changes will rarely cause memory faults (unless there is insufficient memory available).

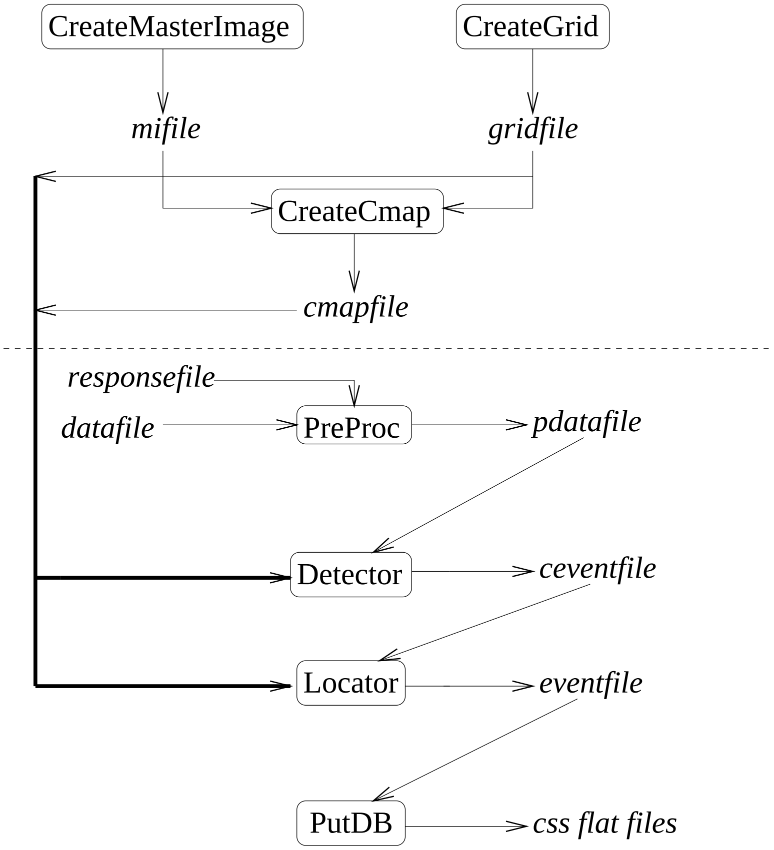


Figure 3.17 Program Flow Diagram illustrating the connectivity of the main programs. Intermediate files are written in binary and italicized in the Figure. Programs above the dashed line need only be run when parameters are changed. Programs below the dashed line are run once for each triggered datastream.

The static programs (*CreateMasterImage*, *CreateGrid*, and *CreateCmap* shown in Figure 3.17) are called individually and generate the static files discussed previously.

They need only be called once for each parameter file. Their output are binary files that are used by the location programs. The three static files are the Master Image, Grid, and Cmap.

The location programs consist of four separate main programs (*PreProc*, *Detector*, *Locator*, and *PutDB* shown in Figure 3.17). They can be called from the command line but to ensure setting the parameters correctly, it is better to call them from a shell script (an example script is provided in Appendix I.8). Each of the four programs are run once for every trigger from the data logger. *PreProc* reads the raw data and the impulse response file of the filter, processes the data, and writes it to a binary processed data file. *Detector* reads the processed data file, Master Image file, grid and Cmap files, performs the coarse correlation and grid search, then writes the coarse location to an intermediate location file. *Locator* reads the processed data, Master Image, grid and Cmap files, and the intermediate location file. Then, if the coarse location is within the fine grid, it performs the fine correlation and grid search, and writes the results to the final event file. If the coarse location is outside the fine grid, then *Locator* will use the coarse hypocenter determined by *Detector* as the final location to write to the file. *PutDB* reads the raw data file headers, the final event file, and the database tables, and places the relevant information into the CSS3.0 flat-file database (see Appendix II). A more complete description of each program is provided in Appendix I.

4. Analysis

4.1 Primary Test Set

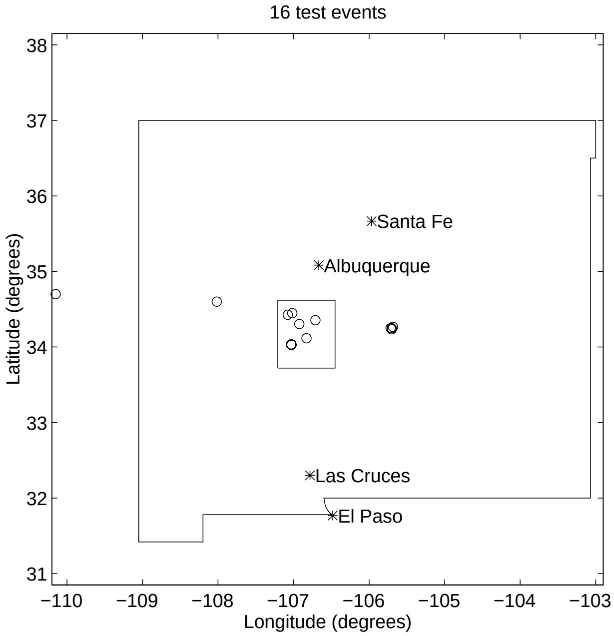


Figure 4.1 Primary Test Set Locations showing the epicenters of the 16 events in the primary test set. Phase arrivals were manually picked by analysts at New Mexico Tech and the events were then located with the generalized least squares program, *Seismos*. The rectangle in the center contains the nominal area of the local grid.

A data set of 16 events was established for primary testing. These events were selected to span a reasonable range of expected event quality and location. Their duration magnitudes (Ake, 1983) range between 0.2 and 3.4. Both local and regional swarms were included as a consistency check. Local events well within the network and near the edge were included, as well as a distant regional event. The earthquakes were originally located by NM Tech personnel using the program *Seismos* (Hartse, 1991). The *Seismos* epicenters are shown in Figure 4.1, and the hypocenters (regionals have underlined evid and 0.0 depth) are listed in Table 4.1. The *LWCEDS* hypocenters are listed in Table 4.2, and the differences are listed in Table 4.3. The event identification number (evid) is a unique tag to aid in selecting a specific event.

Seismos Test Data Hypocenters							
Evid	Latitude	Longitude	Depth	Time	Md	R	RMS
1	34.1158	-106.8282	5.89	Sep 20, 1995 09:10:36.16	0.84	0.869	0.218
2	34.3043	-106.9240	6.35	Sep 22, 1995 10:37:57.84	0.84	1.328	0.362
3	34.3550	-106.7097	8.71	Sep 22, 1995 22:45:56.25	1.35	0.914	0.265
4	34.0333	-107.0288	7.56	Sep 24, 1995 16:04:51.57	1.37	1.415	0.369
5	34.0345	-107.0328	9.22	Sep 24, 1995 16:06:44.74	0.16	0.890	0.242
6	34.0323	-107.0318	8.89	Sep 24, 1995 16:11:25.43	0.33	0.887	0.234
7	34.0297	-107.0282	7.81	Sep 24, 1995 17:58:40.95	1.57	1.225	0.348
8	34.0285	-107.0283	8.80	Sep 24, 1995 18:10:55.02	0.59	0.750	0.202
9	34.4477	-107.0180	4.93	Oct 14, 1995 04:24:03.87	1.43	0.924	0.191
10	34.2667	-105.6833	0.00	Jul 22, 1996 10:06:14.00	3.40	0.976	0.728
11	34.2500	-105.7167	0.00	Jul 22, 1996 10:13:37.00	2.30	1.068	0.799
12	34.2333	-105.7000	0.00	Jul 22, 1996 11:09:32.00	1.60	1.106	0.704
13	34.2500	-105.7000	0.00	Jul 22, 1996 11:13:41.00	1.30	1.011	0.683
14	34.6000	-108.0167	0.00	Aug 01, 1996 16:33:32.00	0.80	0.709	0.347
15	34.7000	-110.1500	0.00	Aug 25, 1996 13:58:45.00	1.80	0.338	0.307
16	34.4275	-107.0768	2.08	Sep 27, 1996 08:35:33.71	1.32	0.954	0.392

Table 4.1 Primary Test Set Seismos Results

LWCEDS Test Data Hypocenters					
Evid	Latitude	Longitude	Depth	Time	Csum
1	34.1346	-106.8415	6.00	Sep 20, 1995 9:10:36.28	16.8
2	34.3058	-106.9168	3.00	Sep 22, 1995 10:37:58.09	13.0
3	34.3599	-106.7101	3.00	Sep 22, 1995 22:45:56.77	19.6
4	34.0354	-107.0260	9.00	Sep 24, 1995 16:04:51.80	29.9
5	34.0445	-107.0259	9.00	Sep 24, 1995 16:06:45.04	14.8
6	34.0445	-107.0259	9.00	Sep 24, 1995 16:11:25.78	18.2
7	34.0354	-107.0260	9.00	Sep 24, 1995 17:58:41.16	31.9
8	34.0354	-107.0260	9.00	Sep 24, 1995 18:10:55.37	21.4
9	34.4770	-107.0032	3.00	Oct 14, 1995 4:24:03.67	24.9
10	34.3355	-105.7628	5.00	Jul 22, 1996 10:06:15.06	8.4
11	34.2427	-106.7434	6.00	Jul 22, 1996 10:13:48.07	30.1
12	34.4257	-105.8671	5.00	Jul 22, 1996 11:09:33.10	7.4
13	34.3355	-105.6542	5.00	Jul 22, 1996 11:13:38.58	5.7
14	34.6059	-108.1471	5.00	Aug 01, 1996 16:33:29.86	3.8
15	33.2539	-109.3563	5.00	Aug 25, 1996 13:58:54.10	2.5
16	34.4500	-107.0795	3.00	Sep 27, 1996 8:35:33.68	24.8

Table 4.2 Primary Test Set LWCEDS Results

LWCEDS/Seismos Test Data Hypocenter Differences						
Evid	Epicentral Difference	Depth Difference	Time Difference	Md	R	Csum
1	2.42	-0.11	-0.13	0.84	0.869	16.8
2	0.69	3.35	-0.25	0.84	1.328	13.0
3	0.54	5.71	-0.52	1.35	0.914	19.6
4	0.35	-1.44	-0.23	1.37	1.415	29.9
5	1.27	0.22	-0.30	0.16	0.890	14.8
6	1.45	-0.11	-0.36	0.33	0.887	18.2
7	0.67	-1.19	-0.21	1.57	1.225	31.9
8	0.80	-0.20	-0.35	0.59	0.750	21.4
9	3.53	1.93	0.20	1.43	0.924	24.9
10	10.58	-5.00	-1.06	3.40	0.976	8.4
11	94.60	-6.00	-11.07	2.30	1.068	30.1
12	26.31	-5.00	-1.10	1.60	1.106	7.4
13	10.39	-5.00	2.42	1.30	1.011	5.7
14	11.98	-5.00	2.14	0.80	0.709	3.8
15	176.42	-5.00	-9.10	1.80	0.338	2.5
16	2.51	-0.92	0.03	1.32	0.954	24.8

Table 4.3 Primary Test Set Difference

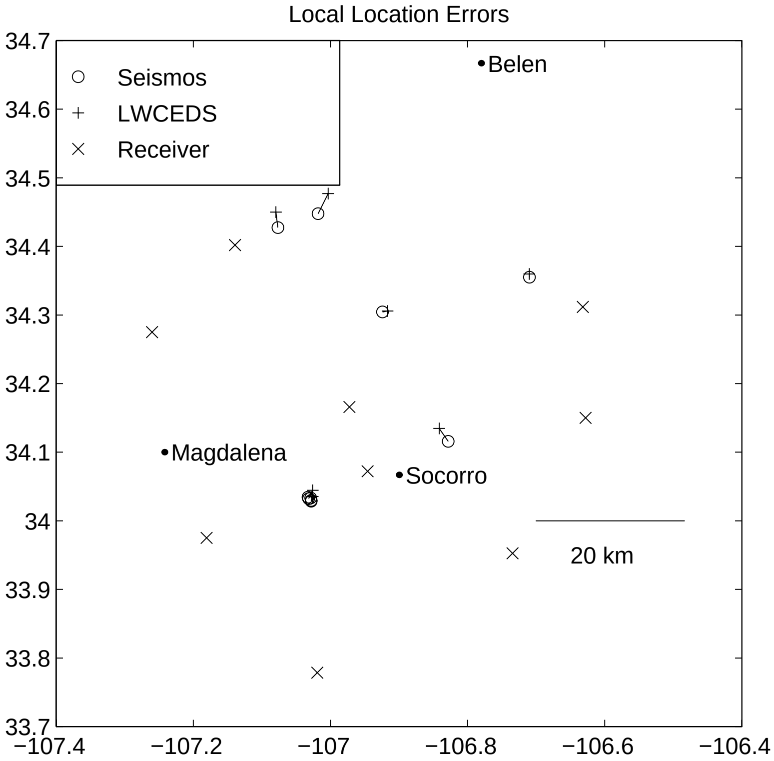


Figure 4.2 Primary Test Set Local Errors comparing epicenters generated by *Seismos* (circles) with epicenters computed using *LWCEDS* (crosses). Station locations are shown with x's. Generally the epicenters from each location method agree to within 2 to 3 km.

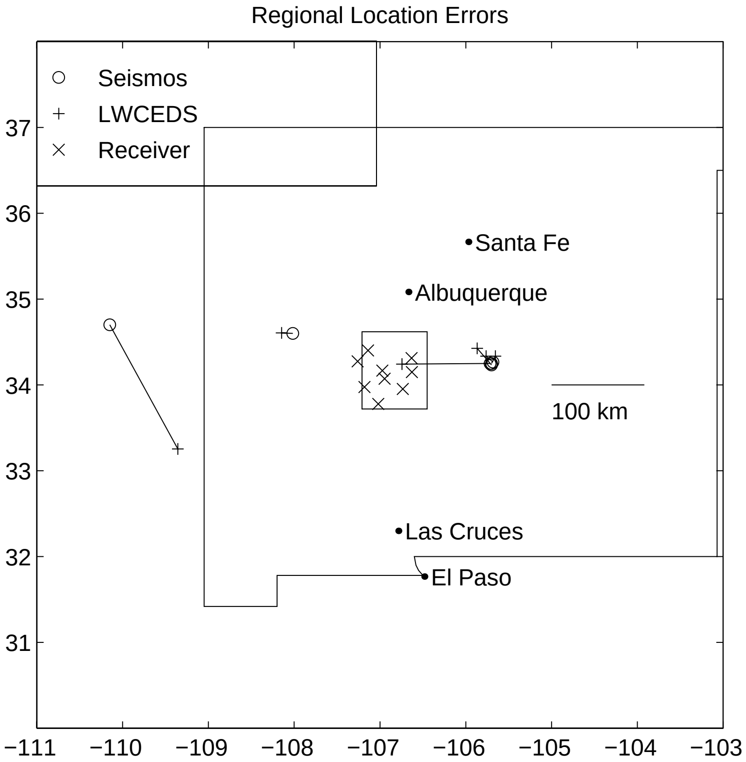


Figure 4.3 Primary Test Set Regional Errors comparing epicenters generated by *Seismos* (circles) with epicenters computed using *LWCEDS* (crosses). Station locations are shown with x's and the nominal local grid is depicted by the rectangle in the center. Generally the epicenters from each location method agree to within 10 to 20 km except for the eastern Arizona event which has poor azimuthal control on the solution, and a Gallinas event which was "pulled" into the local grid by the plane weighting applied to the master image.

In general, *LWCEDS* generates epicenters which are within two or three kilometers (1.4 km mean difference and 1.05 km standard deviation) of the local *Seismos* epicenters (Figure 4.2), and within ten or twenty kilometers (55 km mean difference and 67 km standard deviation) of the regional *Seismos* epicenters (Figure 4.3).

For clarity throughout the rest of this document, all plots of raw data streams have been trace-normalized for amplitude, and all plots of processed data streams have been plot-normalized for amplitude.

The regional event of August 25, 1996 (evid=15) differs from the *Seismos* solution by 176 km. Figure 4.4 is a *MATSEIS* plot with the raw data (top) plotted at the appropriate distance for the *LWCEDS* epicenter, and with the travel time curves overlain. Figure 4.4 (middle) shows the processed data with respect to distance from the *LWCEDS* hypocenter. The processed data are plotted with respect to the *Seismos* hypocenter in Figure 4.4 (bottom). This is an m_d ~1.8 event from eastern Arizona, and was specifically chosen for the large distance (~300 km) from the network. The distance is reasonably well constrained, but azimuth is poorly constrained. This is further illustrated in Figure 4.5 which shows the correlation value for each coarse grid point. Relatively high correlation occurs in a ring about the network which basically acts as a single station due to the small network aperture relative to the source distance of this event.

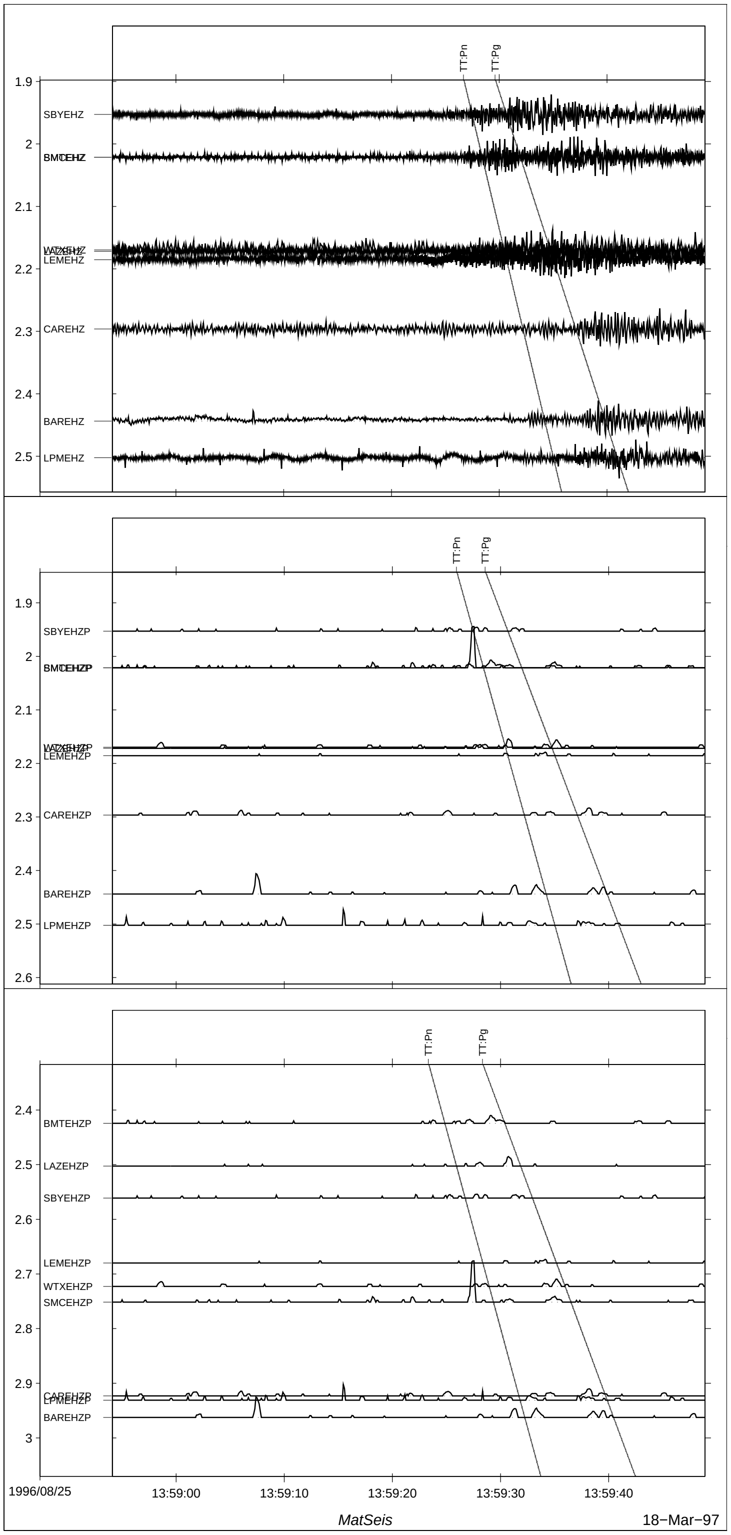


Figure 4.4 Evid 15 Data for an $m_d \sim 1.8$ earthquake from eastern Arizona (Aug 25, 1996 13:58:45 UT) plotted as a function of distance from the *LWCEDS* epicenter for the raw data (top) and processed data (middle). Processed data is also plotted with respect to the *Seismos* epicenter (bottom).

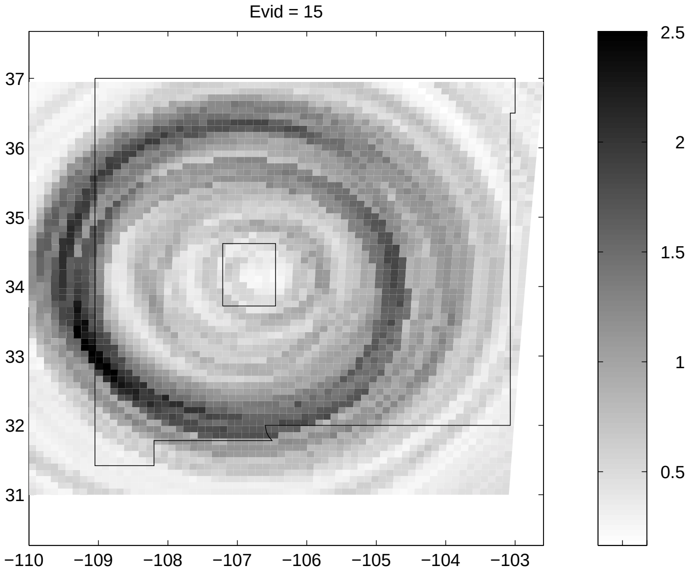


Figure 4.5 Evid 15 Csum Map illustrating the C matrix summation values for each coarse grid point. The poor azimuthal control on the solution is illustrated by the rings of high correlation about each station nearly overlapping to form a single ring. The stations used in the solution are approximately contained within the nominal local grid depicted with the rectangle at the center.

By contrast, the regional event of August 1, 1996 (evid=14) differs from the *Seismos* solution by approximately 12 km. While it is still outside the network, it is much closer than the August 25 event (~100 km). Figure 4.6 (top) shows the raw waveforms overlain on travel time curves at distances appropriate to the *LWCEDS* hypocenter. Figure 4.6 (middle) is a gather of processed waveforms plotted a distance appropriate to the *LWCEDS* hypocenter and Figure 4.6 (bottom) is the processed data with respect to the *Seismos* hypocenter. Figure 4.7 is the correlation value at each grid point, and indicates much better azimuthal control on the epicenter. In this case the epicenter is sufficiently close to the network that correlation rings from individual stations become delineated. The location which best fits the model is where the correlation rings from all the stations overlap. A sense of the error is represented by the area of the dark, high correlation contours. If a method were developed to convert these maps of the correlation values into a quantifiable error estimate, then it may be possible to avoid using error ellipses which in many cases are not an appropriate shape to fully characterize the error.

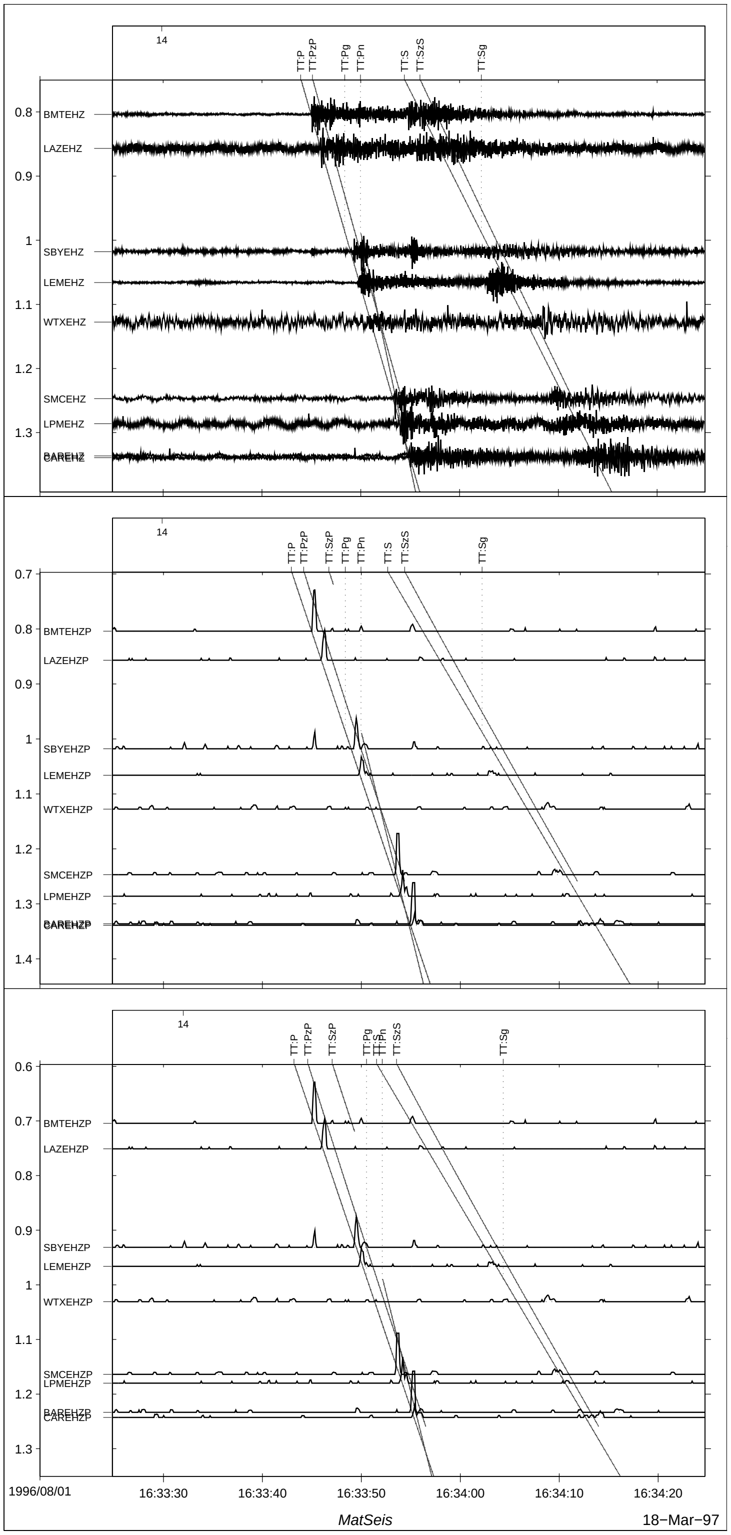


Figure 4.6 Evid 14 Data for an $m_d \sim 0.8$ from north of Pie Town (Aug 1, 1996 16:33:32 UT) plotted as a function of distance from the *LWCEDS* epicenter using the raw (top) and processed (bottom) data. Processed data are also plotted with respect to the *Seismos* epicenter (bottom).

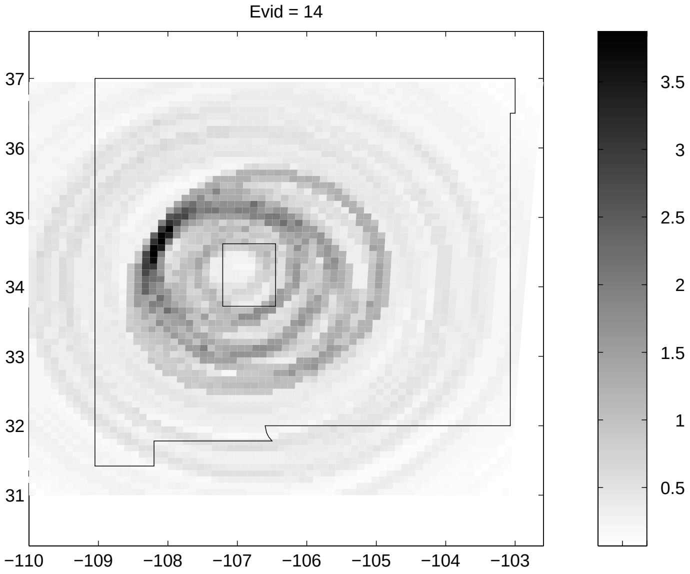


Figure 4.7 Evid 14 Csum Map illustrating the C matrix summation values for each coarse grid point. Azimuthal control on the solution for this regional event is much better than the eastern Arizona event and illustrated by the high correlation rings for each station overlapping in a smaller area.

The local event of September 24, 1995 at approximately 18:11 UT (evid=8) has good agreement between the *Seismos* and *LWCEDS* solutions. This is illustrated in Figure 4.8 (top), which shows the raw waveforms plotted at distances appropriate to the *LWCEDS* epicenter. The processed waveforms are plotted in Figure 4.8 (middle) at distances appropriate to the *LWCEDS* hypocenter, and Figure 4.8 (bottom) is a gather of processed waveforms plotted at distance with respect to the *Seismos* hypocenter. Comparison of Figure 4.8 middle and bottom also illustrates the constant time lag of the *LWCEDS* hypocenters which have multiple sources. First, to reduce processing, the data on which *LWCEDS* operates has been decimated from 100 sps to 10 sps so errors of a tenth of a second are to be expected. Second, the master image envelopes are centered on the predicted travel time curves. This, combined with the finite rise time of the STA/LTA, falsely shifts the *LWCEDS* origins backward in time. The easiest fix for this would be to force the master image envelopes to begin at or after the predicted travel time curve. Alternatively, providing a static offset to the data could account for the rise time of the STA/LTA but the amount of the offset will depend on the window length which is not constant. Another, more complicated, but perhaps more robust correction would be to convolve the travel time curve with an appropriate response function and then have the adaptive STA/LTA operate on these synthetic data streams producing a master image which looks more like the expected processed data. An empirical master image would also correct the problem assuming processing is consistent between the Master Image and data matrices.

The correlation values for the September 24 local, Figure 4.9, are well constrained for the coarse grid in both distance and azimuth. Figure 4.10 is a map of the correlation values at the local scale where the seismic stations are the x's, The most prominent rings of high correlation values are concentric about the stations with the largest SNR and these rings overlap at the *LWCEDS* location (small white circle southwest of Socorro). There are also other, lower correlation, rings about each station which are associated with observed arrivals at each station aligning incorrectly with theoretical arrivals at various distances. The false-correlation rings do not have contributions from as many stations as the true correlation arrangement and thus sum to a smaller overall value.

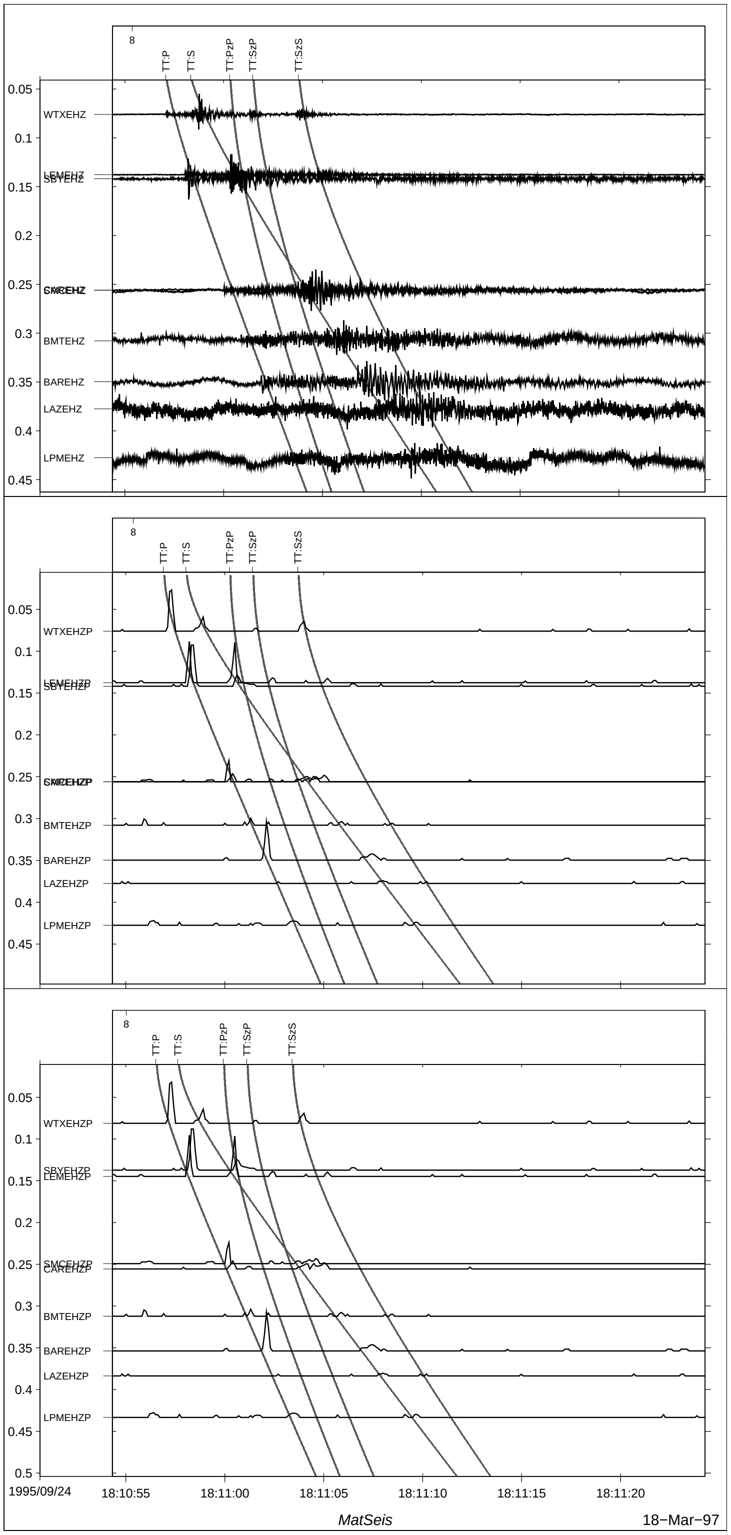


Figure 4.8 Evid 8 Data for an $m_d \sim 0.6$ from southwest of Socorro (Sep 24, 1996 18:10:55 UT) plotted as a function of distance from the *LWCEDS* epicenter using raw data (top) and processed data (middle). The processed data is also plotted as a function of distance from the *Seismos* epicenter (bottom).

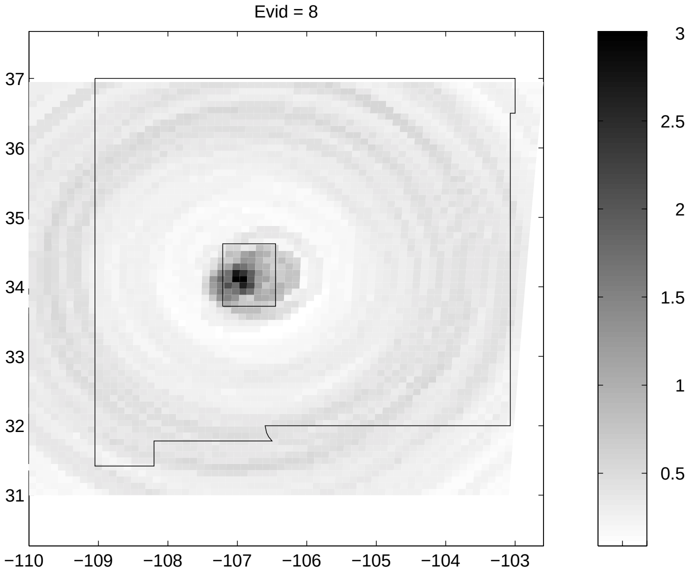


Figure 4.9 Evid 8 Csum Map illustrating the C matrix summation values for each coarse grid point. At this scale, the rings of high correlation for this local event appear as a point within the network. The nominal local grid is contained within the central rectangle.

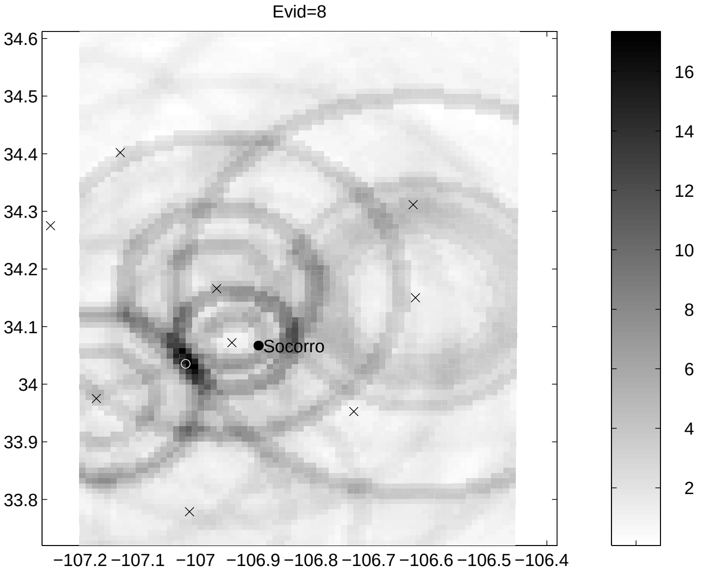


Figure 4.10 Evid 8 Local Csum Map illustrating the C matrix summation for each fine grid point. At this scale, the most prominent rings of high correlation are concentric about the stations which recorded the largest SNR signals and these rings overlap at the *LWCEDS* location (small white circle southwest of Socorro). Station locations are plotted as the X's.

The event of July 22, 1996 at about 10:13 UT (evid=11) is part of a swarm near Gallinas. Figure 4.11 (top) shows the raw waveforms for the *LWCEDS* hypocenter. Processed waveforms are shown in Figure 4.11 (middle) for the *LWCEDS* location and shown in Figure 4.11 (bottom) for the *Seismos* hypocenter. This event has been mislocated by the coarse search, and because the coarse location is within the local grid, the fine search was performed. No reflections are included in the coarse master image, yet the waveform alignment illustrated in Figure 4.11 (middle) yields a larger correlation sum than the alignment in Figure 4.11 (bottom). This is true because of the plane weighting applied to the master image (see Figure 3.9), which reduces the weight of the master image with increasing distance and travel time. The processed data represents signal to noise ratio and the plane is applied to more closely model expected signal attenuation which is not observed for this event. Indeed, the location of this event can be improved by applying a different weighting plane to the master image, but not without deleterious effects on locations for other events. The maximum correlation values as a function of time are shown in Figure 4.12 with plane weighting applied to the master image (top) and without (bottom). Particularly at regional distances and with events that have only a single defining phase, only slight changes in the maximum correlation can dramatically change the automatic location which is why the time-distance plane weighting is applied. Only P, S, Pn, Pg, and Sn are used in the coarse Master Image and there are a number of peaks corresponding to various alignments of these theoretical arrivals with the observed arrivals. Times 1 and 2 in Figure 4.13, within about 10 s of the trace start, are alignments of observed P with theoretical S. As the model is stepped through the data, observed P arrivals begin aligning for various locations with theoretical P near the correct origin time (3 in the bottom of figure 4.12) at about 20 seconds. Later sufficient observed P arrivals align with theoretical P at small travel distances to produce a peak in the correlation at time 4 which corresponds with a location within the aperture of the network. Note also in the bottom of Figure 4.13 that both origin times at 1 and 2 yield a higher correlation value than the "true" origin time at 3. This is a result of locating an event with essentially one phase (Pn). The STA/LTA is so large across all stations that the difference in phase weighting between P and S is relatively minor. For this event to be properly located automatically, an attenuation plane somewhere between 0 and the current plane should be used. In this case an empirical Master Image might be more appropriate because it should better model signal attenuation than plane weighting. Not surprisingly, the map of the C summation values for the *LWCEDS* location (time 4 in Figure 4.12) are shown in Figure 4.13 and look similar to maps for local events. The map of Csum values for an earlier origin time (Figure 4.14) looks more like the maps for the other regional events. The outer high correlation ring represents the false correlation at time 1 from Figure 4.13 and the ring just inside the outer one is the false correlation at time 2. The budding inner ring is the "true" correlation at time 3 and will contain higher values at the true origin time (though not the highest without the appropriate weighting). In fact, a movie of C summation maps would begin with a large circle at large distance that slowly contracts about the network as time advances. As the radius of the circle approaches the aperture of the network, the single circle will diverge into a circle for each station. These station specific circles will continue to contract about the appropriate station, peak, then rapidly disappear at times dependent on the first arrival at each station.

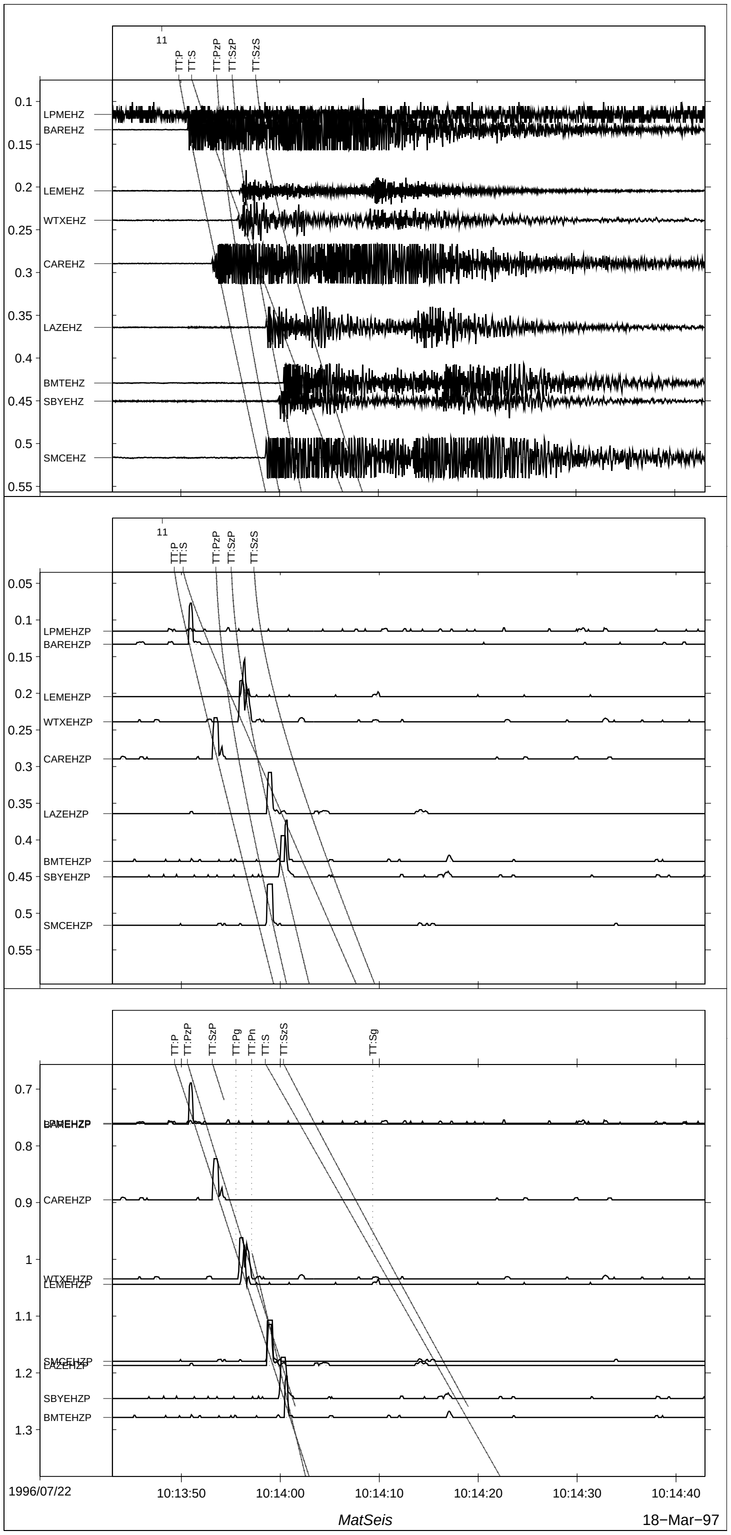


Figure 4.11 Evid 11 Data for an $m_d \sim 2.3$ near Gallinas (July 22, 1996 10:13:37 UT) plotted as a function of distance from the *LWCEDS* epicenter using both raw (top) and processed (middle) data. The processed data are also plotted with respect to the *Seismos* epicenter (bottom).

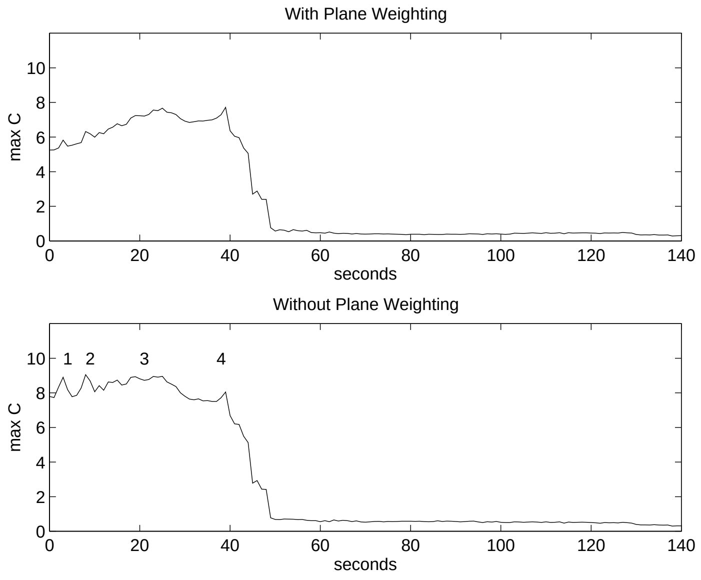


Figure 4.12 Maximum Correlation as Function of Time with plane weighting applied to the master image (top) and without (bottom). Because the plane weighting is time dependent, later origin times are given more weight because the waves have shorter travel times. Removing the plane has a deleterious effect on estimating the best origin time. At times 1 and 2, observed P is aligned with theoretical S (2 is aligned at a closer distance). Time 3 (the "true" origin time) is where observed P is aligning with various combinations of theoretical Pn, and Pg. At time 4, observed P arrivals are able to align with sufficient theoretical P arrivals at a point within the aperture of the network that a peak is produced in the correlation.

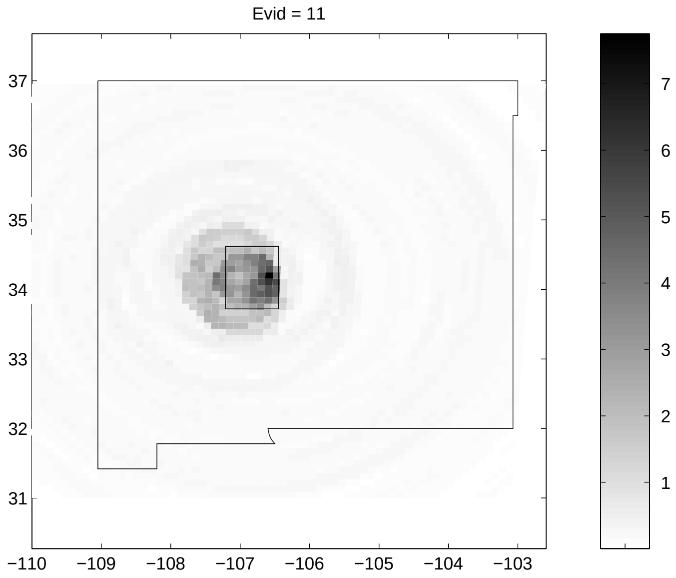


Figure 4.13 Evid 11 Csum Map illustrating the C matrix summation values for each coarse grid point. This Gallinas regional produced a large correlation value within the local grid (central rectangle).

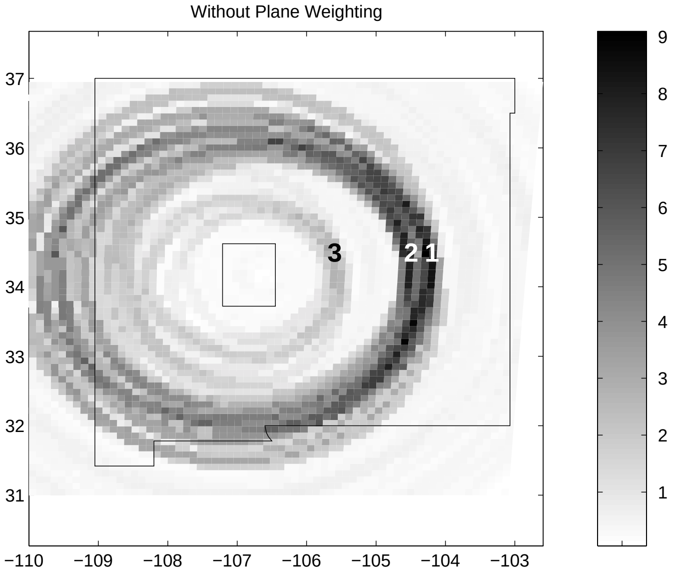


Figure 4.14 Evid 11 Csum Map Without Plane Weighting still mislocates this Gallinas event at point 1 (corresponds with time 1 from figure 4.12). The "true" epicenter is at point 3.

The map of the maximum Correlation value does not directly represent error in the epicenter location, but it is closely related to how well the processed data for a given location fit the model. If the maximum Correlation value (Csum) for each of the test events is plotted as a function of the epicentral difference between the *LWCEDS* and *Seismos* epicenters (Figure 4.15), there does not appear to be a trend of decreasing Csum with increasing epicentral difference. There are two groups in Figure 4.15 which correspond with local (x's) and regional events (o's). In general, the local events have higher Csum values because there are more phases in the master image which provide more opportunities for non-zero summations. The outlier at approximately 100 km epicentral difference is the Gallinas regional which located as a local. This event has a large Csum value because the p-arrivals align well with the local phases. Because of the apparent lack of a relationship between the Csum values and the location errors, the Csum value should not be construed as an appropriate measure of error. But a map of the Csum values as shown in Figure 4.12 does give an indication of the variation in the Correlation with changes in the assumed hypocenter.

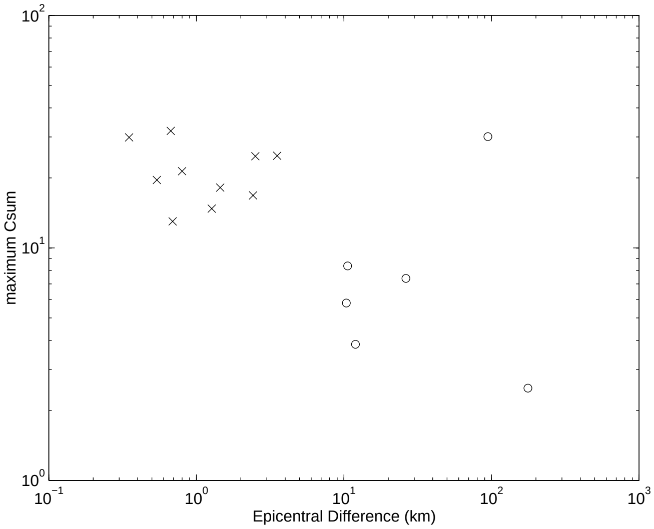


Figure 4.15 Max Correlation vs. Error showing the lack of dependence of the maximum summation through C on epicentral difference between *Seismos* and *LWCEDS* locations. Two distinct groups (Local x's, Regional o's) are formed and the data are scattered within these groups. Plot is log-log.

To quickly summarize, of the 10 local events in the primary test set, 90% were automatically located to within 3 km of the *Seismos* epicenter. One local event has a *LWCEDS* to *Seismos* difference of 3.53 km and this event is on the edge of the aperture of the network. Of the 6 regional events in the set, 66 % were automatically located to within 30 km of the *Seismos* epicenters. Of the remaining 33 %, one regional had poor azimuthal control due to lack of station coverage and another had only p-phases and was mislocated due to crude modeling of attenuation.

4.2 Expanded Test Set

To test the performance of operating *LWCEDS* online, a cron job was initiated to operate on the data which are transferred daily from the PC data logger to the local unix network of workstations. Testing began on December 3, 1996 and continued through January 7, 1997. During this time 156 triggers were recorded by the data logger and event type classified by seismic analysts. Of these triggers, 33 were teleseisms, 102 were regional events (including explosions), and 21 were local earthquakes. The 21 local events were located using traditional means (i.e., picks made using Xpick, and entered into *Seismos*) to determine "ground truth" (Table 4.4) for comparison with the automatic locations (Table 4.5). While the events in the primary data set were culled from a data set generated for another study, and were known to have clean time series with well constrained locations, membership in the expanded data set was governed only by events which happened to occur during the time period of the online experiment. Consequently, the expanded data set is, in some cases, not as well located due to poor signal to noise ratio and contamination by electronic spikes.

Seismos Test Data Hypocenters							
Evid	Latitude	Longitude	Depth	Time	Md	R	RMS
15	34.1527	-106.8538	6.25	Dec 6,1996 3:46:29.55	0.4	0.969	0.305
22	34.0183	-107.0547	1.84	Dec 8,1996 6:34:09.29	-0.5	0.684	0.209
39	33.9827	-106.9538	1.37	Dec 11,1996 20:33:34.04	0.5	0.734	0.192
56	34.1630	-107.0173	2.46	Dec 14,1996 18:53:42.07	1.0	1.077	0.286
59	33.9477	-106.9183	5.00	Dec 16,1996 15:38:25.34	0.9	1.465	0.640
64	34.0488	-106.9312	10.33	Dec 17,1996 7:38:18.22	-0.3	0.729	0.155
65	34.0415	-106.9208	6.68	Dec 17,1996 11:56:44.89	-0.2	1.694	0.431
85	34.3495	-106.7088	8.57	Dec 21,1996 18:57:03.98	0.4	1.370	0.328
89	34.3753	-106.9947	5.00	Dec 23,1996 4:02:15.72	-0.1	1.597	0.487
92	34.1637	-107.0178	3.38	Dec 24,1996 4:11:02.39	1.0	1.067	0.228
93	34.1598	-107.0108	4.87	Dec 24,1996 4:25:47.96	0.2	1.326	0.364
98	34.1777	-106.9800	4.90	Dec 24,1996 20:19:11.28	0.0	1.461	0.271
99	34.1102	-106.6852	4.53	Dec 25,1996 4:59:44.15	0.5	0.760	0.207
100	34.3838	-107.0307	6.42	Dec 25,1996 12:55:42.00	0.4	0.778	0.172
111	34.2282	-106.8978	8.06	Dec 28,1996 21:58:58.46	0.2	2.052	0.729
113	34.0378	-106.9965	5.56	Dec 30,1996 9:32:14.26	-0.1	0.369	0.088
114	34.1585	-106.8530	8.08	Dec 30,1996 9:57:02.04	0.6	1.491	0.582
120	34.1698	-106.8497	7.56	Dec 31,1996 6:14:06.21	0.3	1.117	0.381
133	34.3378	-106.7257	2.51	Jan 2,1997 6:03:15.18	0.6	1.531	0.593
141	34.1848	-106.8258	6.61	Jan 3,1997 8:47:11.54	0.3	0.801	0.289
151	34.1848	-106.8230	5.37	Jan 7,1997 5:38:16.54	0.2	0.391	0.117

Table 4.4 Expanded Test Set Seismos Results

LWCEDS Test Data Hypocenters					
Evid	Latitude	Longitude	Depth	Time	Csum
15	34.1616	-106.8522	6.00	Dec 6,1996 3:46:29.81	13.25
22	35.1466	-108.2446	5.00	Dec 8,1996 6:33:46.76	1.90
39	33.9994	-106.9395	6.00	Dec 11,1996 20:33:34.27	24.42
56	34.1706	-107.0040	3.00	Dec 14,1996 18:53:42.29	33.62
59	33.9543	-106.9072	6.00	Dec 16,1996 15:38:25.65	28.87
64	34.0535	-106.9176	9.00	Dec 17,1996 7:38:18.59	18.24
65	34.0535	-106.9176	9.00	Dec 17,1996 11:56:44.86	23.64
85	34.3148	-106.6235	3.00	Dec 21,1996 18:57:05.43	13.01
89	34.4049	-107.1339	3.00	Dec 23,1996 4:03:38.81	12.95
92	34.1706	-107.0040	3.00	Dec 24,1996 4:11:02.70	27.06
93	34.1706	-106.9931	3.00	Dec 24,1996 4:25:48.35	16.71
98	34.1706	-106.9931	3.00	Dec 24,1996 20:19:11.50	10.36
99	34.1075	-106.6358	9.00	Dec 25,1996 4:59:43.99	19.50
100	34.4140	-107.0251	6.00	Dec 25,1996 12:55:42.07	15.09
111	34.3238	-106.6343	3.00	Dec 28,1996 21:58:48.98	12.12
113	34.0445	-106.9826	3.00	Dec 30,1996 9:32:14.72	19.55
114	34.1616	-106.8522	6.00	Dec 30,1996 9:57:02.43	22.01
120	34.1616	-106.9498	3.00	Dec 31,1996 6:14:08.00	13.42
133	34.3599	-106.6666	9.00	Jan 2,1997 6:03:14.87	17.43
141	34.1887	-106.8196	6.00	Jan 3,1997 8:47:11.74	16.86
151	35.0565	-107.5890	5.00	Jan 7,1997 5:38:00.07	2.27

Table 4.5 Expanded Test Set LWCEDS Results

LWCEDS/Seismos Test Data Hypocenter Differences						
Evid	Epicentral Difference	Depth Difference	Time Difference	Md	R	Csum
15	1.00	0.25	-0.26	0.4	0.969	13.25
22	166.13	-3.16	22.53	-0.5	0.684	1.90
39	2.28	-4.63	-0.23	0.5	0.734	24.42
56	1.49	-0.54	-0.22	1.0	1.077	33.62
59	1.26	-1.00	-0.31	0.9	1.465	28.87
64	1.36	1.33	-0.37	-0.3	0.729	18.24
65	1.36	-2.32	0.03	-0.2	1.694	23.64
85	8.74	5.57	-1.45	0.4	1.370	13.01
89	13.22	2.00	-83.09	-0.1	1.597	12.95
92	1.49	0.38	-0.31	1.0	1.067	27.06
93	2.02	1.87	-0.39	0.2	1.326	16.71
98	1.44	1.90	-0.22	0.0	1.461	10.36
99	4.57	-4.47	0.16	0.5	0.760	19.50
100	3.39	0.42	-0.07	0.4	0.778	15.09
111	26.49	5.06	9.48	0.2	2.052	12.12
113	1.48	2.56	-0.46	-0.1	0.369	19.55
114	0.35	2.08	-0.39	0.6	1.491	22.01
120	9.28	4.56	-1.79	0.3	1.117	13.42
133	5.96	-6.49	0.31	0.6	1.531	17.43
141	0.71	0.61	-0.20	0.3	0.801	16.86
151	119.55	0.37	16.47	0.2	0.391	2.27

Table 4.6 Expanded Test Set Difference

4.2.1 Mislocated Events

A few examples of mislocated events follow:

- The event on December 8, 1996 6:34:09 UT (evid=22) has an *LWCEDS* location which differs from the *Seismos* location by 166.13 kilometers. As seen in the raw and processed waveform gathers of Figure 4.16 (top and middle respectively) the LPM data stream is contaminated with positive and negative spikes which have a separation which is a reasonable S-P time for a regional event. These spikes on LPM govern the correlation distance for the event and make it impossible, without somehow filtering the spikes, to automatically locate this event properly. Processed waveforms for the *Seismos* location are shown in Figure 4.16 (bottom).

- The event from December 23, 1996 4:02:15 UT (evid=89) is contaminated by a spike on LAZ about 85 seconds prior to the origin time of the event. This can be seen in the unprocessed traces in Figure 4.17 (top). This is the largest "signal" of any channel so LAZ is forced to correlate as the closest station and the first arrival is essentially ignored. The processed waveforms are displayed in Figure 4.17 (middle) for the *LWCEDS* hypocenter, and in Figure 4.17 (bottom) for the *Seismos* location.

- The event from December 28, 1996 21:58:58 UT (evid=111) is contaminated by unknown impulsive noise on station LPM (Figure 4.18 top) which creates a large peak in the processed data (Figure 4.18 middle) and causes LPM to be the closest station, when in fact, LEM should be the closest station (Figure 4.18 bottom).

- The event of January 7, 1997 5:38:16 UT (evid=151) is mislocated by about 120 kilometers. This is a result of a large noise spike on WTX which has a reasonable "arrival" time for a regional S-P time. WTX thus fixes the distance and forces a mislocation. The raw waveforms, are shown in Figure 4.19 (top). The processed gather is shown in Figure 4.19 (middle) for the *LWCEDS* location and in Figure 4.19 (middle) for the *Seismos* location.

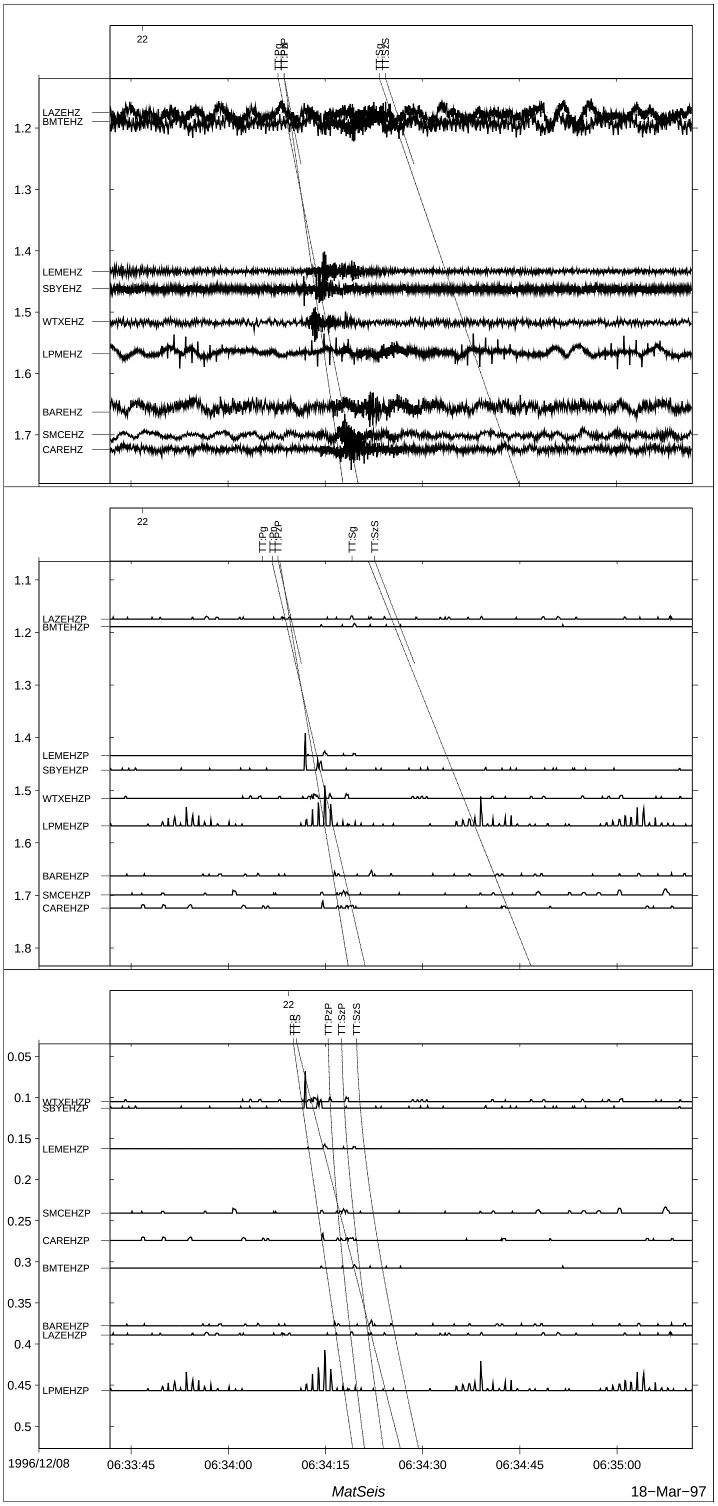


Figure 4.16 Evid 22 Data for an $m_d \sim -0.5$ southwest of Socorro (Dec 8, 1996 6:34:09 UT) plotted as a function of distance from the spike corrupted *LWCEDS* epicenter for the raw (top) and processed data (middle). Processed data are plotted as a function of distance from the *Seismos* epicenter in the bottom gather.

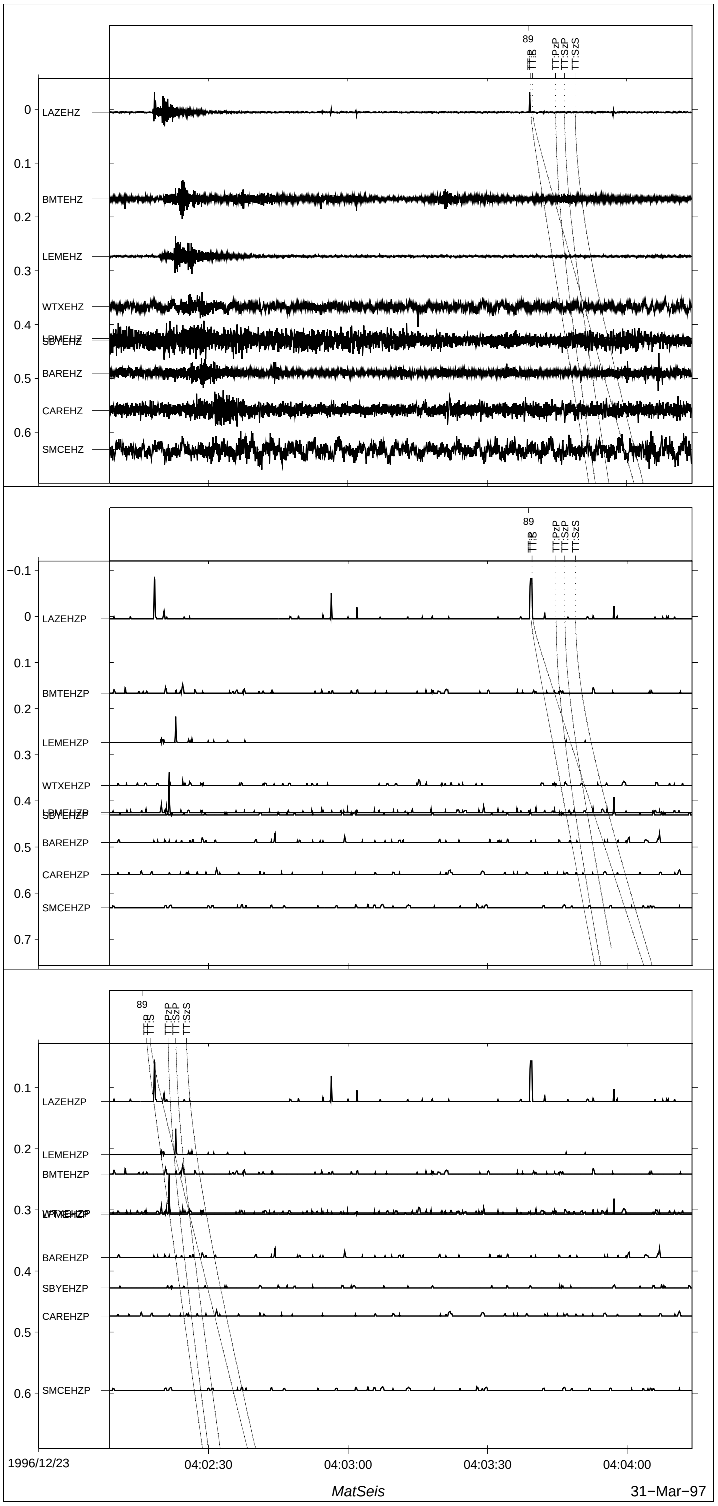


Figure 4.17 Evid 89 Data for an $m_d \sim -0.1$ near Ladrones Peak (Dec 23, 1996 4:02:15 UT) plotted as a function of distance from the spike corrupted *LWCEDS* epicenter for the raw (top) and processed data (middle). Processed data are plotted as a function of distance from the *Seismos* epicenter in the bottom gather.

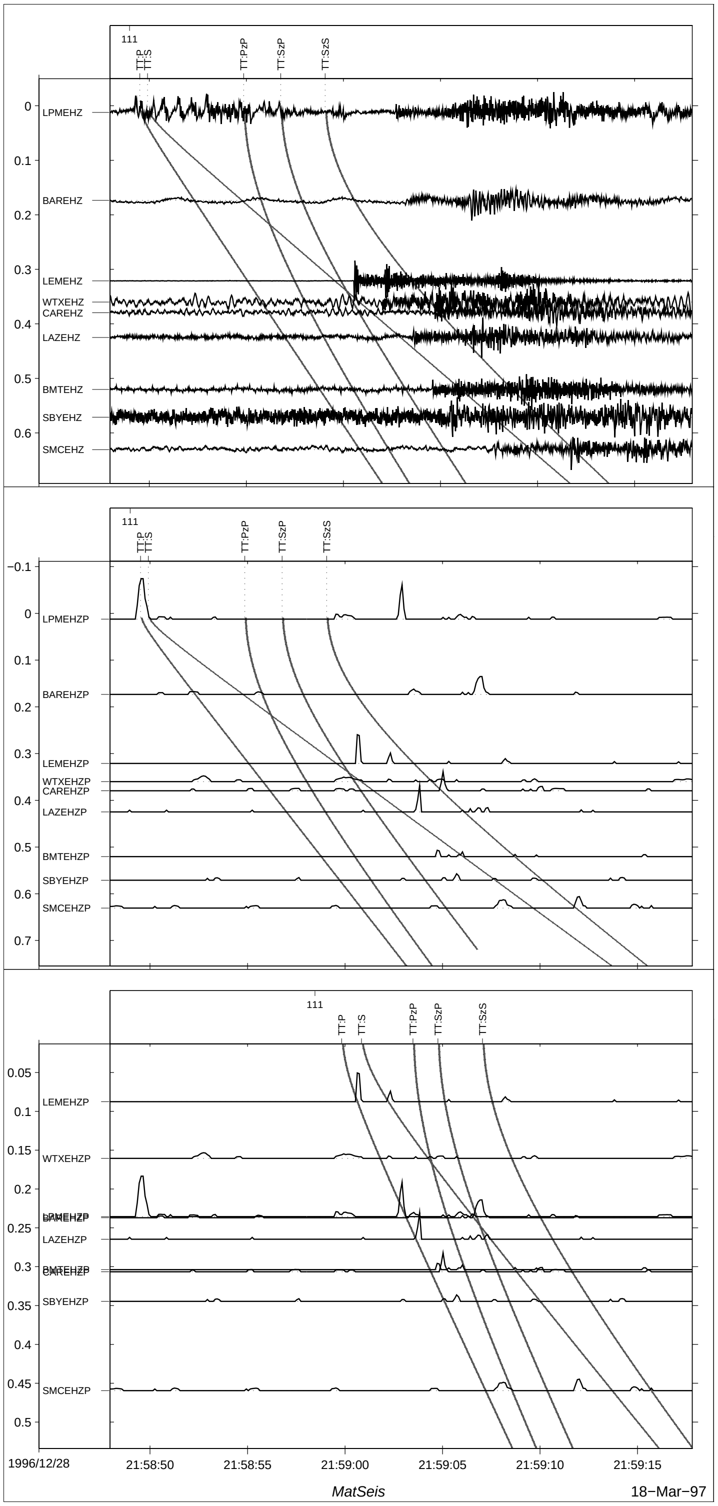


Figure 4.18 Evid 111 Data for an $m_d \sim 0.2$ near Bernardo (Dec 28, 1996 21:58:58 UT) plotted as a function of distance from the noise corrupted *LWCEDS* epicenter for the raw (top) and processed data (middle). Processed data are plotted as a function of distance from the *Seismos* epicenter in the bottom gather.

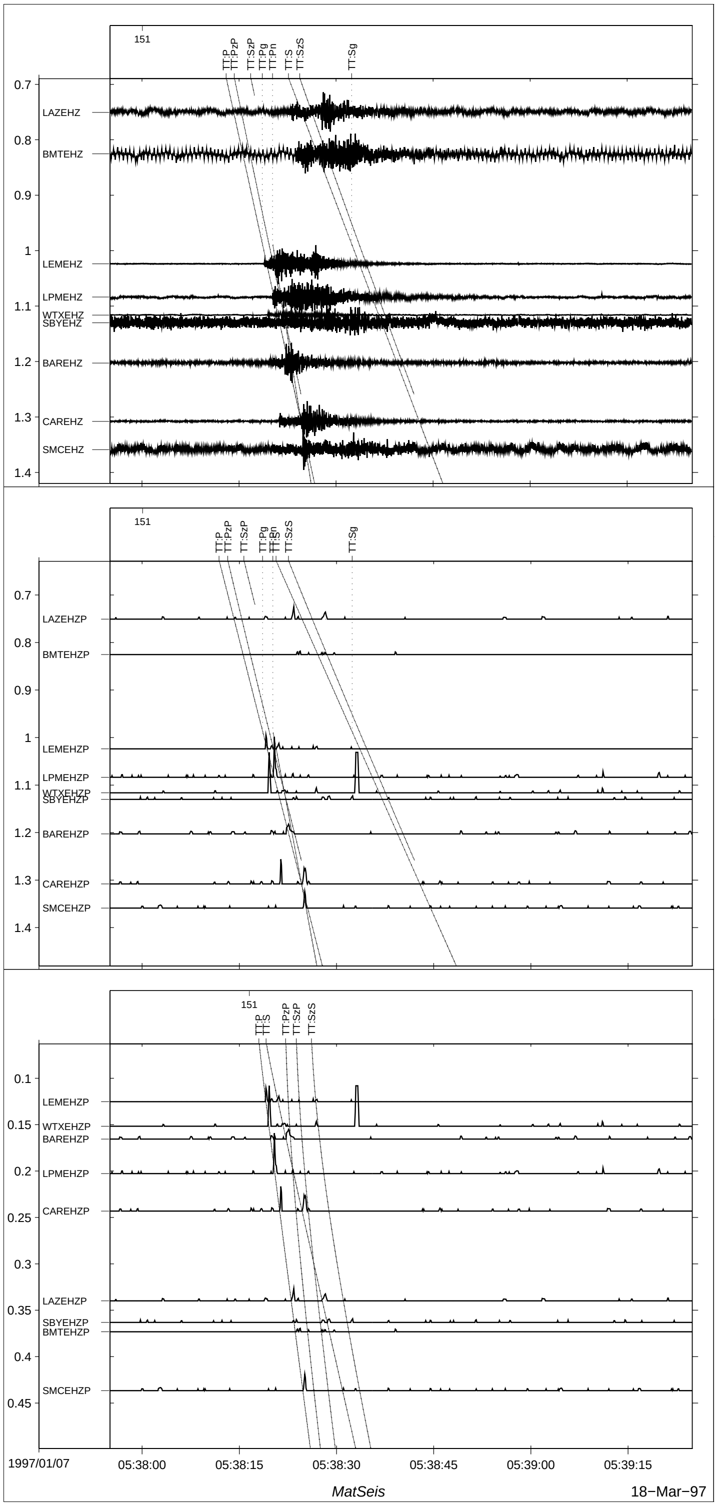


Figure 4.19 Evid 151 Data for an $m_d \sim 0.2$ east of Polvadera (Jan 7, 1997 5:38:16 UT) plotted as a function of distance from the spike corrupted *LWCEDS* epicenter for the raw (top) and processed data (middle). Processed data are plotted as a function of distance from the *Seismos* epicenter in the bottom gather.

4.2.2 Well Located Events

An example of an event that is uncontaminated by the noise problems of the previous examples is the local earthquake of December 24, 1996 20:19:11 UT (evid=98) with md$\sim$0.0 and a rich suite of phases (Figure 4.20, top), This event was well located by *LWCEDS* (differing from *Seismos* by about 1.5 kilometers), is typical of the results for other local events in the test set with uncontaminated data streams, and given development of a "spike" filter, shows the potential success of *LWCEDS*. The *LWCEDS* gather is shown in Figure 4.20 (middle) and the *Seismos* gather is shown in Figure 4.20 (bottom).

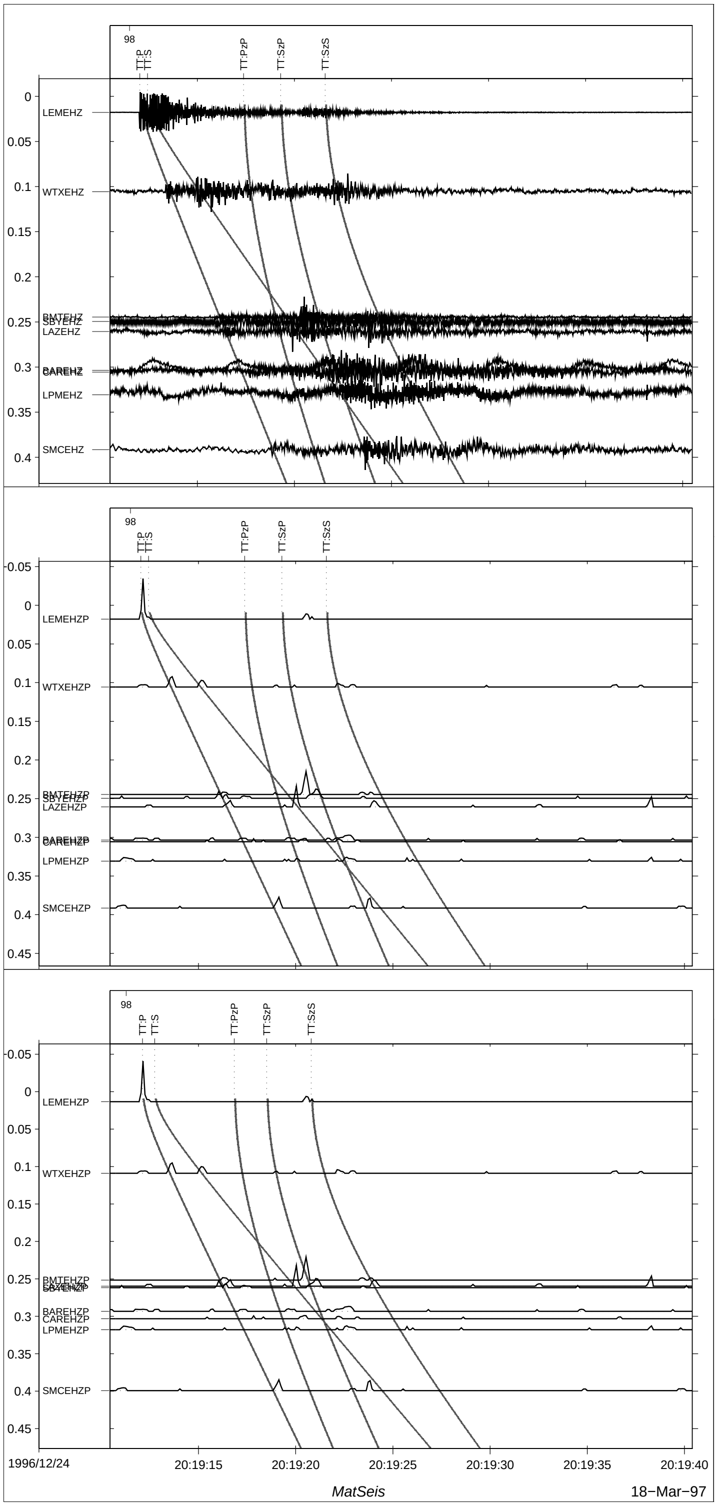


Figure 4.20 Evid 98 Data for an $m_d \sim 0.0$ near Polvadera (Dec 24, 1996 20:19:11 UT) plotted as a function of distance from the well located *LWCEDS* epicenter for the raw (top) and processed (middle) data. Processed data are also shown as a function of distance from the *Seismos* epicenter (bottom).

4.2.3 Depth Estimates

Estimates of source depth for both the primary and expanded test sets have been made with mixed results. Automatic depth estimates are generated by performing the grid search over multiple C matrices that have been calculated using Master Images appropriate for the each depth. *LWCEDS* currently uses fine grid Master Images for depths of 3, 6, and 9 km, and a single coarse grid Master Image for 5 km depth. As a means of examining the variability of *LWCEDS* results with depth, hypocenters and maximum correlation values were calculated for fixed source depths ranging from 0.5 to 9.5 km in 0.5 km increments. As seen in Figure 4.21, the epicenters for the local events in the Primary test set are generally stable with variation in source depth. Station locations are plotted with x's and epicenters are plotted with o's (one epicenter for each event at each depth). Multiple epicenters are superimposed for the swarm west of Socorro but the other clusters are single events where an epicenter has been plotted for each of the 19 fixed depths. All 19 epicenters are not obvious because of the grid based search which is capable of producing identical epicenters.

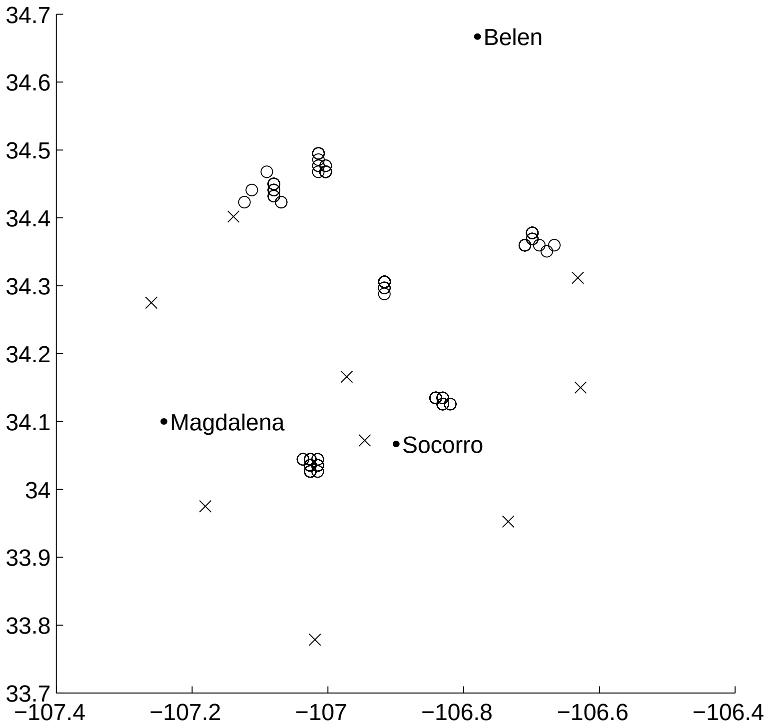


Figure 4.21 Epicenter vs. Depth for the local events in the primary test set. Station locations are delineated with x's and epicenters are plotted with o's. An epicenter for each event is plotted for a range of fixed source depths from 0.5 km to 9.5 km in increments of 0.5 km. Multiple events are superimposed for the swarm west of Socorro but the other clusters are single events with 19 overlapping circles (one for each depth) and illustrate the general stability of the *LWCEDS* epicenter with variable source depth.

As one might expect, the calculated origin time is strongly dependent on the depth of the source. Generally, as the fixed source depth increases, the origin time at which the maximum correlation occurs moves backward in time (i.e. a longer postulated travel distance requires an earlier postulated origin time). This is born out in Figure 4.22 where the relative origin time is plotted as a function of depth for each of the 10 events in the primary test set. To reduce processing, the fine search is only performed over a fixed length time window centered on the origin time from the coarse search, rather than for the entire processed data stream. Currently *LWCEDS* only computes the C matrices for origin times that are within ±6 s of the coarse origin time. The time plotted in Figure 4.22 thus represents time with respect to the coarse grid search hypocenter. A time of 0 is 6 seconds prior to the coarse origin time, of 6 is precisely the coarse origin time, and of 12 is 6 s after the coarse origin time. Figure 4.22 then illustrates that the maximum correlation value generally moves backward in time with increasing fixed source depth (small relative time is early, large is later). Thus there is a large tradeoff between depth and origin time and the epicenter is not significantly affected.

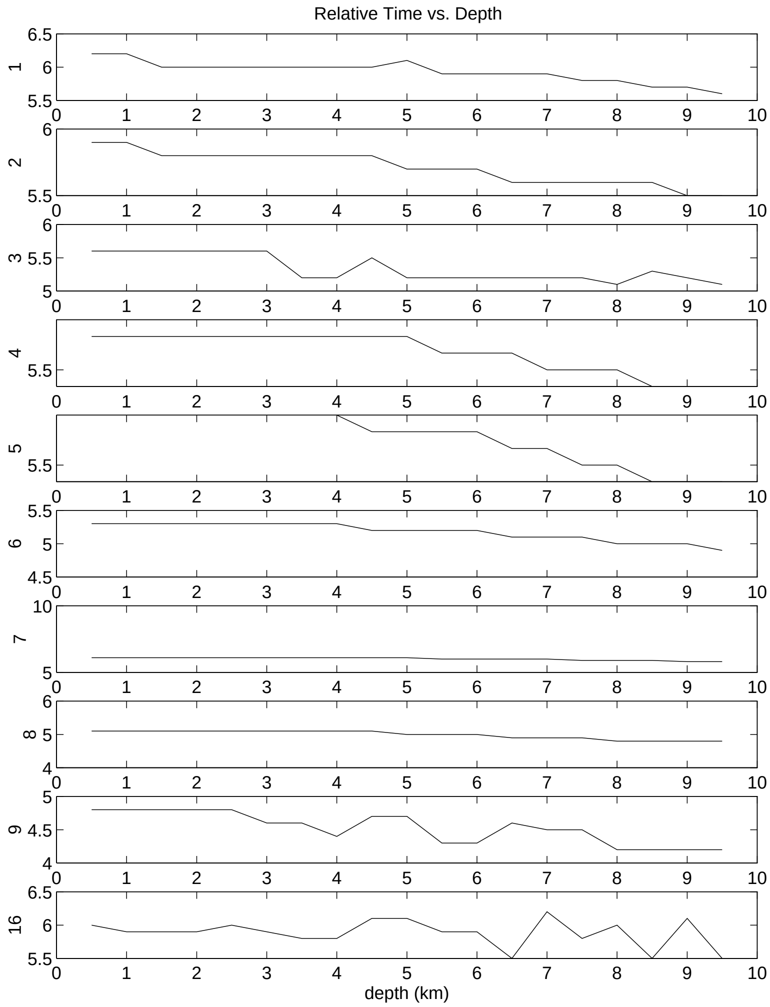


Figure 4.22 Relative Origin Time vs. Depth for the local events in the Primary test set. In the current Implementation of *LWCEDS* only origin times within a 12 second window centered on the coarse search origin time are considered. The time on the vertical axis is with respect to the origin calculated in the coarse search yielding early origins for small relative time and late origins for large relative time. Thus a relative time of 0 s in the fine grid search is 6 s prior to the coarse search hypocenter and a relative time in the fine search of 0 s is precisely the coarse grid origin time. Generally the maximum correlation occurs at earlier origin times for increasing source depths. (Evid is indicated to the left of each plot).

The maximum correlation value as a function of fixed source depth is a wide, generally smooth envelope for the local events from the primary test set (Figure 4.23). The *Seismos* depth estimate is plotted as an x in Figure 4.23 and the *LWCEDS* depth estimate is plotted as an o. It is possible to gain insight on three aspects of the automatic depth estimates from Figure 4.23. First, the wide envelopes are indicative of a poorly constrained estimate. Second, the automatic depth estimates are not located at the maximum correlation value in Figure 4.23 because of the relatively course resolution of 3, 6, and 9 km. It is not clear, however, that increasing this resolution would improve the agreement with the *Seismos* hypocenters, and significant computational savings can be gained by restricting the depth search to three bins because a complete correlation and grid search must be performed for each Master Image (i.e. each depth). Third, deeper events appear to have automatic depth estimates which agree more closely with the *Seismos* estimates.

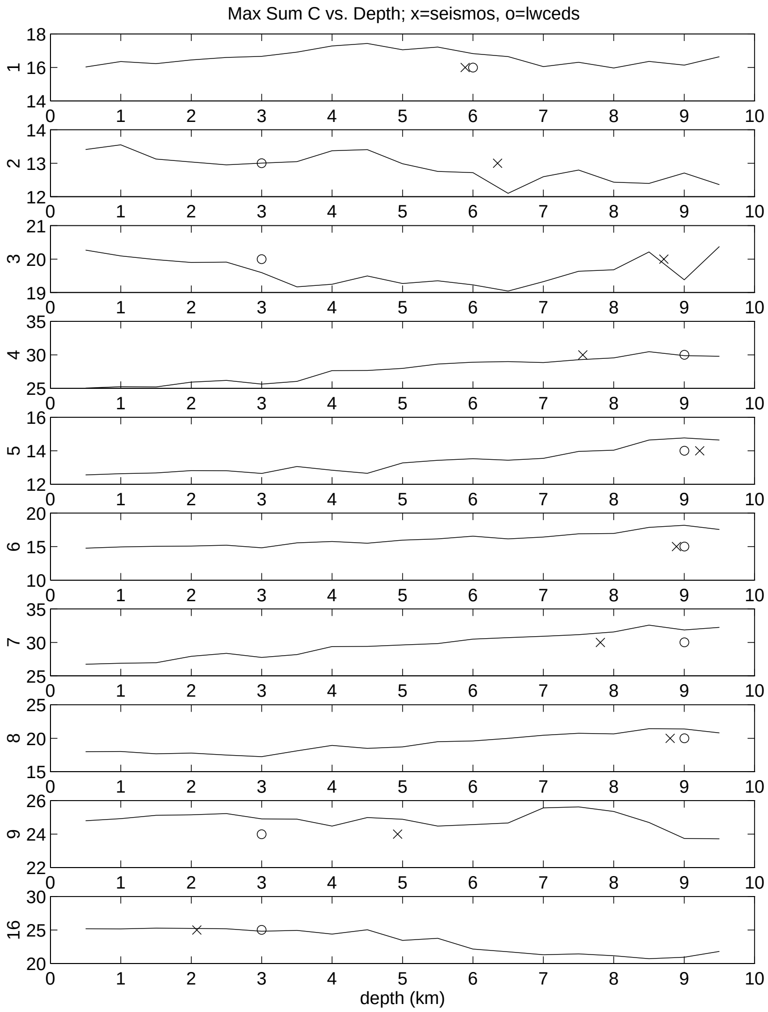


Figure 4.23 Maximum Correlation Value vs. Depth for the local events in the primary test set using fixed source depths in the correlation ranging from 0.5 km to 9.5 km in 0.5 km increments showing broad envelopes (poor constraint) on the automatic depth estimates. The *Seismos* depth estimate is plotted as an x, and the *LWCEDS* depth estimate (considering only possible source depths of 3, 6, and 9 km) is indicated with an o. Agreement between the two methods appears to improve with increasing depth.

Greater insight can be gained by examining the relationship between the *LWCEDS* and *Seismos* depth estimates. This is plotted in Figure 4.24 for the local events in both the primary test set (o's) and the expanded test set (x's; mislocated events have been excluded). There appears to be a relationship (barring two outliers) between depth error and source depth if only the primary test set (o's) is considered: increasing agreement between *LWCEDS* and *Seismos* with increasing source depth. When the expanded test set is included (x's) the relationship is lost. Recall that the primary test set was culled from a larger set that was specifically selected for high quality magma body reflections (Balch, 1997) but membership in the expanded test set only required an origin time that happened to fall within the time of the online experiment. Thus the data from the expanded test set includes fewer reflected phases and hence less accurate depth estimates (for both *LWCEDS* and *Seismos*). For the primary test set, the linear trend of increasing accuracy with increasing depth is likely due to at least three factors. First as depth increases travel distance decreases for reflected phases and hence less signal attenuation and geometric spreading occurs. Second, for small horizontal travel distances, the difference between the S and P-arrivals increases with depth making it more likely that the S-arrival for the closest station will contribute to the correlation (i.e. it will be less likely to be obscured by P-coda yielding a greater SNR). Third, and perhaps most significant, the S-SzP and S-PzP crossover distance decreases with increasing source depth. Thus, as the source depth increases it becomes more likely that these reflected phases will be observed, and not obscured by S-coda. Finally, undersampling during the depth search (i.e. using only three bins) can significantly affect the accuracy of individual hypocenters but does produce generally poorer estimates (*'s in Figure 4.24).

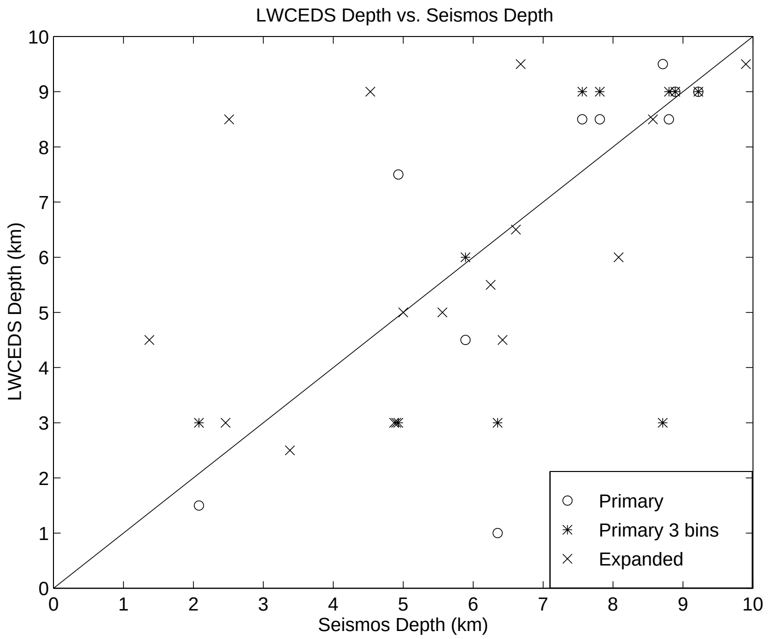


Figure 4.24 LWCEDS Depth vs. Seismos Depth for local events in both the primary (o's) and expanded (x's) test sets. Estimates for the 3 depth bin primary hypocenters are not significantly worse than the high resolution depth search (o's). For the primary test set events, there is an approximate linear trend of increasing agreement between the *LWCEDS* and *Seismos* depth estimates with increasing source depth. This is likely due to the decreasing crossover distance between S and reflected arrivals with increasing source depth. This trend is not seen in the expanded data set due to fewer observations of reflected phases.

4.3 Results Summary

The 156 triggers in the expanded test set were automatically located by *LWCEDS* and the 21 local events were relocated using *Seismos*. All 156 epicenters are mapped in Figure 4.25. Here a regional swarm near Gallinas and expected patterns of mining and quarrying activity can be seen. A suitable filter to remove teleseisms should be developed since they will falsely locate somewhere within the grid. Increasing the low corner of the preprocessing passband enough to filter the teleseismic signals produces deleterious effects in the seismograms for other events.

Checking for fast apparent velocity across the network is likely to be undependable for "filtering" teleseisms, particularly when surface waves are the only observed teleseismic arrivals, because their velocity may be similar to the apparent velocity of some regional phases. Further, deep local events centered on the network may have apparent velocities that approach vertically traveling teleseismic body waves. It is likely that spectral analysis or comparison with global bulletins will be necessary to discriminate teleseismic triggers from local or regional activity. A comparison of the *LWCEDS* and *Seismos* local epicenters from the expanded test set is plotted in Figure 4.26 showing variable results. For data which are uncontaminated by spikes (Figure 4.27), the two methods agree to within about 2 or 3 km for local events.

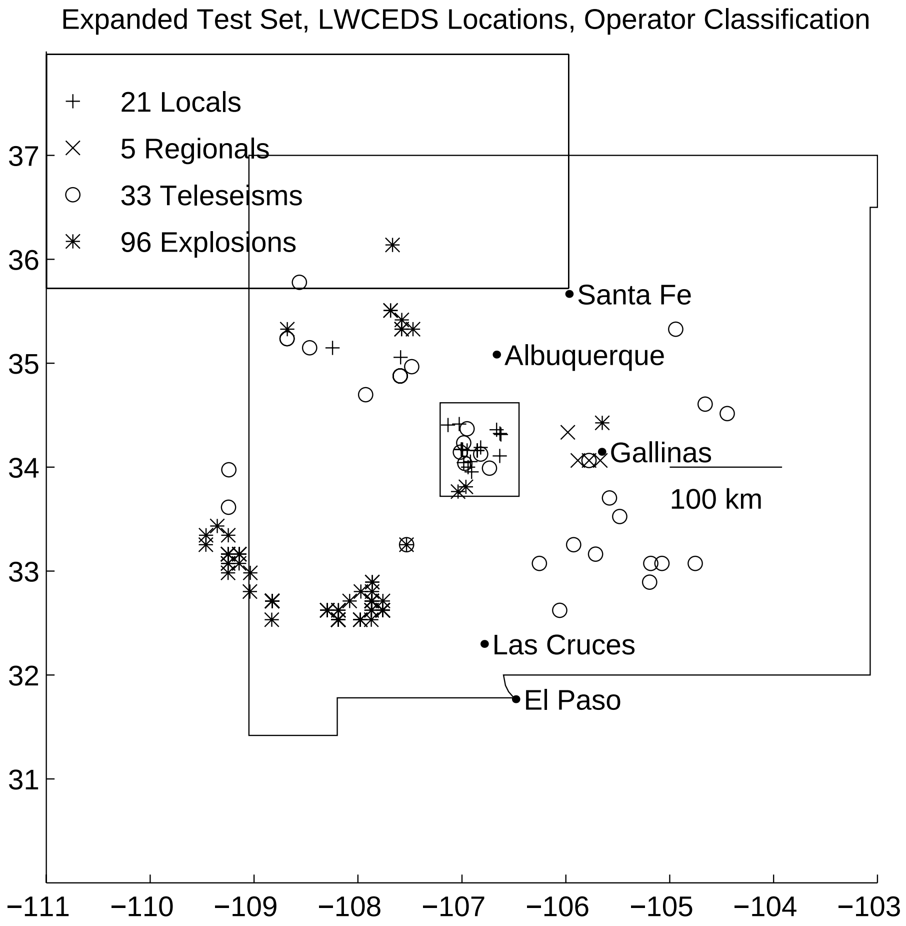


Figure 4.25 Expanded Set Locations showing the epicenters of the expanded test set which are all triggers that were recorded from December 3, 1996 through January 7, 1997 and located using *LWCEDS*. The events were type (local/regional/explosion/teleseism) classified manually.

Of the 21 local events in the expanded test data set, 19% are mislocated by more than 10 km due to electronic spikes, 19% are mislocated by 3 to 9 km because of poor azimuthal constraint (the events on the eastern, northern, and northeastern boundary of the network aperture in Figure 4.27), one event (5%) is mislocated by 9.28 km because of poor signal-to-noise ratio at all but one station, and the remaining 57% have epicenters that are within 3 km of the *Seismos* epicenters. Significant gains could be made by expanding the network station coverage and by developing a routine to mask electronic spikes (a Q dependent exponential?). The spikes are usually no more than a few points wide (often a single point) and are one-sided (i.e. either positive or negative).

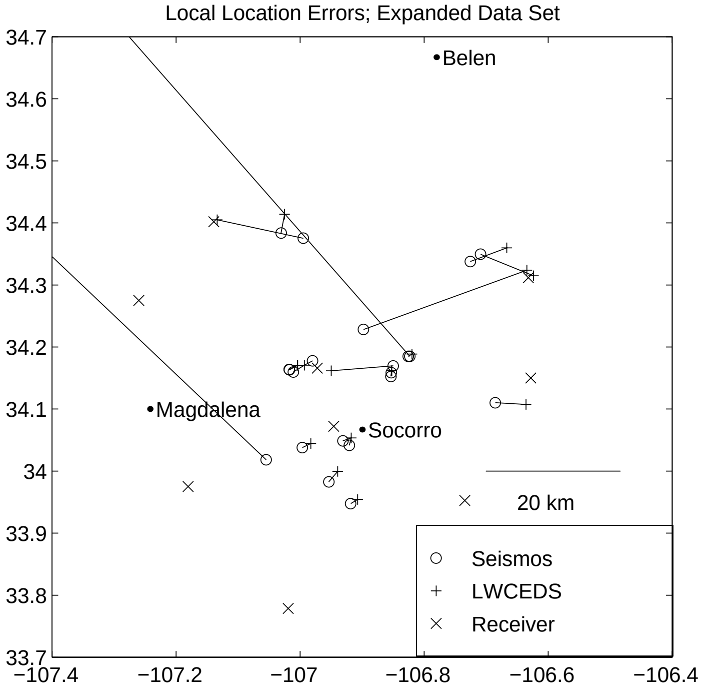


Figure 4.26 Expanded Set Local Errors showing the epicenters for the 21 local events as determined by *LWCEDS* with lines to link these locations to the epicenters determined using *Seismos*. The seismograms are frequently contaminated by electronic and telemetry spikes causing mislocations.

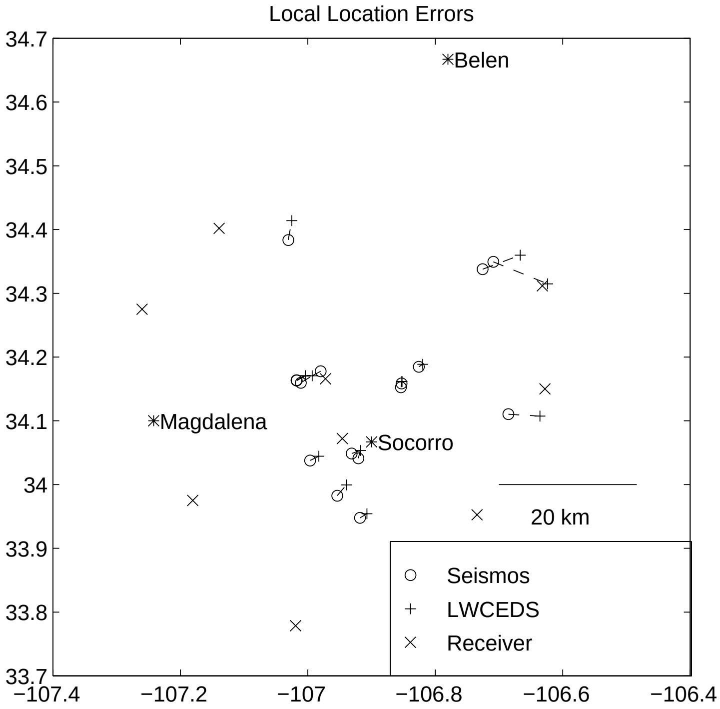


Figure 4.27 Well Located Expanded Set Local Errors showing the epicenters for the 76% of local expanded test set epicenters as determined by *LWCEDS* with lines to link these locations to the epicenters determined by *Seismos*. The four noise contaminated mislocated events and the one poor SNR event are excluded from this plot.

In summary, we have automatically located 57% of the local events in the expanded test set and 90% of the local events in the primary test set to within 3 km of the *Seismos* locations. The percentage of well located events in the expanded test set may be increased to 95% given development of suitable spike masking and expanded station coverage.

5. Conclusions

The goal of this research was to develop an automated local seismic event location system based on waveform correlation. It is intended to complement similar work being performed by the CTBT R&D group at Sandia National Laboratories and we have shown the local implementation of *WCEDS* (*LWCEDS*) to be a viable automatic seismic event location algorithm. The local locations are within 2 to 3 kilometers of manually located events when the raw data are not contaminated by spikes. This system has an advantage over other systems because of its lack of dependence on explicit phase identification, which makes it easier to implement in a variety of seismogenic regions. The maps of the maximum summed correlation values may provide a more realistic means of obtaining location error than error ellipses if one can develope a method for quantifying these images. The *LWCEDS* unreviewed bulletins will allow more efficient manual relocation and phase identification. Given that effective solutions to present shortcomings can be implemented (see Section 5.1), *LWCEDS* presents the possibility of an alarm system for potentially damaging earthquakes because locations can be generated automatically and unreviewed bulletins can be provided in near real-time.

To summarize the *LWCEDS* system:

- The data are preprocessed with an adaptive STA/LTA which allows enhancement of short-period first arrivals and longer period later arrivals with a single algorithm.

- A master image is generated from theoretical travel time curves. The zero elements in the sparse master image are kept to facilitate future implementation of an empirical stack and to retain flexibility for velocity models from different seismogenic provinces.

- The matrix multiplication scheme of Young et al. (1996) is then used to perform the correlation and generate the C matrix.

- A two step grid search is performed summing across the stations within the C matrix at distances appropriate for each station. A coarse search is performed over the entire state of New Mexico, followed by a fine search, when appropriate, over the Socorro Seismic Anomaly.

- Results are placed into a CSS3.0 ASCII database. The database allows access to the hypocenter and waveform data without detailed knowledge of the storage format and location. It also allows a central location to store the data which simplifies security, updates, and dissemination.

5.1 Solutions to Present Shortcomings

For *LWCEDS* to be used as a regional location system and to improve the quality of the local event bulletins, the following improvements are in progress:

- Increase station coverage within the state of New Mexico.

- Install a filter or mask to remove spikes from the raw data streams to prevent mislocations, perhaps by "looking" for impulsive one-sided arrivals with no coda. In some cases, it may be possible to mitigate the effects of electronic spikes and localized noise transients by excluding single station locations (i.e. solutions which are dominated by a one data stream). This may be accomplished by estimating the relative contribution of each station to the correlation value for each grid point while performing the summation through the C-matrix.

- Install a filter for teleseisms which checks for apparent velocity across the network, frequency content, and/or uses other sources of information. Care must be used when checking for slowness across the array otherwise local events may be removed as teleseisms (particularly for deep events centered on the network). Relying on other teleseismic bulletins may not provide a comprehensive teleseismic list in near real time. Spectral analysis of the unprocessed data may be required.

- Install a module to estimate duration magnitude. This may be accomplished by comparing the average post-event energy with the average pre-event energy to estimate the coda duration. Severe underestimation has been troublesome perhaps because of the inability of the automatic method to project the body wave coda into the background noise and the lack of frequency content being used in the estimate. Overestimates could be a problem in the presence of surface waves or multiple events. Detailed studies should be performed to ensure the automatic estimates agree with previous estimates.

- Install a user interface to refine locations and interface with *Seismos* and the CSS3.0 database. This will allow repicking and maintenance of the database without directly editing the ASCII files, but will require modifying *Seismos* or developing an interface that allows it to read from and write to the database.

5.2 Future Work

This research presents interesting possibilities for future projects. The author and others have noticed unidentifiable phases in the data which may be shown to be systematic by generating an empirical stack. The processed waveforms may be averaged for appropriate distances and used to replace the theoretical master image. As more digital data become available, these images may be constructed for various areas of the monitoring region, for different azimuths within a region, for various depth bins, and for other refining criteria. This will improve the shape and weight of the phase envelopes as well as help identify new phases. Further, different master images (including theoretical ones) may be used to more accurately represent various aspects of the model (e.g. depth to Moho), thus eliminating the reliance on a single layer cake model for the entire state. As improved master images become available, refined locations can be obtained and iterations can be performed over successively generated images and locations.

Great care must be used in interpreting the maximum correlation value. In this study we have found no relationship between this value and epicentral distance between the *LWCEDS* and *Seismos* locations. Nevertheless, a visual representation of the accuracy of the solution is contained within a map of the correlation value for each grid point. This presents the possibility of using these maps to quantify the quality of the location (perhaps through the area of the 95% contour, or some similar scheme).

Experienced analysts are often able to qualitatively locate an event by looking at the waveforms, sometimes a single waveform. It may be possible to generate a database of unprocessed waveforms and in a tree-like fashion, search this database for locations that have historically produced similar waveforms. This would not be entirely dissimilar to producing a suite of master images from unprocessed waveforms where azimuth, province, and depth are incorporated and delineated.

6. References

Ake, J. (1983). A Magnitude Scale for Central magnitude scale for central New Mexico based on signal duration, *Geophysics Open-file Report,* **45,** 8 pages.

Allen, R. (1978). Automatic Earthquake Recognition and Timing from Single Traces, *Bull. Seism. Soc. Am.,* **68,** 1521-1532.

Bache, T., Bratt, S., Wang, J., Fung, R., Kobryn, C., and J. Given (1990). The intelligent monitoring system, *Bull. Seism. Soc. Am.,* **80,** 1833-1851.

Balch, B., Hartse, H., Sanford, A., and K. Lin (1997). A New Map of the Geographic Extent of the Socorro Midcrustal Magma Body, *Bull. Seism. Soc. Am.,* **87,** 174-182.

Beal, G., Le Bras, R., Nagy, W., Sereno, T., and H. Swanger (1995). Global Association System Phase 2: Conflict Resolution, *SAIC Report,* **SAIC-95/1029,** 46 pages.

Console, R. editor (1994). Special issue on the workshop: planning and procedures for GSETT-3, *Annali Di Geofisica,* **37 #3,** 241-457.

Evans, J. and S. Allen (1983). A teleseismic-specific detection algorithm for single short period traces , *Bull. Seis. Soc. Am.,* **73,** 1173-1186.

Gutenberg, B., and C. Richter (1954). Seismicity of the earth and associated phenomena, Princeton Univ. Press, Princeton.

Hartse, H. (1991). Simultaneous Hypocenter and Velocity Model Estimation using Direct and Reflected Phases from Microearthquakes within the Central Rio Grande Rift; New Mexico, *PhD Dissertation, New Mexico Tech,* 251 pages.

Hartse, H., Sanford, A., and J. Knapp (1992). Incorporating Socorro Magma Body Reflections into the Earthquake Location Process, *Bull. Seism. Soc. Am.*, **82,** 2511-2532.

Johnson, C., Lindh, A., and B. Hirshorn (1994). Robust Regional Phase Associations, *U.S. Geological Survey Open File Report,* **94-621,**

Johnson C., Bittenbinder, A., Bogaert, B., Dietz, L., and W. Kohler (1995). Earthworm: A Flexible Approach to Seismic Network Processing, *IRIS Newsletter,* **14 #2,** 1-4.

Le Bras, R., Swanger, H., Sereno, T., Beall, G., Jenkins, R., and W. Nagy (1994a). Global Association; Design Document and User's Manual, *SAIC Report,* **SAIC-94/1142,** 67 pages.

Le Bras, R., Swanger, H., Sereno, T., Beall, G., Jenkins, R., Nagy, W. and A. Henson (1994b). Global Association; Final Report, *SAIC Report,* **SAIC-94/1155,** 28 pages.

Lee, W., Dodge, D. (1992). A Course on: PC-Based Seismic Networks, *U.S. Geological Survey Open File Report,* **92-441,** 535 pages.

Lin, K., Sanford, A., and I. Tsai (1996).. Probabilistic Seismic Hazard of the Socorro, New Mexico Area of the Rio Grande Rift, *New Mexico Tech Geophysical Research Center Open-File Report,* **82,** 30 pages.

Press, W., Teukolsky, S., Vettering, W., and B. Flannery (1992). Numerical Recipes in C, Cambridge University Press, 994 pages.

Sanford, A., Balch, R., and K. Lin (1995). A seismic anomaly in the Rio Grande rift near Socorro, New mexico, *Seis. Res. Letters,* **66 #2,** 44.

Shearer, P. (1991). Imaging Global Body Wave Phases by Stacking Long Period Seismograms, *J. Geophys. Res.,* **96,** 20353-20364.

Shearer, P. (1994). Global Seismic Event Detection Using a Matched Filter on Long-Period Seismograms, *J. Geophys. Res.,* **99,** 13713-13725.

Skov, M. (1994). Digital Seismic Data Acquisition and Processing as Applied to Seismic Networks in the Rio Grande Rift and on Mount Erebus, Antarctica, *MS Thesis,* New Mexico Tech, 97pp.

Thomas, P. (1965). Geodesic Arc Length on the Reference Ellipsoid to Second-Order Terms in the Flattening, *J. Geoph. Res.,* **70 #14,** 3331-3340.

Tong, C. (1995). Charactarization of Seismic Phases–an Automatic Analyser for Seismograms, *Geop. J. Int.,* **123,** 937-947.

van der Vink, G., Simpson, D., Butler, R., Hennet, C., and T. Wallace(1996). CTBT... At last!, *IRIS Newsletter,* **XV, #3,** 1-3.

Young, C., Harris, M., Beiriger, J., Moore, S., Trujillo, J., Withers, M., and R. Aster (1996). The Waveform Correlation Event Detection System Project, Phase 1: Issues in Prototype Development and Testing, *Sandia Report,* **Sand96-1916,** 36 pages.

Appendix I. Programming Guide

LWCEDS is composed of multiple main programs that are called from a script and get input from data files and a parameter file. All intermediate files are binary and generated automatically. The parameter file name and path are given as input when calling a main program and most data file names and paths are given as parameters in the parameter file. Generally the main programs are called from a shell script but need not be. An example parameter file is shown at the end of this section. A complete parameter file must be provided or spurious errors will occur. Parameters can be overridden from the command line (e.g. with the call *Detector par=LWCEDS.par phase=0,1*). The general flow of the program is illustrated in Figure I.1.

CreateMasterImage, *CreateGrid*, and *CreateCmap* are run only once and need not be run again, unless parameters change. *PreProc*, *Detector*, *Locator*, and *PutDB* are run for every trigger. *PreProc* generates binary files of preprocessed data. *Detector* performs the coarse grid search. *Locator* performs the fine grid search, and *PutDB* places the results into the CSS3.0 flat-file database.

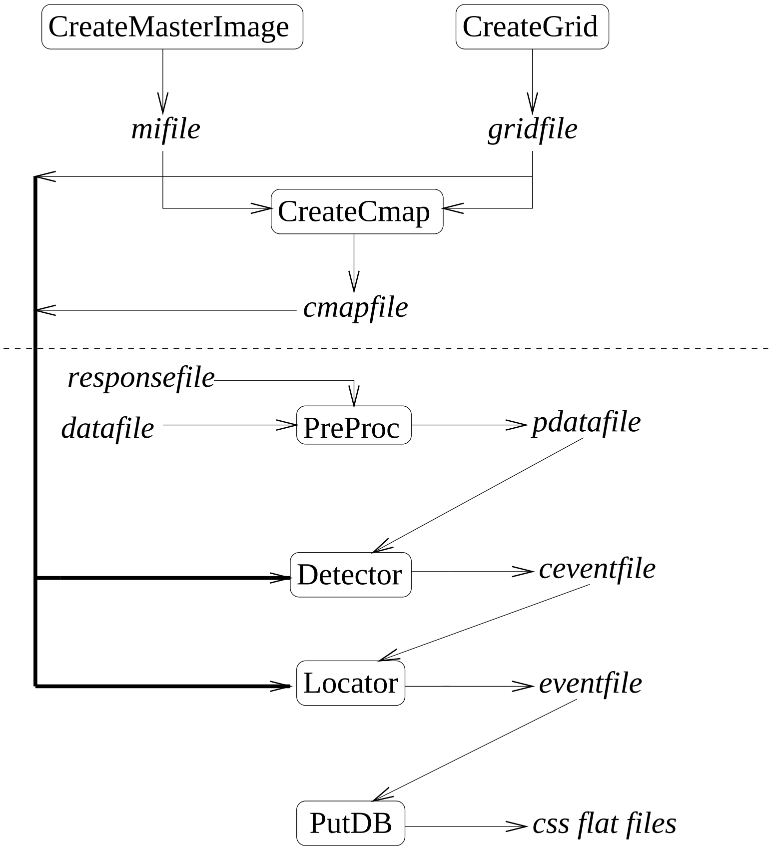


Figure I.1 Program Flow Diagram illustrating the connectivity of the main programs. Intermediate files are written in binary and italicized in the figure. Programs above the dashed line need only be run when parameters are changed. Programs below the dashed line are run once for each triggered datastream.

I.1 CreateMasterImage

Appropriate master image files must be created using *CreateMasterImage*, prior to running *LWCEDS*. These files need only be created once, and are called, for example, with the command *CreateMasterImage par=LWCEDS.par*. The parameter file must be included in the call to *CreateMasterImage*, and the parameters listed in this section must be provided in the file.

CreateMasterImage generates coarse and fine master images in exactly the same manner using the different coarse and fine parameters specified in the parameter file. The images are sized the same for each depth. The travel time curves are then calculated using *curvfnd* (Hartse 1991) which is a ray tracing algorithm. It is assumed that physically reasonable minimum and maximum distances are given for each phase; *curvfnd* may hang or give erroneous results if this assumption is violated. These time distance pairs are then fed to the subroutine *GetPhaseWidth*. Within the maximum and minimum width parameters specified in the parameter file, the width of each curve is determined by the grid spacing, the slope of the travel time curve, and the discretization of the Master Image, whichever is greatest. The grid spacing is used as the distance dx, and the width, dt, is determined by the formula for slope: slope=dt/dx. Each phase is then added into the master image matrix with a call to the subroutine *AddPhaseToMI*. For each distance in the travel time curve, the width is set to span the curve equally in time, the weight is determined by the envelope (rectangles are the only shapes currently supported), the width, and the weight parameter. Finally the entire master image is scaled by a plane whose slope and intercepts are defined in the parameter file and it is written to a binary file.

Parameters

Cphaselist: list of phase codes to be included in coarse master image (uses the *Seismos* standard, 2=Pg, 3=Sg, 6=Pn, 7=Sn, etc). A comma separated list of integers.

Cphaseweight: weight for each coarse image phase. A comma separated list of floats. A float for each entry in Cphaselist must be provided; order counts.

Fphaselist: phase codes to be include in fine master image (uses the *Seismos* standard, 0=P, 1=S, 8=SzS, etc). A comma separated list of integers.

Fphaseweight: weight for each fine image phase. A comma separated list of floats. A float for each entry in Fphaselist must be provided; order counts.

Cdepthlist: a master image for each of these depths is used in the coarse search. A comma separated list of floats (in kilometers).

Fdepthlist: a master image for each of these depths is used in the fine search. A comma separated list of floats (in kilometers).

Cshape: code to indicate the shape of the phase envelopes in the coarse master image. Currently only 0 (rectangle) is supported. A comma separated list of integers, one entry for each entry in CphaseList.

Fshape: code to indicate the shape of the phase envelopes in the fine master image. Currently only 0 (rectangle) is supported. A comma separated list of integers, one entry for each entry in FphaseList.

curvepath: absolute path to find tdpairs.dat generated by *curvfnd*. A string.

LD: maximum distance for local (fine) master image. Be careful because this value is reflected in the size of the fine grid and this parameter could be nullified in some cases without appropriate (yet still physically realistic) changes in FmaxT and FmaxD. A float (in kilometers).

RD: maximum distance for regional (coarse) master image. Be careful this value is reflected in the size of the coarse grid and this parameter could be nullified in some cases without appropriate (yet still physically realistic) changes in CmaxT and CmaxD. A float (in kilometers).

plane: slope and intercepts of plane by which to multiply the master image. Intended to crudely mimic signal attenuation with increasing travel distance. Comma separated list of four floats. Can produce large variations in the results so change only as a last resort. Is normalized by the size of the master image so entries are unitless. The time, t, and distance, d, dependent plane, y, uses the following parameters: $y = ad + bt + c + etd$. The order of the user defined paramaters is plane=a,b,c,e where e was included to allow non-linearity in the plane but has been set to 0 for all tests in this document.

Coffset: distance spacing of bins in the coarse master image (in kilometers). A float.

Foffset: distance spacing of bins in the fine master image (in kilometers). A float.

CminTerr: minimum timing error (envelope width) for each phase in coarse master image (in seconds). A comma separated list of floats, an entry for each entry in Cphaselist; order counts.

FminTerr: minimum timing error (envelope width) for each phase in fine master image (in seconds). A comma separated list of floats, an entry for each entry in Fphaselist; order counts.

CmaxTerr: maximum timing error (envelope width) for each phase in coarse master image (in seconds). A comma separated list of floats, an entry for each entry in Cphaselist; order counts.

FmaxTerr: maximum timing error (envelope width) for each phase in fine master image (in seconds). A comma separated list of floats, an entry for each entry in Fphaselist; order counts.

CminT: coarse master image will have legal arrival times between CminT and CmaxT (it will be CmaxT-CminT seconds in length). Travel times from curves outside this bound will not be used. Intended to prevent the master image from becoming inadvertently large. A float (in seconds).

CmaxT: see CminT.

FminT: fine master image will have legal arrival times between FminT and FmaxT (it will be FmaxT-FminT seconds in length). A float (in seconds).

FmaxT: see FminT.

CminD: minimum distance (in kilometers) for each phase in coarse master image (a comma separated list, with a distance for each phase in CphaseList). Intended to prevent calculating inappropriate arrival times (e.g. less than the critical distance for a refracted phase).

CmaxD: maximum distance (in kilometers) for each phase in coarse master image (a comma separated list, with a distance for each phase in CphaseList).

FminD: minimum distance (in kilometers) for each phase in fine master image (a comma separated list, with a distance for each phase in FphaseList).

FmaxD: maximum distance (in kilometers) for each phase in fine master image (a comma separated list, with a distance for each phase in FphaseList). Intended to prevent calculating inappropriate arrival times (e.g. greater than twice the distance from the source to the farthest reflection point for PzP).

CMIout: path and file prefix for binary output of coarse master image data filename will be appended with the depth (e.g. /home/user/Image/MI will cause a file to be created for the 5 km depth called /home/user/MI.5.00). A string.

FMIout: path and file prefix for binary output of fine master image data filename will be appended with the depth. A string.

I.2 CreateGrid

The coarse and fine grids are created in exactly the same manner using different coarse and fine parameters from the parameter file. Minimum and maximum latitudes and longitudes are determined from the parameters. A recursive procedure for calculating the grids is employed using the Andoyer-Lambert approximation to first order flattening as described by Thomas (1965). The four corners of the grid are specified in degrees and the grid spacing is specified in kilometers. First, a point is determined that is a given distance (specified by the grid spacing) due north of the southwest corner. Next, the point directly north of the previously calculated point is computed, continuing until the maximum latitude has been reached. In this way lines of latitude that are all separated by the grid spacing are determined. For each line of latitude, the grid is filled by determining the coordinates for the point which is the given distance due east of the previous point. This continues for each line of latitude until the grid is completely filled in. This procedure provides nodes which form squares of equal area, but is not effective for grids covering the entire globe since the eastern edge of the grid is designed to keep the entire grid rectangular in a cartesian sense, rather than follow any particular meridian.

Parameters

refelev: elevation in kilometers to add to the second order approximation of the surface geoid for distance calculations (generally elevation at WTX is used; 1.55 km).

Cgrdsp: spacing of coarse grid points in kilometers. A float.

Fgrdsp: spacing of fine grid points in kilometers. A float.

Fminlat: minimum latitude of fine grid (negative south of equator). A float.

Fmaxlat: maximum latitude of fine grid. A float.

Fminlon: minimum longitude of fine grid (negative west of prime). A float.

Fmaxlon: maximum longitude of fine grid. A float.

Cminlat: minimum latitude of coarse grid (negative south of equator). A float.

Cmaxlat: maximum latitude of coarse grid. A float.

Cminlon: minimum longitude of coarse grid (negative west of prime). A float.

Cmaxlon: maximum longitude of coarse grid. A float.

I.3 CreateCmap

Each grid point has associated with it, a unique summation path through the C matrix. By summing and normalizing by the number of stations, we get the combined correlation of each station with the model predictions for the specified grid point. These summation paths could be calculated on the fly using the distance between each station and the grid point, but the Cmap provides a much faster means of indexing. For each grid point, the distance and appropriate index into the C matrix are calculated for each station and stored in a file to be used by the location programs. Changing grid parameters and some master image parameters (particularly size related parameters) requires a new Cmap. Note that depth need not be included in the distance used to generate the Cmap since this is already incorporated into the Master Image. That is, the travel times in the master images are indexed by horizontal travel distance.

Parameters

CGridFile: path and file of coarse grid. A string.

StaFile: path and file of station coordinates. A string.

CCmapFile: path and file prefix (depth gets appended to filename) to store binary coarse cmap file. There is a separate Cmap for each depth in case different Master Images are used (probably unnecessary). A string.

FGridFile: path and file of fine grid. A string.

FCmapFile: path and file prefix (depth gets appended to filename) to store binary fine cmap file. A string.

I.4 PreProc

Preprocessing can be performed on either single component or three component data. A station file is used to specify raw data file names and whether the station is single or three component. The raw data are University of Alaska Fairbanks ad hoc format. PreProc tests whether the data files actually exist, and keeps track of the number of stations for which there is data. After reading header and data for each station, the mean is calculated for each and removed, and the data then are bandpass filtered using a time-domain convolution with the impulse response for the filter. We currently use the *Matlab Signal Processing Toolbox* to generate this file. Next a vector of zero crossings is generated, the data are squared, and the adaptive STA/LTA is generated. If a station is three component, the data for each channel of the station are squared, summed point by point, and the resulting data stream is divided by three prior to application of the adaptive STA/LTA. The adaptive STA/LTA stream is then decimated and any points below a specified threshold are set to zero. The processed data streams are then written to a headerless binary file.

Parameters

NumWfdisc: maximum number of stations. An integer used to size arrays.

WfdiscPath: absolute path to ah files. A string.

WfdiscStaFile: absolute path and name to file containing data file names and preprocessing codes. Preprocessing codes currently supported are 0 (single component) and 1 (3-component). A string.

ResponseFile: filename and path of vector containing impulse response of bandpass filter which get convolved with the data. A string.

NumZC: number of zero crossings for the short window. An integer.

LTRatio: ratio of length of long term window over length of short window. An integer.

Cdecrate: decimation rate for the Coarse search data. A float in samples per second.

Fdecrate: decimation rate for the Fine search data (currently must be the same as Cdecrate). A float in samples per second. Because only a single processed data stream is currently generated, the coarse and fine decimation rates must be equal.

ZeroThreshold: "Water level" for output of adaptive STA/LTA. Any STA/LTA values below this threshold will be set to zero. A float.

samprate: nominal sampling rate of the raw data in samples per second.

I.5 Detector

The *Detector* performs the coarse grid search. After the Master Image, Cmap, and processed data are loaded into memory, the C matrix is populated for the entire time series at once by performing the correlation in the frequency domain using the Numerical Recipes routine *correl*. For an assumed origin time, the maximum summation through C is found for a given depth. The maximum summation at the next depth is then compared with the previous and the larger is chosen. Once the maximum summation in three spatial dimensions has been determined, the assumed origin time is incremented by a specified time step, and the process is repeated comparing the new maximum summation with the previous assumed origin time. If the ultimate maximum summation value is greater than a threshold, an event is declared. A few diagnostic files recording correlation values are written in addition to the location, but these are only appropriate for the most recent trigger on which the program was run and will likely be disabled in the final version of the code.

Parameters

DetectThreshold: maximum summation through C-matrix must be greater than this value to declare an event. A float.

CTimeStep: step (in seconds) through the data by this value, performing a coarse grid search at each step. A float.

FTimeStep: step (in seconds) through the data by this value, performing a fine grid search at each step. A float.

CCofTFile: path and file to output max Correlation as function of time with resolution of CTimeStep. A string.

CGofTFile: path and file to output grid point of max Correlation as function of time with resolution of CTimeStep. A string.

CDofTFile: path and file to output depth of max Correlation as function of time with resolution of CTimeStep. A string.

CEventFile: intermediate file containing hypocenter resulting from Detector. A string.

FCofTFile: path and file to output max Correlation as function of time with resolution of FTimeStep. A string.

FGofTFile: path and file to output grid point of max Correlation as function of time with resolution of FTimeStep. A string.

FDofTFile: path and file to output depth of max Correlation as function of time with resolution of FTimeStep. A string.

FEventFile: intermediate file containing hypocenter resulting from Locator. A string.

GlitchWeight: height of glitch phase; intended to "filter" glitches and teleseisms. A float. Not currently used because it "filters" too many locals.

GlitchWidth: width of glitch phase in seconds.

I.6 Locator

The *Locator* program performs the correlation and search for the fine grid. The coarse location is read from the output file of the *Detector* and if this location is not within the four corners of the fine grid, the fine search is not performed. Otherwise, the operation of the *Locator*, is very similar to that of the *Detector*, except the grid search is only performed for an intrinsically defined time window centered on the coarse origin time (currently 12 seconds). The results, whether from the coarse or fine search are written to a final event file.

I.7 PutDB

PutDB places the hypocenter results and waveform pointers into the CSS flat-file database. Origins are read from the *Locator* output file and the appropriate data file header information is obtained. The lastid table from the database determines the orid and evid for the new event, and the appropriate lines are added to the origin, wfdisc, and lastid tables.

I.8 Example Shell Script

Generally, the main programs are called from a shell script which uses a parameter file, but also overrides some parameters. The WfdiscPath, for example, provides the path to the data and specifying this string from the command line within the script, allows looping over a number of triggers.

```
#!/bin/tcsh
#
# set the working directory
#
set dir=/mk2/mmw/LWCEDS
#
# make a tempory data directory
#
cd $dir/ONLINE_TEST/temp
#
# don’t mess with the raw data, just copy it into
# the tempory directory, convert to ah, and clean.
#
foreach event (‘ls -d /wren6/data/*.wvm‘)
   set myevent = ‘echo $event | colrm 1 12‘
```

```
    echo $myevent
    cp $event $dir/ONLINE_TEST/temp/$myevent
    /usr/local/bin/pcx2ah $myevent
    rm $dir/ONLINE_TEST/temp/$myevent
end
#
# find out what directories we have to work with
#
cd $dir/ONLINE_TEST/temp
ls -d * >! ../current.event.list
cd $dir
#
# loop over number of triggers sending output to a logfile
#
foreach event (‘cat $dir/ONLINE_TEST/current.event.list‘)
    mv $dir/ONLINE_TEST/temp/$event $dir/ONLINE_TEST/data
    set event = $dir/ONLINE_TEST/data/$event/
    echo working on $event/
    echo working on $event >> $dir/ONLINE_TEST/logfile
    #
    # preprocess the data
    #
    cd $dir/PreProc
    PreProc par=$dir/ONLINE.par WfdiscPath=$event >>
                                                   $dir/ONLINE_TEST/logfile
    echo PreProc done
    #
    # do the coarse search
    #
    cd $dir/Detector
    Detector2 par=$dir/ONLINE.par WfdiscPath=$event >>
```

```
                                        $dir/ONLINE_TEST/logfile
#
# keep the coarse locations in a separate file
#
cat $dir/ONLINE_TEST/coarse.event.dat >>
                                        $dir/ONLINE_TEST/Cevent.dat
echo Detector done
#
# do the fine search
#
cd $dir/Locator
Locator2 par=$dir/ONLINE.par WfdiscPath=$event >>
                                        $dir/ONLINE_TEST/logfile
echo Locator done
#
# run the magnitude estimator
#
cd $dir/Magnitude
GetMag par=$dir/ONLINE.par WfdiscPath=$event >>
                                        $dir/ONLINE_TEST/logfile
#
# add the event to a final location file
#
cat $dir/ONLINE_TEST/fine.event.dat >>
                                        $dir/ONLINE_TEST/event.dat
#
# update the database
#
cd $dir/DBLink
PutDB2 par=$dir/ONLINE.par WfdiscPath=$event >>
                                        $dir/ONLINE_TEST/logfile
```

```
end
```

I.9 Example Parameter File

The parameter file values used for the calculations in this document are provided here. No Defaults exist and presence of this file and the parameters is not checked for. If a complete parameter file is not specified on the command line, unpredictable errors will occur. Certain parameters can be overridden from a shell program (e.g. WfdiscPath), which allows looping through a number of parameter values (e.g. multiple triggers).

```
MagDelay=5.0
NumWfdisc=11
WfdiscPath=”/mk2/mmw/LWCEDS/data/950920/950920090959/”
WfdiscStaFile=”/mk2/mmw/LWCEDS/PreProc/WfmFile.Names”
ResponseFile=”Filters/IP.1.35”
NumZC=5
LTRatio=5
Cdecrate=10.0
Fdecrate=10.0
ZeroThreshold=2.0
samprate=100.0
Cphaselist=0,1,2,3,6
Cphaseweight=1.2,1.00,1.20,1.00,1.10
Fphaselist=0,1,8,9,10
Fphaseweight=1.4,1.20,1,1,1
```

```
Cdepthlist=5.00
Fdepthlist=3.0,6.0,9.0
Cshape=0,0,0,0,0
Fshape=0,0,0,0,0
curvepath=”/mk2/mmw/socorro/Model/curves/”
LD=110
RD=510
plane=-0.6,-0.4,1.0,0.0
Coffset=10
Foffset=1
CminTerr=0.1,0.2,0.2,0.2,0.2
FminTerr=0.1,0.2,0.2,0.2,0.2
CmaxTerr=10.0,10.0,10.0,10.0,10.0
FmaxTerr=10.0,10.0,10.0,10.0,10.0
CminT=0.0
CmaxT=150.0
FminT=0.0
FmaxT=35.0
CminD=0,0,110,110,110
CmaxD=110,110,510,510,510
FminD=0,0,0,0,0
FmaxD=110,110,140,140,80
CMIout=”/mk2/mmw/LWCEDS/Images/Coarse/MI”
```

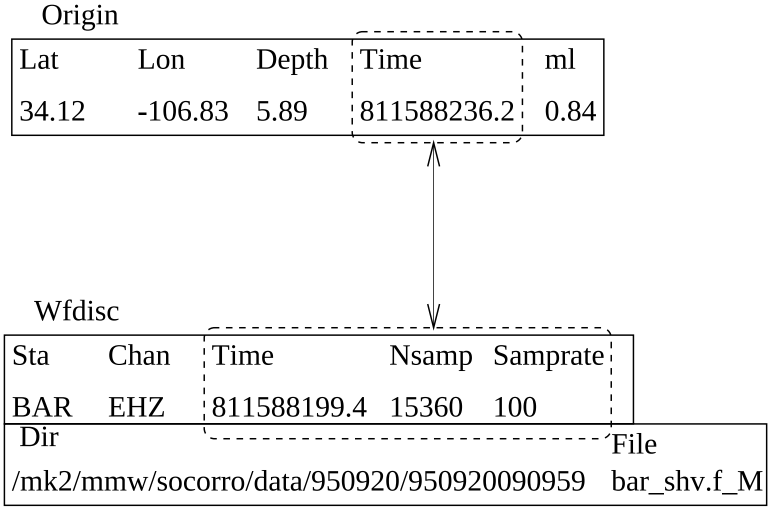
```
FMIout=”/mk2/mmw/LWCEDS/Images/Fine/MI”
refelev=1.555
Cgrdsp=10
Fgrdsp=1
Fminlat=33.72
Fmaxlat=34.62
Fminlon=-107.21
Fmaxlon=-106.45
Cminlat=31.00
Cmaxlat=37.00
Cminlon=-110.00
Cmaxlon=-103.00
CGridFile=”/mk2/mmw/LWCEDS/Images/Coarse/Grid.dat”
StaFile=”/mk2/mmw/LWCEDS/Grid/sta.dat”
CCmapFile=”/mk2/mmw/LWCEDS/Images/Coarse/Cmap”
FGridFile=”/mk2/mmw/LWCEDS/Images/Fine/Grid.dat”
FCmapFile=”/mk2/mmw/LWCEDS/Images/Fine/Cmap”
GlitchWeight=0.00
GlitchWidth=1.0
DetectThreshold=0
CTimeStep=1.0
FTimeStep=0.1
CCofTFile=”/mk2/mmw/LWCEDS/Images/Coarse/CofT2”
```

```
CGofTFile=”/mk2/mmw/LWCEDS/Images/Coarse/GofT2”
CDofTFile=”/mk2/mmw/LWCEDS/Images/Coarse/DofT2”
CEventFile=”/mk2/mmw/LWCEDS/coarse.event.dat2”
FCofTFile=”/mk2/mmw/LWCEDS/Images/Fine/CofT2”
FGofTFile=”/mk2/mmw/LWCEDS/Images/Fine/GofT2”
FDofTFile=”/mk2/mmw/LWCEDS/Images/Fine/DofT2”
FEventFile=”/mk2/mmw/LWCEDS/fine.event.dat2”
Tmag=0.95
FinalEventFile=”/mk2/mmw/LWCEDS/fine.event.dat2”
DBPath=”/mk2/mmw/LWCEDS/CSS3.0/socorro2/socorro”
updateDB=”y”
```

Appendix II. Data Management

LWCEDS automatically updates a CSS3.0 flat-file database. This database consists of a number of ASCII files in the CSS3.0 format (or schema). CSS3.0 is a standard set of files, each containing a standard set of fields (i.e. columns). Each record in the tables (i.e. row) is unique, but fields may be duplicated between files (indeed, should be duplicated if the database is to be relational). Even within the confines of CSS3.0 considerable customization exists at some institutions (particularly with the use, non-use, or addition of unique identifier keys) so not all CSS3.0 databases are compatible. Nevertheless, the CSS3.0 format is arguably the most common database schema in seismology, hence our decision to use it for *LWCEDS*. A flat-file (i.e. standard ASCII) database requires more storage capacity than binary tables, but is also more accessible to the knowledgable user. Additionally, a no cost flat-file relational database management package is available from the IRIS/USGS Joint Seismic Program (*DATASCOPE*). A system which works on binary files (*postgres*) is also available from IRIS, but it was determined that ASCII files were preferable for a first system to allow system administrators more direct access to the data.

Consider the two database tables shown in Figure II.1: Origin and Wfdisc. The Origin table provides hypocenter information with one hypocenter per line, and the Wfdisc table provides waveform data with one seismogram per line. A single record is shown for both tables. To obtain waveform data for the hypocenter shown in the Origin table, the two records are linked using timing fields. The Wfdisc table has a time field (epoch time: seconds past 00:00:00, January 1, 1970) for the start of the waveform data, as well as the number of samples and sample rate. From this information the time of the end of the record can be calculated. To relate the records from the Origin and Wfdisc tables, we search the appropriate Wfdisc fields for records that have start and end times which bracket the time from the Origin table.



Origin

Lat	Lon	Depth	Time	ml
34.12	-106.83	5.89	811588236.2	0.84

Wfdisc

Sta	Chan	Time	Nsamp	Samprate
BAR	EHZ	811588199.4	15360	100

Dir	File
/mk2/mmw/socorro/data/950920/950920090959	bar_shv.f_M

Figure II.1 Example Database Tables illustrating methods of joining two tables. A single record (i.e. row) from two tables (Origin and Wfdisc) are shown. Times are given in epoch time (seconds past 00:00 UT, January 1, 1970) to facilitate arithmetic operations.

The Origin table will typically be populated with hundreds or even thousands of events but we are often only interested in select events (above a given magnitude for example). Using the database management system, it is fairly easy to create a subset of the entire list which contains only events with magnitude greater than a user specified value. Additional inclusive or exclusive searches (logical AND and logical OR, among other searches) can be performed on any field and joined with appropriate common fields to other tables.

The specific tables used and their fields are known as the database schema (e.g. CSS3.0 schema), while the program used to access the database is the data management system. The schema provides a standard way to store data, and the data management system provides an interface for users to access the data, which ideally insulates the user from the details of where and how the data are stored. A relational database management system (rdbms, e.g. *Datascope* or *postgres*) uses schemas with multiple tables and allows users to join, or relate, records from different tables. Most data management systems also allow the creation of subsets, such as limiting the number of Origin table records to hypocenters with depth greater than 5 km, and with magnitudes greater than 1.3. *Datascope* has the advantage of online help and a community of users available through email.

The database used by *LWCEDS* is still in its infancy and some fields (particularly unique id fields) are not currently used. The tables presently supported by *LWCEDS* are listed in Appendix II. Only the Wfdisc, Origin, and Lastid tables are currently updated automatically. No intrinsic rdbms commands are used by *LWCEDS* so it is independent of the particular data management system used (assuming a CSS3.0 flat-file schema).

The following is a detailed list of the tables used for the New Mexico Tech Seismic Network CSS3.0 database. Not all listed tables are currently used, but are included for potential usefulness (eg. *LWCEDS* does not pick arrivals, so can't update the arrivals table). Fields which are currently set to NULL are indicated by an X in the NULL? column. The format indicates the number of characters in each field; do not deviate from this or *Datascope* and other schema dependent programs will not work.

II.1 Arrival

The **arrival** table gives information characterizing seismic phases observed at a particular station. Not currently used by *LWCEDS* since explicit phase identification and timing is not performed.

field	format	NULL?	Description
sta	%6s	X	station name acronym
time	%17.5f	X	phase arrival epoch time (seconds past 1 Jan 1970)
arid	%8d	X	unique arrival identifier
jdate	%8d	X	year and julian day
stassid	%8d	X	id to join arrivals that are from a single event
chanid	%8d	X	another useless id used for the sitechan table
chan	%8s	X	the channel of the station (see IRIS station book
		X	for naming conventions)
iphase	%8s	X	phase that arrived
stype	%1s	X	signal type: l (local), r (regional), t (teleseism),
		X	m (mixed or multiple), g (glitch), c (cal pulse),
		X	s (signal), n (noise)
deltim	%6.3f	X	standard deviation of a detection time
azimuth	%7.2f	X	station to event azimuth measured clockwise from
		X	north in degrees
delaz	%7.2f	X	standard deviation of the azimuth
slow	%7.2f	X	observed slowness across array
delslow	%7.2f	X	standard deviation of slowness

ema	%7.2f	X	emergence angle of arrival measured in degrees
		X	from vertical
rect	%7.3f	X	rectilinearity
amp	%10.1f	X	zero to peak instrument corrected displacement in nm
per	%7.2f	X	period of arrival in seconds
logat	%7.2f	X	log of amplitude/period ratio (not used if amp and
		X	per are not null)
clip	%1s	X	indicates clipping (c) or not clipped (n)
fm	%2s	X	first motion compression and dilation on short period
		X	use c, d, and compression dilation on long period use
		X	u, r; empty char denoted with '.' eg compression
		X	with no long period would be 'c.'
snr	%10.2	X	signal to noise ratio
qual	%1s	X	denote sharpness of phase onset: i (impulsive, accurate
		X	to 0.2s); e (emergent accurate to 1.0s); w (weak
		X	uncertainty ¿1.0s)
auth	%15s	X	originator of arrival info, eg NEIC, IRIS, SEISMOS.
commid	%8d	X	id used to point to remarks table
lddate	%17.5f	X	epoch time record was last modified

II.2 Assoc

The **assoc** table associates arrivals with origins.

field	format	NULL?	Description
arid	%8d	X	unique id assigned to each arrival at each sta and chan

orid	%8d	X	unique id assigned to each origin
sta	%6s	X	code for station
phase	%8s	X	standard seism code for the particular phase
belief	%4.2f	X	qualitative estimate (0 to 1) of confidence that
		X	the phase has been correctly identified
delta	%8.3f	X	source to receiver arc length in degrees
seaz	%7.2f	X	receiver to source azimuth in degrees clockwise from N
esaz	%7.2f	X	source to receiver azimuth in degrees clockwise from N
timeres	%8.3f	X	travel time residual in seconds of the arrival
timedef	%1s	X	flag indicating whether phase is defining (d) or not (n)
		X	ie, whether it was used to locate the event
azres	%7.1f	X	azimuth residual in degrees between observed and "true"
azdef	%1s	X	same as timedef only for azimuth
slores	%7.2f	X	slowness residual in sec/deg bewteen observed and theo
slodef	%1s	X	same as timedef for slowness
emares	%7.1f	X	emergence angle residual
wgt	%6.3f	X	final weight of arrival used in location code (0 to 1)
vmodel	%15s	X	velocity model used eg iasp91, hartse
commid	%8d	X	unique id linking to a comment
lddate	%17.5f	X	epoch time record was last modified

II.3 Instrument

The **instrument** table contains nominal instrument calibrations for a single frequency, and pointers to frequency dependent calibration constants and exact calibration curves.

field	format	NULL?	Description
inid	%8d		unique id linking a sensor with an instrument type
insname	%50s		name of the particular instrument eg sts2:RT72A would be a streckeisen with a RefTek
instype	%6s		instrument type eg SRO, WWSSN
band	%1s		qualitative indicator of pass band s (short period), m (mid period), i (intermediate period), l (long period), b (broadband), h (high frequency, very short period), and v (very long period).
digital	%1s		flag denoting whether record describes analog (a) or digital (d) recording system
samprate	%11.7f		nominal sample rate in samples/sec
ncalib	%16.6f		nominal conversion factor in nm/count for conversion at period specified by ncalper
ncalper	%16.6f		period in sec at which ncalib is valid
dir	%64s		path (excluding file), I guess to the dfile
dfile	%32s		instrument response file (different from dfile in wfdisc which would be the waveform filename)
rsptype	%6s		natural units for the instrument (A, V, or D)
lddate	%17.5f		epoch time of last record modification

II.4 Netmag

The **netmag** table summarizes estimates of network magnitudes (as opposed to magnitude from a single station) for an event.

field	format	NULL?	Description
magid	%8d	X	unique id for this table
net	%8s	X	network used for calculation eg WWSSN, socorro
orid	%8d	X	unique origin id linking to origin table
evid	%8d	X	unique event id (if there is more than one orid
		X	for an evid, then differing opinions on the hypocenter
		X	and/or origin time exist)
magtype	%6s	X	specifies type of magnitude for magid, eg mb, ms,
		X	m0, ml, md etc
nsta	%8d	X	number of stations used to compute magnitude
magnitude			next line
	%7.2f	X	the actual number
uncertainty			next line
	%7.2f	X	standard deviation of magnitude
auth	%15s	X	originator of an arrival (used in the mag calc)
		X	eg NEIC, seismos, balch
commid	%8d	X	unique id linking to a comment
lddate	%17.5f	X	epoch time of last modification of record

II.5 Origerr

The **origerr** table provides error estimates associated with the origin table records.

field	format	NULL?	Description
orid	%8d	X	unique id linking to a calculated origin
sxx	%15.4f	X	element of the covariance matrix for the
		X	origin in orid. The matrix is symmetric
		X	and (x,y,z,t) refer to latitude, longitude,
		X	depth, and origin time. Given in km^2
syy	%15.4f	X	see above
szz	%15.4f	X	see above
stt	%15.4f	X	see above, given in sec^2
sxy	%15.4f	X	see above
sxz	%15.4f	X	see above
syz	%15.4f	X	see above
stx	%15.4f	X	see above, given in km*sec
sty	%15.4f	X	see above
stz	%15.4f	X	see above
sdobs	%9.4f	X	square root of sum of squares of time residuals
		X	divided by the number of degrees of freedom. The
		X	time residuals are for those phases used to locate
		X	the event. The number of degrees of freedom is
		X	the number of defining observations (ndef) less
		X	the dimensions of the solution (eg 4 if depth is
		X	a free parameter, 3 if it is fixed).
smajax	%9.4f	X	length of semi-major axis of location error ellipse in km

sminax	%9.4f	X	length of semi-minor axis of location error ellipse in km
strike	%6.2f	X	strike of semi-major axis of location error ellipse in
		X	degrees cw from north
sdepth	%9.4f	X	max error of depth estimate in kilometers
stime	%8.2f	X	origin time uncertainty in sec.
conf	%5.3f	X	confidence (0 to 1) attached to smajax, sminax,
		X	sdepth, and stime
commid	%8d	X	unique id linking to comment
lddate	%17.5f	X	epoch time of last modification of record

II.6 Origin

The **origin** table contains information describing a computed origin for a particular event. Note that an event (identified by evid) is associated with waveforms, stations, channels, et cetera and if more than one origin (identified by orid) exists for an event then there are differing opinions as to the "true" hypocentral solution. This will result in more than one orid for a given evid.

field	format	NULL?	Description
lat	%9.4f		geographic latitude of solution in degrees (-90 to 90)
lon	%9.4f		geographic longitude of solution in degrees (-180 to 180)
depth	%9.4f		depth to hypocenter from surface (eg ¿0) in km
time	%17.5f		epoch time (seconds past 0 hr of 1 JAN 1970)
			I guess well need another digit of accuracy in
			200 years or so
orid	%8d		unique id for the origin

evid	%8d		unique id for the event
jdate	%8d		julian date, note that counting begins with 1 not 0 so jdate for january 1, 1970=1970001 and Jan 1 0001 AD is 0001001 (I guess there is no 0000 AD?), and only the year goes backward BC so jan 10, 10BC is -0010010
nass	%4d	X	number of arrivals associated with this origin
ndef	%4d	X	number of arrivals used to compute this origin
ndp	%4d	X	number of depth phases used to compute this hypocenter
grn	%8d	X	geographic region number after Flinn, Engdahl and Hill (BSSA v64, pp 771-992, 1974).
srn	%8d	X	seismic region number see above reference.
etype	%7s	X	type of seismic event: local (l), regional (r), teleseism (l), explosion (e).
depdp	%9.4f	X	estimated depth from depth phases.
dtype	%1s	X	flage indicating method of calculating depth such as f (free parameter), d (from depth phases), r (restrained by location program), g (restrained by geophysicist).
mb	%7.2f	X	body wave magnitude
mbid	%8d	X	unique id linking to netmag
ms	%7.2f	X	surface wave magnitude
msid	%8d	X	unique id linking to netmag
ml	%7.2f		local magnitude such as duration magnitude or richter

mlid	%8d	X	unique id linking to netmag
algorithm	%15s	X	brief description of algorithm used for processing eg computing a seismic origin or a beam
auth	%15s		originator of origin eg NEIC, seismos, balch
commid	%8d	X	unique id linking to comment
lddate	%17.5f		epoch time of last modification of this record

II.7 Sensor

The **sensor** table provides record of updates for calibration factor or clock error of a specific instrument.

field	format	NULL?	Description
sta	%6s	X	common station code name
chan	%s	X	common channel code name
time	%17.5f	X	first epoch time update is valid
endtime	%17.5f	X	last epoch time update is valid
inid	%8d	X	unique id for the instrument type
chanid	%8d	X	unique id linking to sitechan table
jdate	%8d	X	julian date of time
calratio	%16.6f	X	dimensionless calibration correction factor permitting small refinements to calib and calper
calper	%16.6f	X	period in sec for which calib, ncalib, and calratio are valid
tshift	%6.2f	X	corrects discrepencies between actual time and recorded time, actual time=recorded time+tshift in seconds

instant	%1s	X	y means snapshot of at time of discrete procedural change and n means snapshot of continuously changing drift
lddate	%17.5f	X	epoch time of last modification of this record

II.8 Site

The **site** table contains station location information describing a point on the earth such as an instrument location, array location, or reference location.

field	format	NULL?	Description
sta	%6s		commone station code name
ondate	%8d		julian start date when station etc began operation
offdate	%8d		julian end date when staion was pulled, moved, etc
lat	%9.4f		geographic latitude in degrees (-90.0 to 90.0)
lon	%9.4f		geographic longitude in degrees east of prime meridian (-180.0 to 180.0)
elev	%9.4f		elevation above mean sea level in km
staname	%50s		full name of station eg ANMO=Albuquerque, New Mexico
statype	%4s		station type eg ss (single station), ar (array), ss3c (single station 3 component), ss1c (single station single component)
refsta	%6s	X	specifies reference station for relative array array element locations
dnorth	%9.4f	X	relative position of array element in km north of refsta
deast	%9.4f	X	relative position of array element in km east of refsta
lddate	%17.5f		epoch time of last modification of this record

II.9 Sitechan

The **sitechan** table provides information about various channels of a station.

field	format	NULL?	Description
sta	%6s		common station code name
chan	%8s		common channel code name
ondate	%8d		year and julian day when this record became valid
chanid	%8d		unique id for this sta and chan (seems a bit renundant with sta, chan, ondate, offdate doesn't it)
offdate	%8d	X	julian date when this record is no longer valid
ctype	%4s		type of data channel: n (normal, a normal instrument response), b (beam, a coherent beam formed from array), i (incoherent beam or energy stack)
edepth	%9.4f		depth of instrument in km relative to elev in site table
hang	%6.1f		horizontal orientation of the instrument in degrees clockwise from north
vang	%6.1f		vertical orientation of the instrument in degrees from up
descrip	%50s		description of the data channel
lddate	%17.5f		epoch time of last modification of this record

II.10 Wfdisc

The **wfdisc** provides a pointer to the waveform files stored on disk.

field	format	NULL?	Description
sta	%6s		common station code name
chan	%8s		common channel code name

time	%17.5f		epoch time of start time of trace
wfid	%8d	X	unique id for segment of digital data
chanid	%8d		unique id linking to sitechan
jdate	%8d		julian date of trace start
endtime	%17.5f	X	epoch time of last sample in trace
nsamp	%8d		number of samples in trace
samprate	%11.7f		sample rate of data in samples/sec
calib	%16.6f	X	conversion factor to ground motion (nm/s)/count for a velocity sensor
calper	%16.6f	X	period of validity of calib in sec
instype	%6s		instrument type eg SRO,WWSSN etc
segtype	%1s		natural units of sensor (A, V, D). This was originally used to indicate whether the data were segmented, virtual original etc, but JSPC change this.
datatype	%2s		format of binary data storage s2 (short integer), s4 (long integer), t4 (32 bit float) in Sun format Note that CSS standard defines many other formats not supported by JSPC software
clip	%1s	X	trace is either clipped (c) or not (n).
dir	%64s		directory path of trace (can be relative but be careful)
dfile	%32s		filename of the waveform file
foff	%10d		byte offset of trace start segment generally used when data are multiplexed

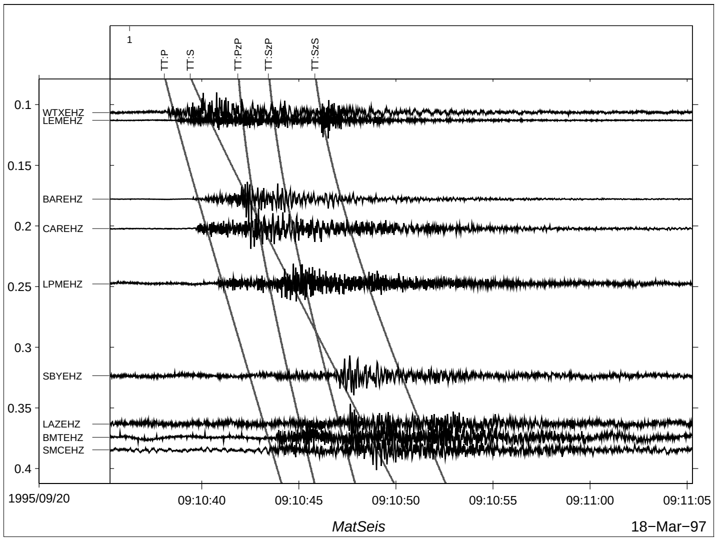
commid	%8d	X	unique id linking to a comment record
lddate	%17.5f		epoch time of last modification of this record

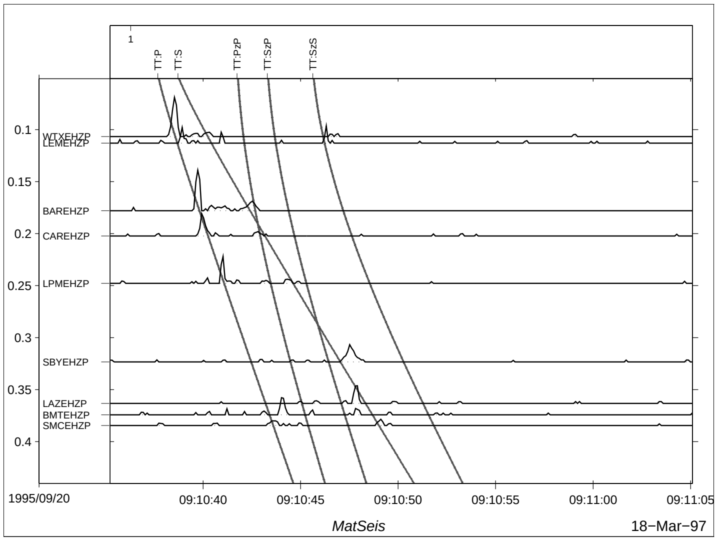
Appendix III. Additional Event Figures

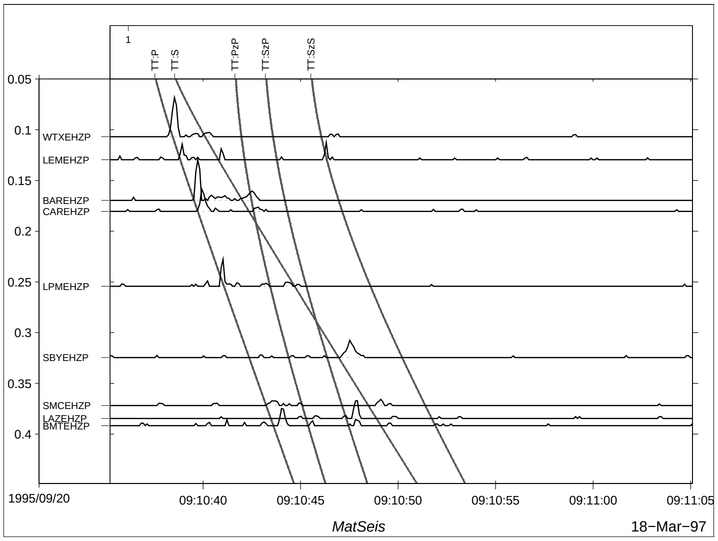
Data plots are provided for the 16 events in the Primary Test Set and for the 21 events in the Expanded test set, as a means of qualitatively judging the quality of the location. In all plots, time in seconds is given on the horizontal axis, and distance in degrees is given on the vertical axis. Traces are plotted vertically to represent distance from the event hypocenter. Raw traces are at the top, preprocessed traces with respect to the *LWCEDS* hypocenter are plotted in the middle, and preprocessed traces with respect to the *Seismos* hypocenter are on the bottom. In some respects, these data plots provide more information than the difference tables (4.3 and 4.6).

In all plots the raw data have been trace normalized for amplitude (i.e. decreasing signal amplitude with distance is not apparent) and the processed data have been normalized by the largest amplitude trace (i.e. decreasing signal-to-noise ratio with distance, when appropriate, is apparent).

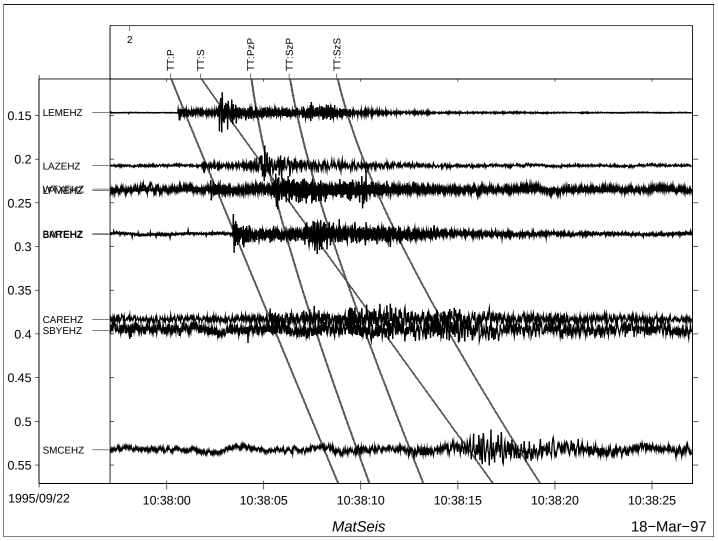
Primary Set Evid=1

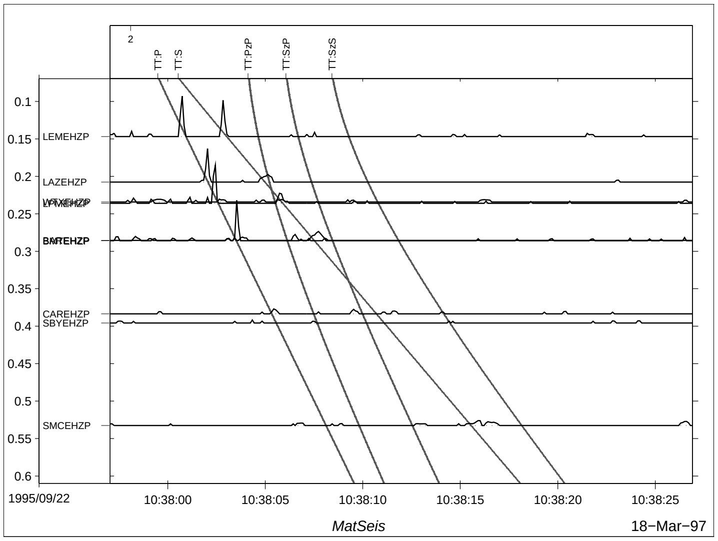


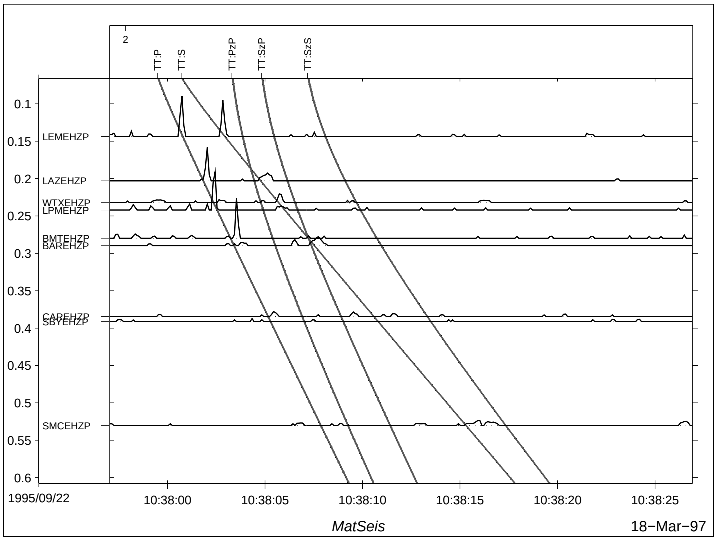




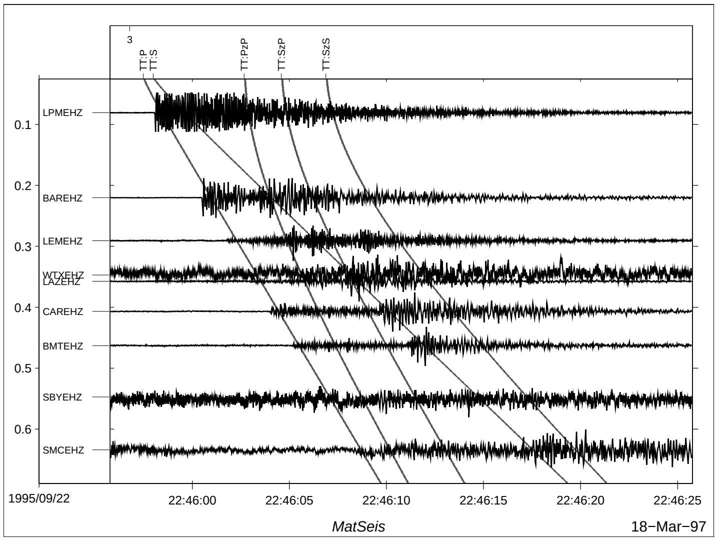
Primary Set Evid=2

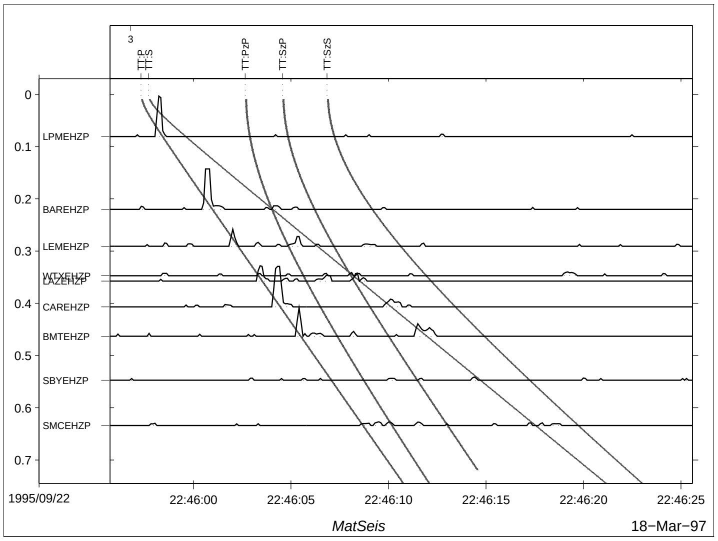


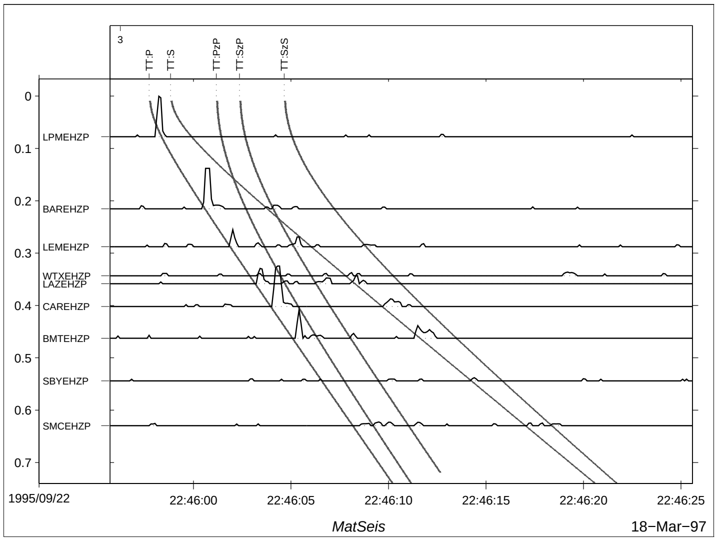




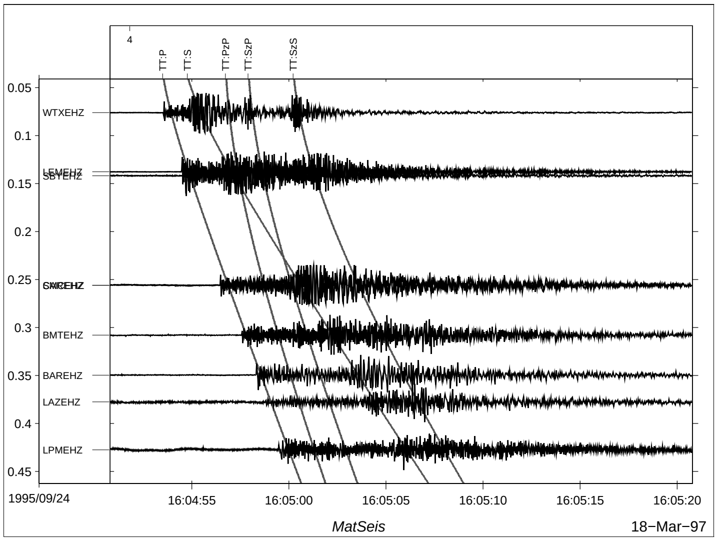
Primary Set Evid=3

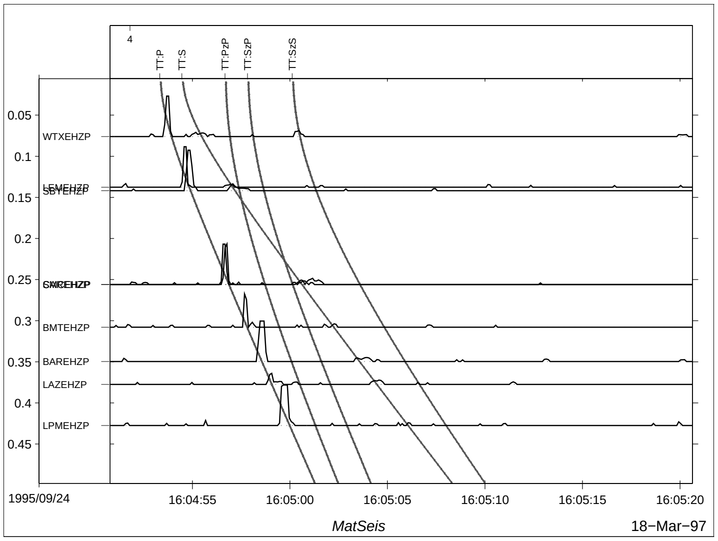


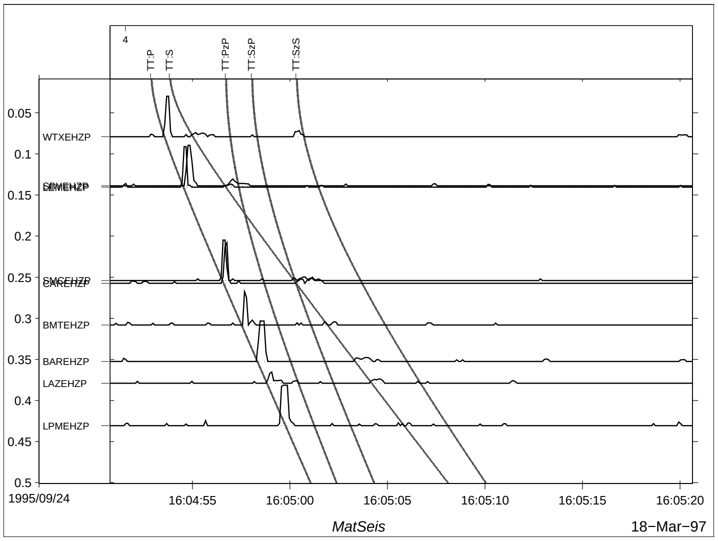




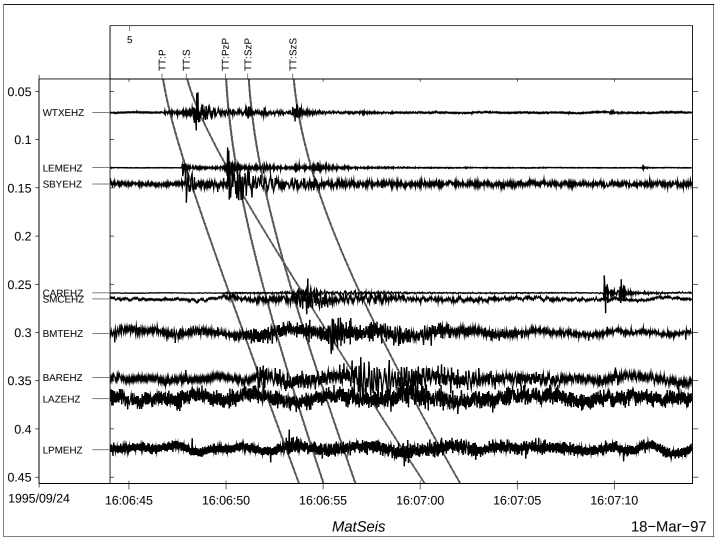
Primary Set Evid=4

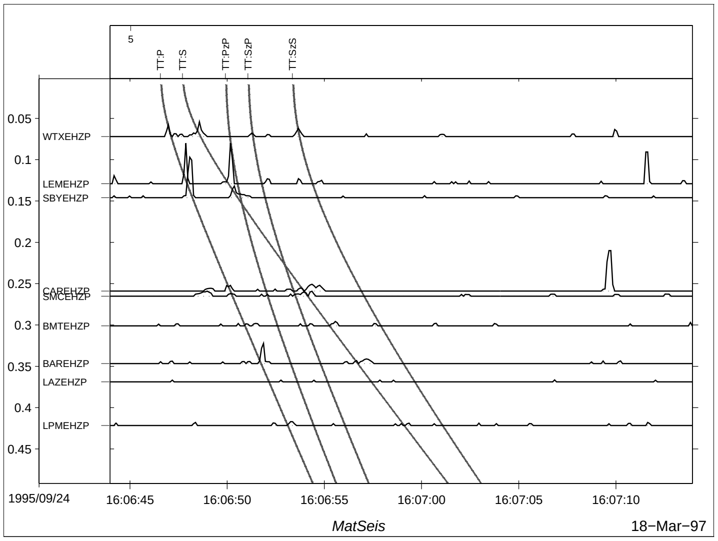


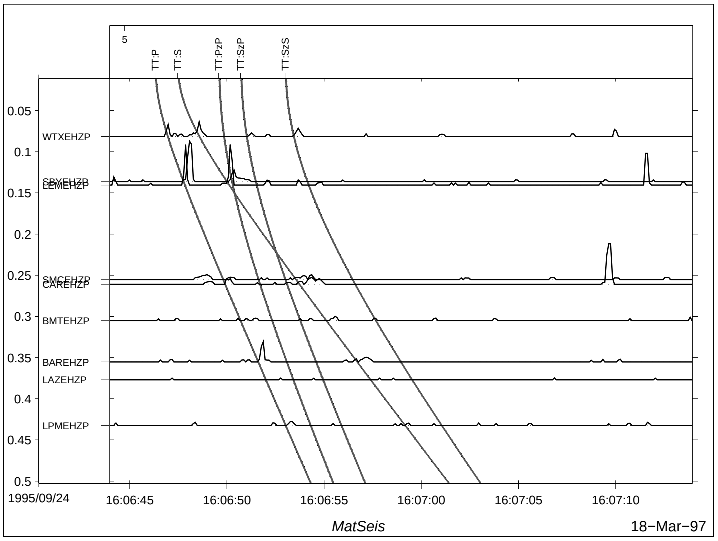




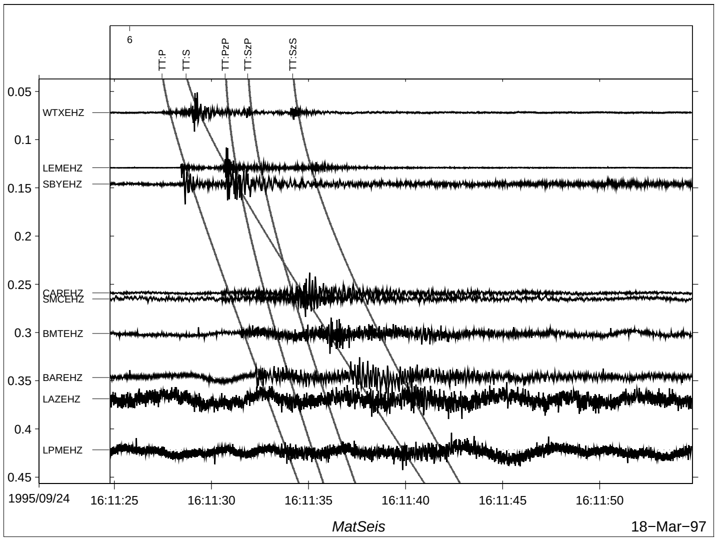
Primary Set Evid=5

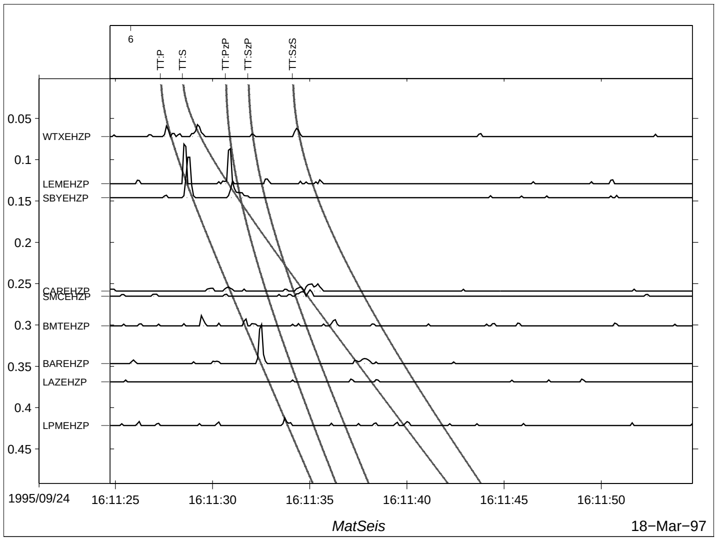


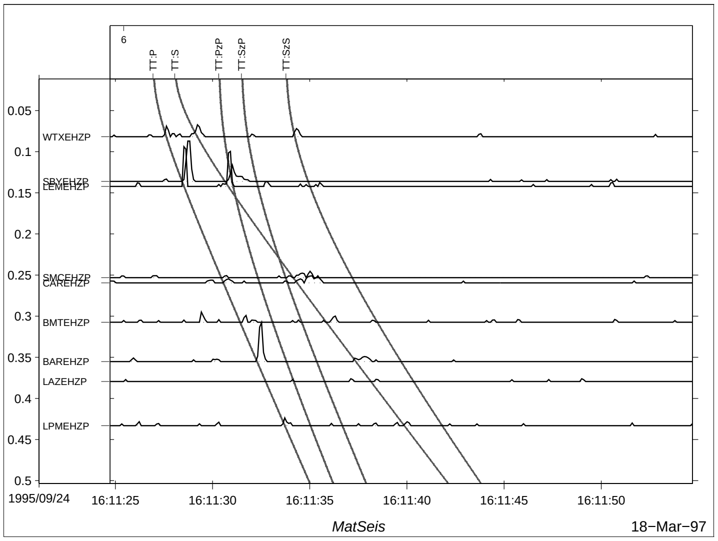




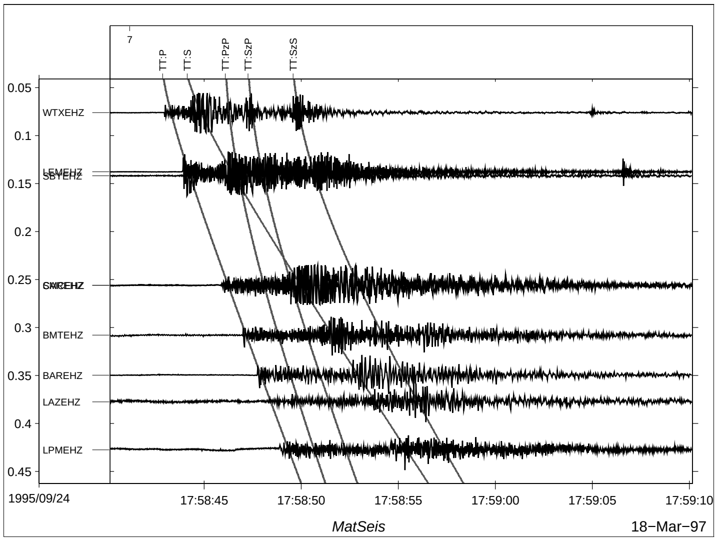
Primary Set Evid=6

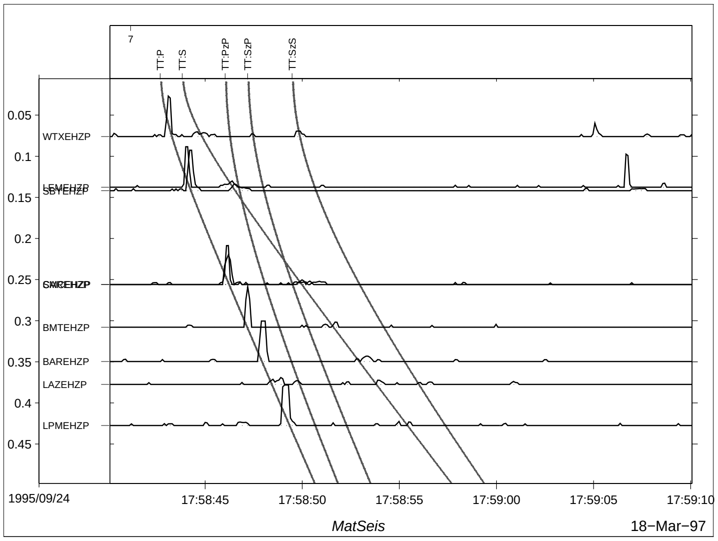


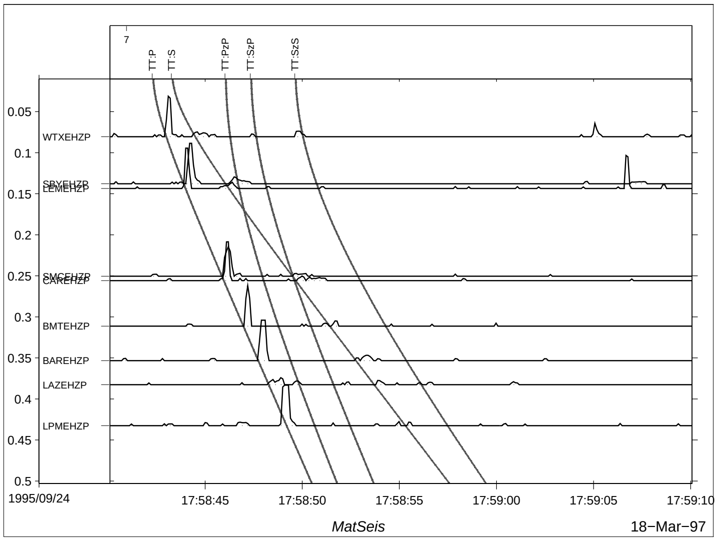




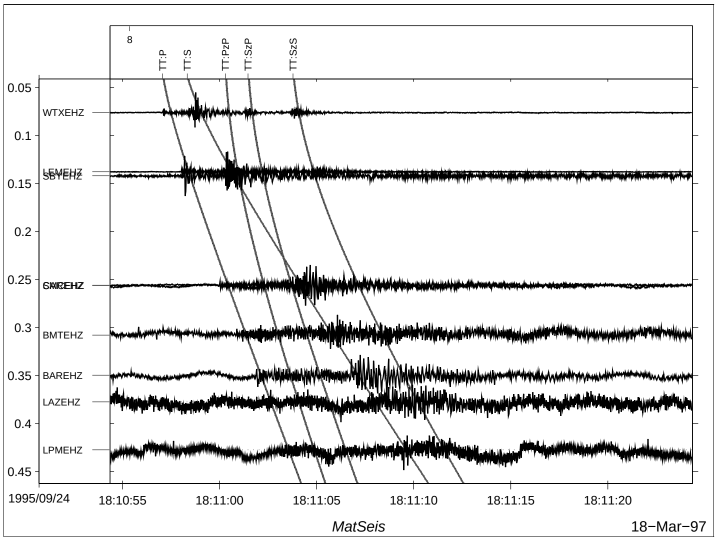
Primary Set Evid=7

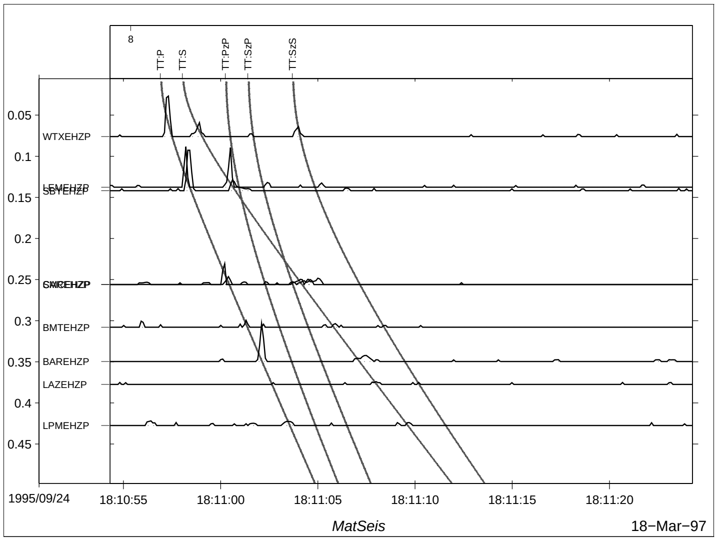


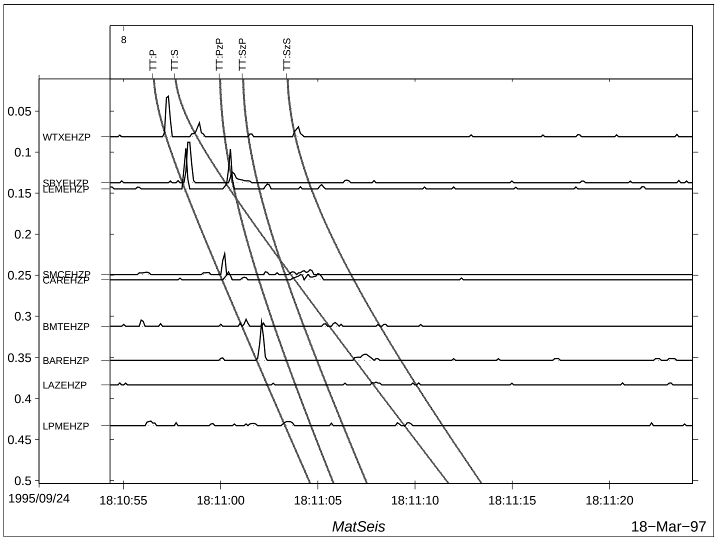




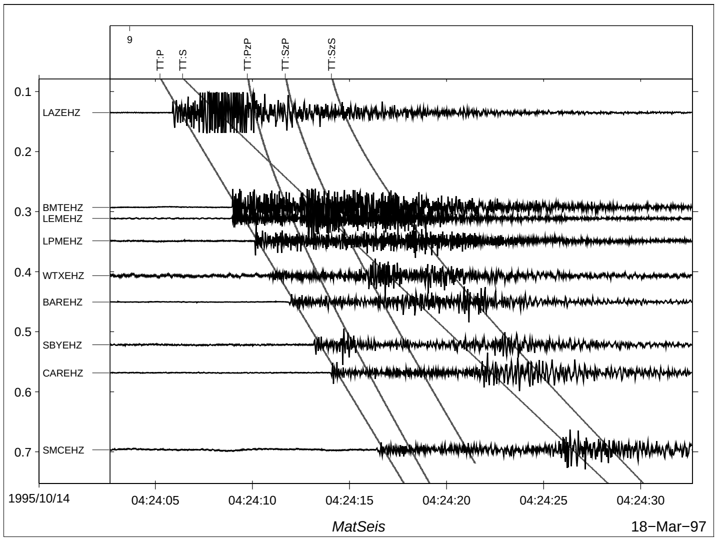
Primary Set Evid=8

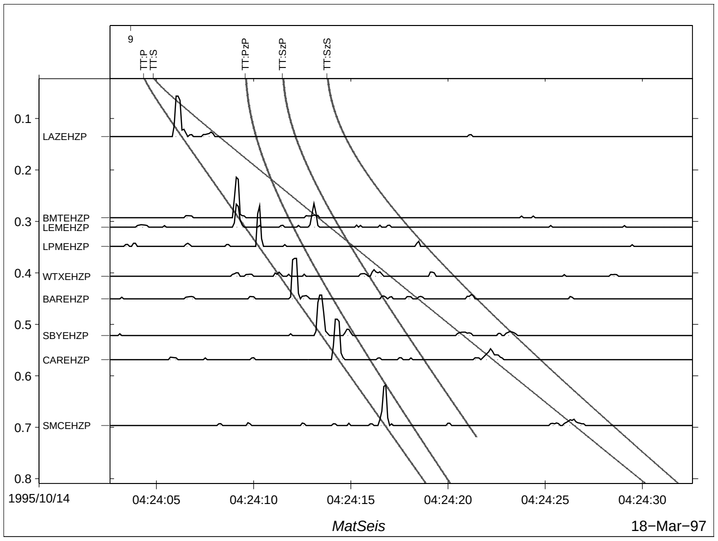


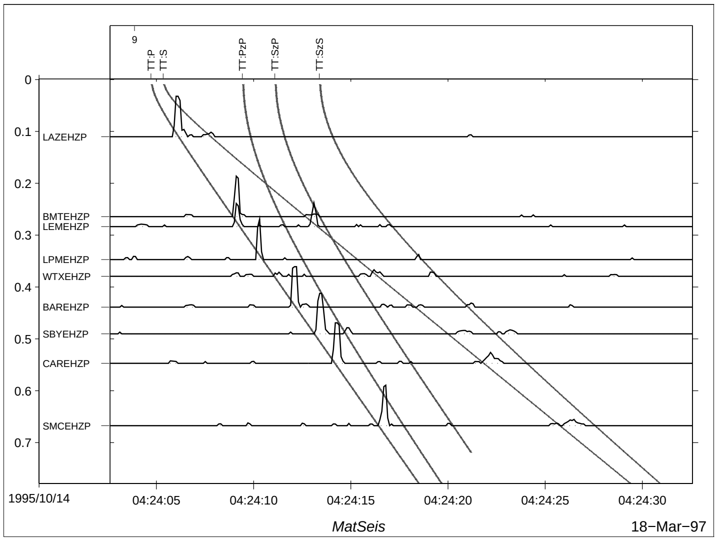




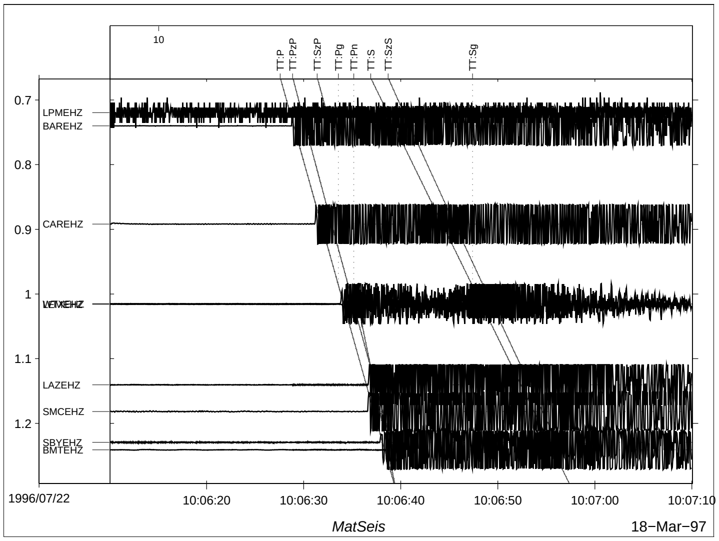
Primary Set Evid=9

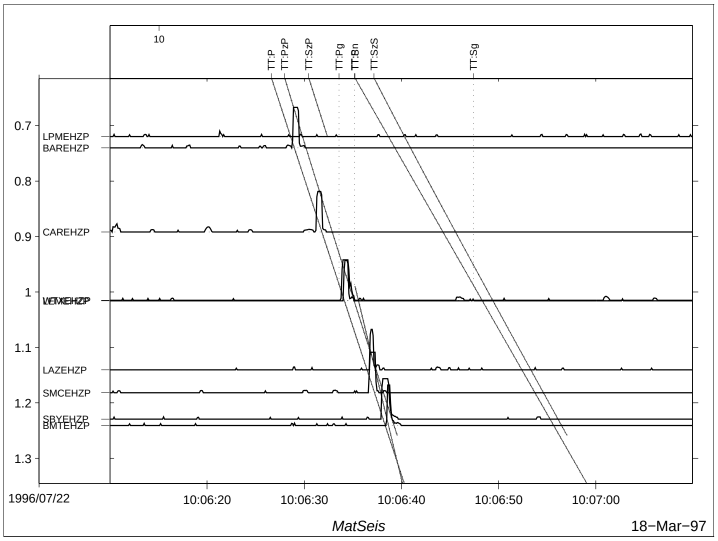


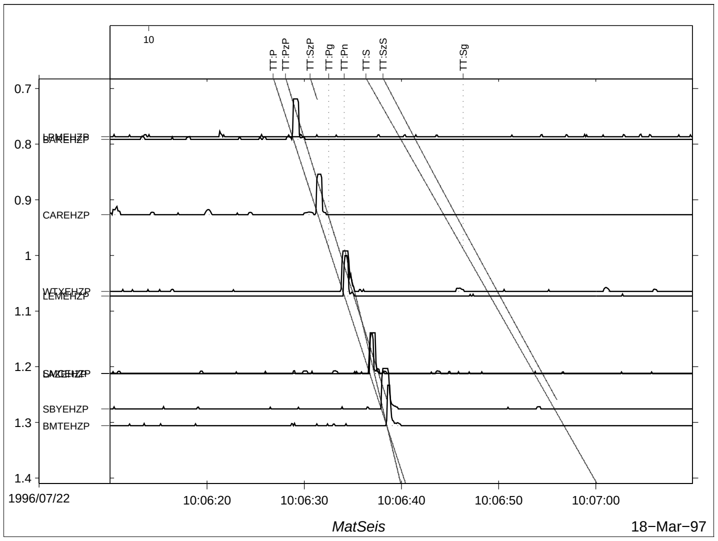




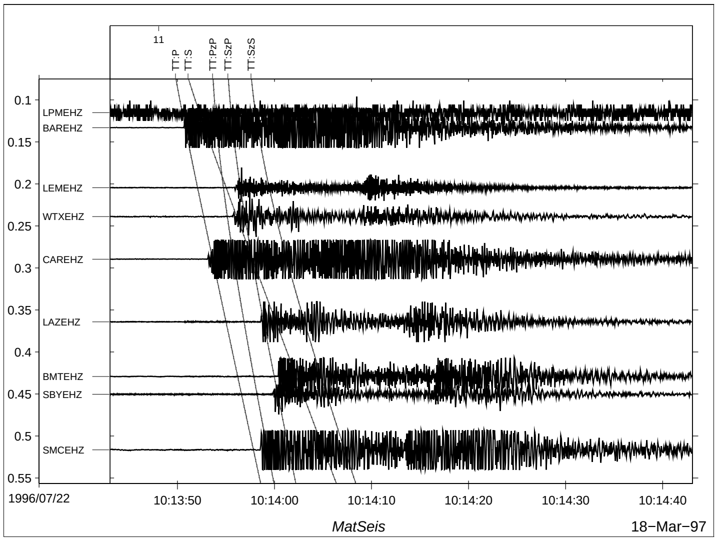
Primary Set Evid=10

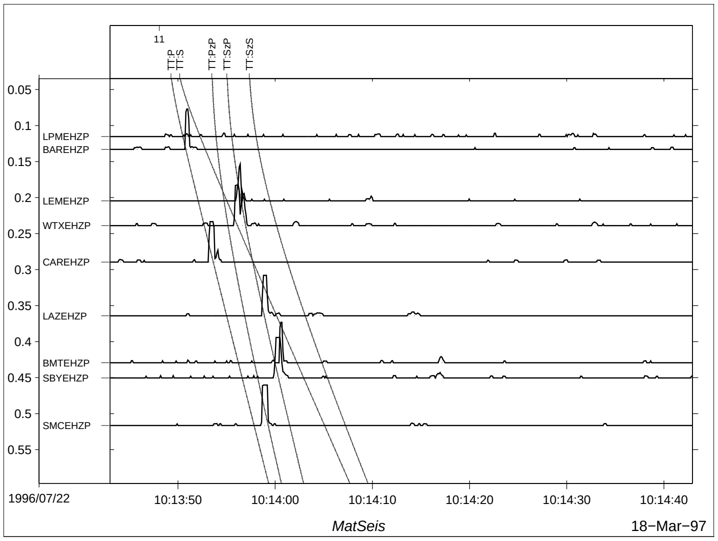


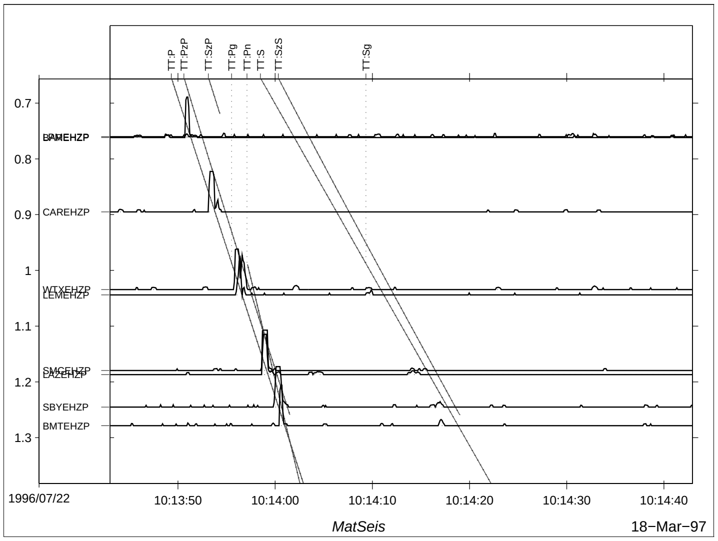




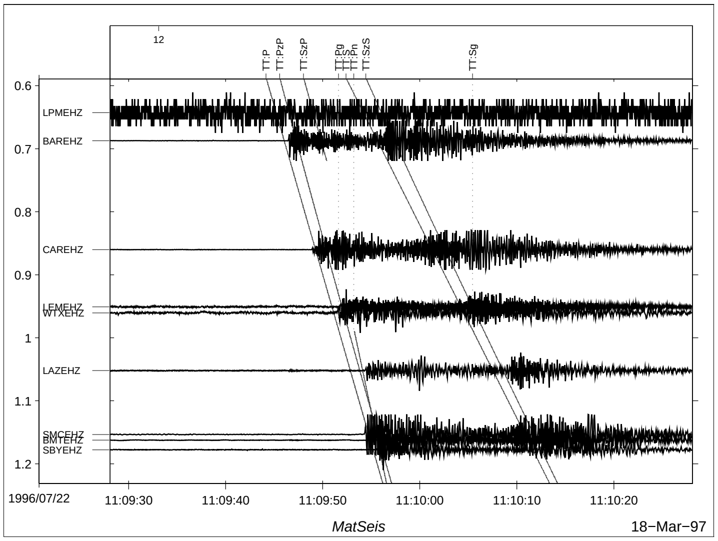
Primary Set Evid=11

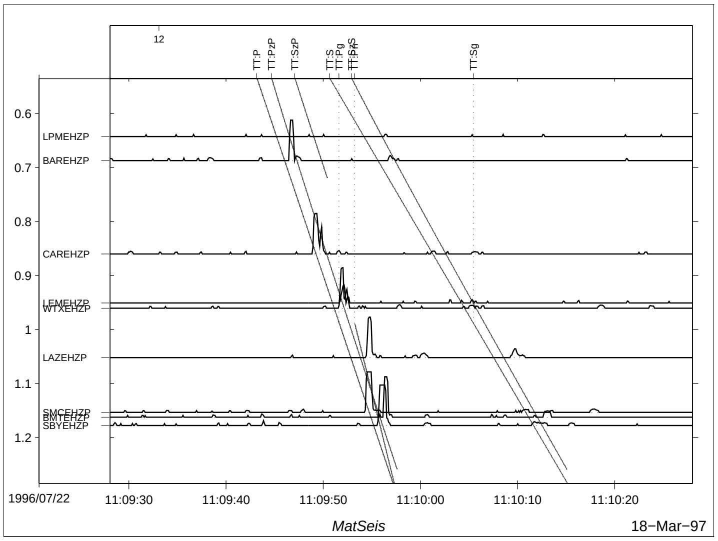


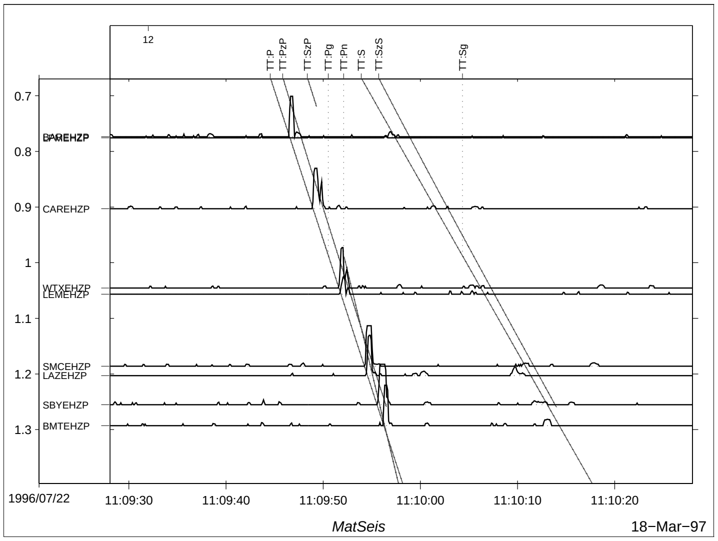




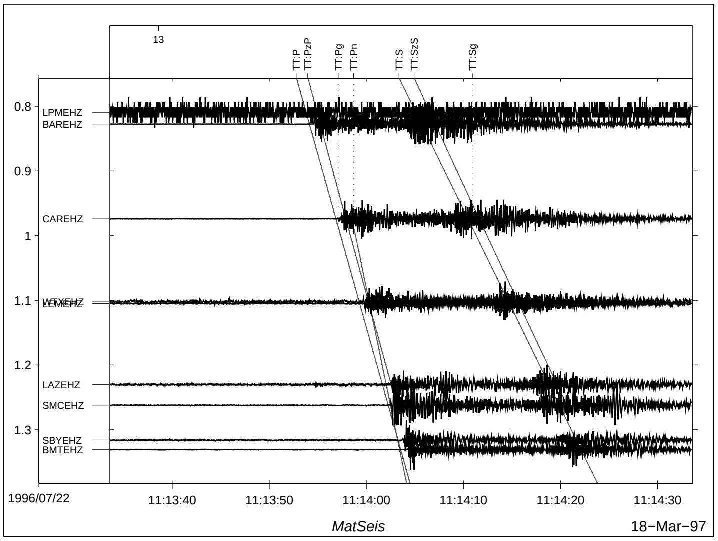
Primary Set Evid=12

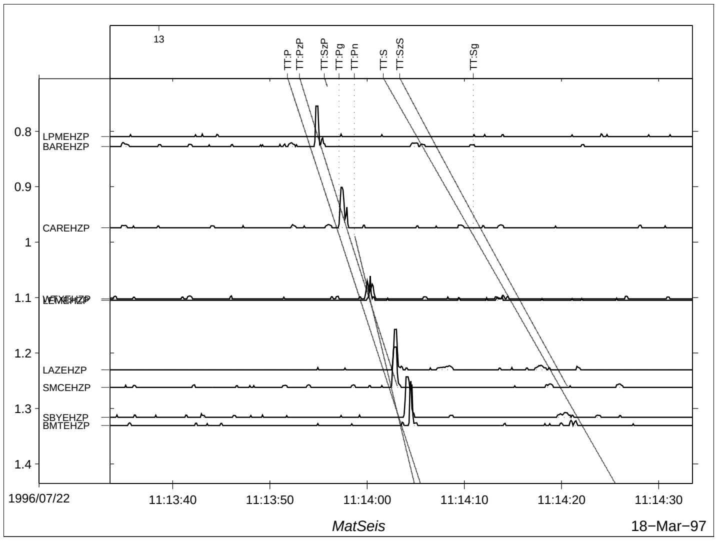


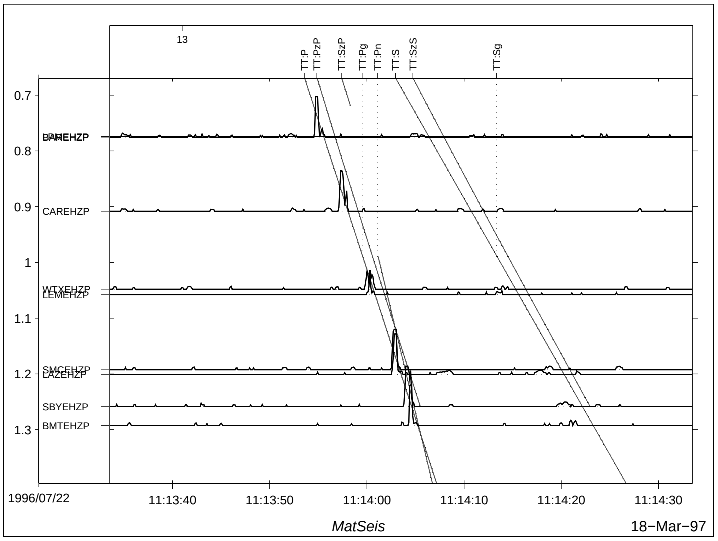




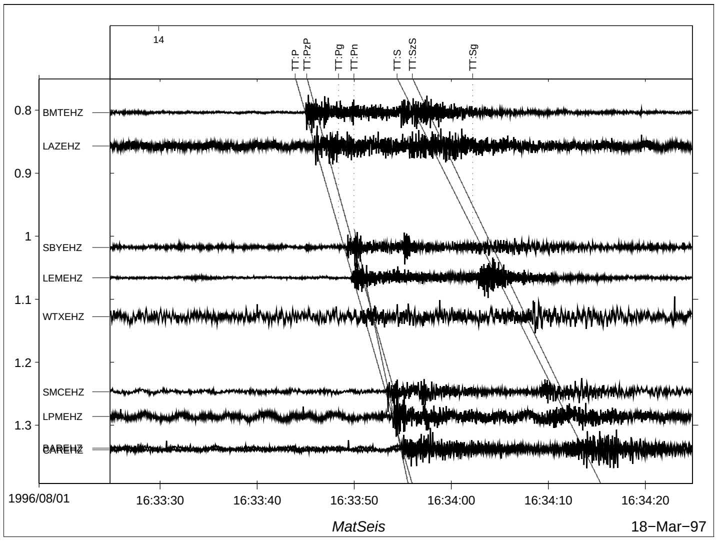
Primary Set Evid=13

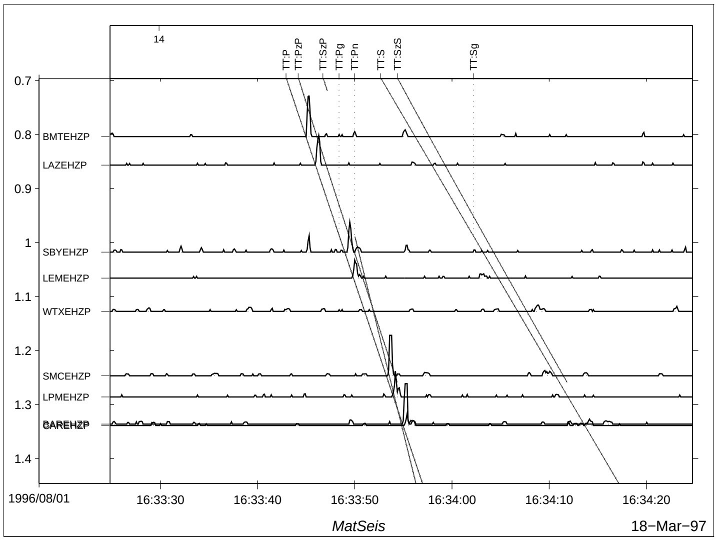


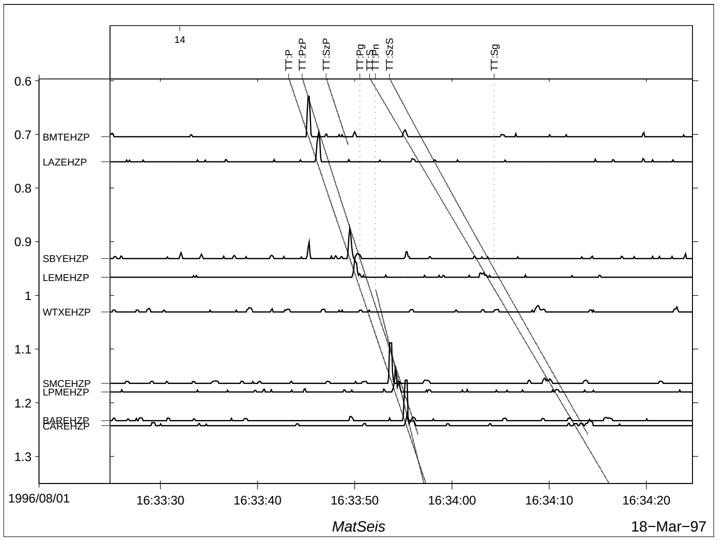




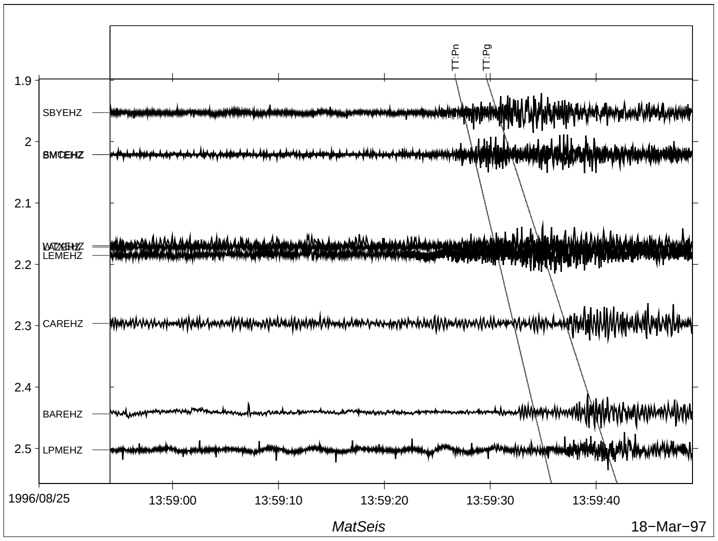
Primary Set Evid=14

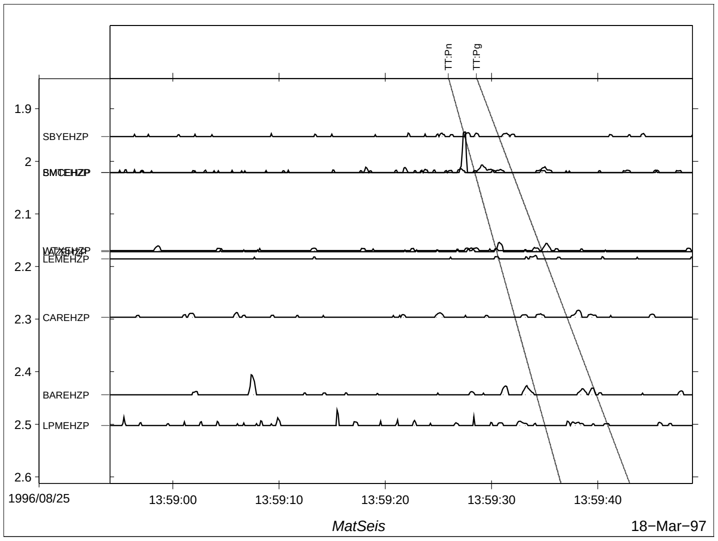


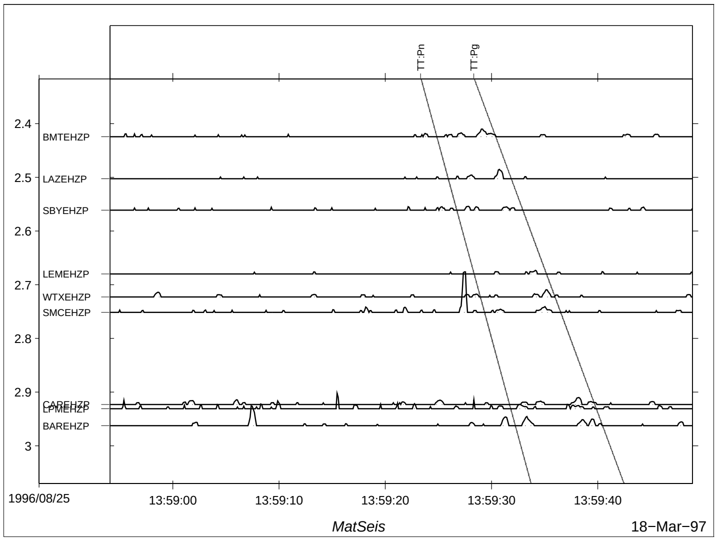




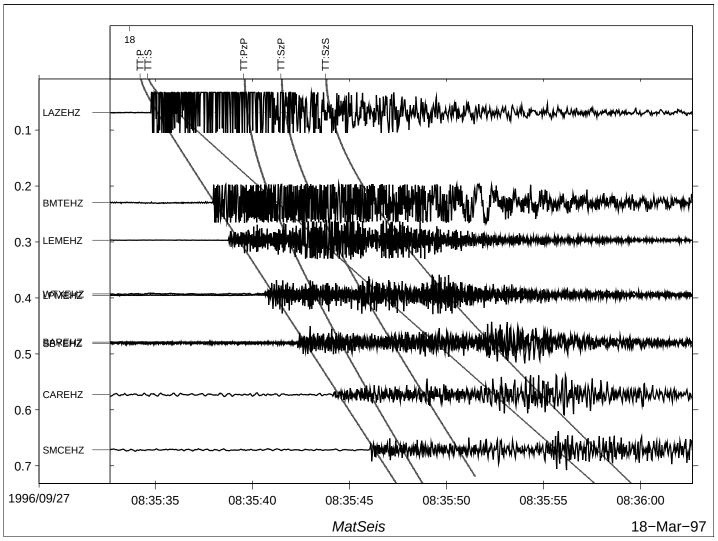
Primary Set Evid=15

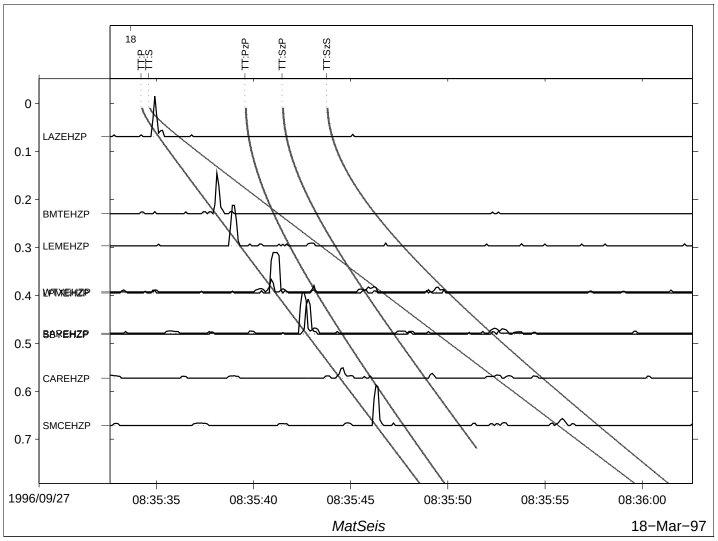


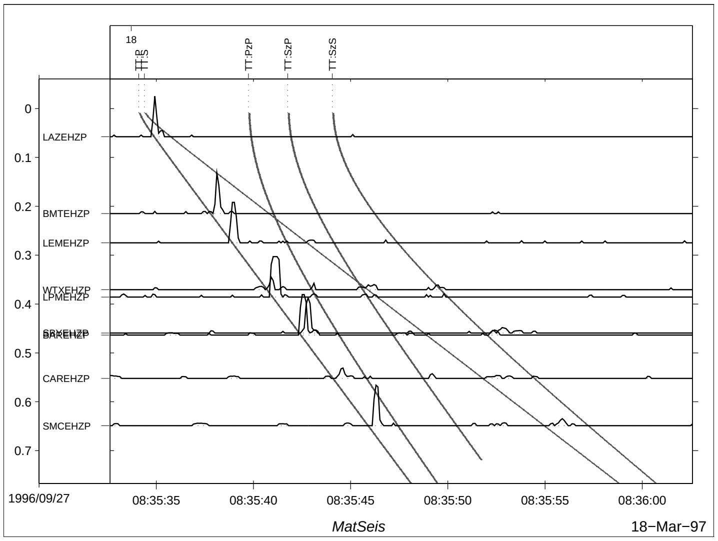




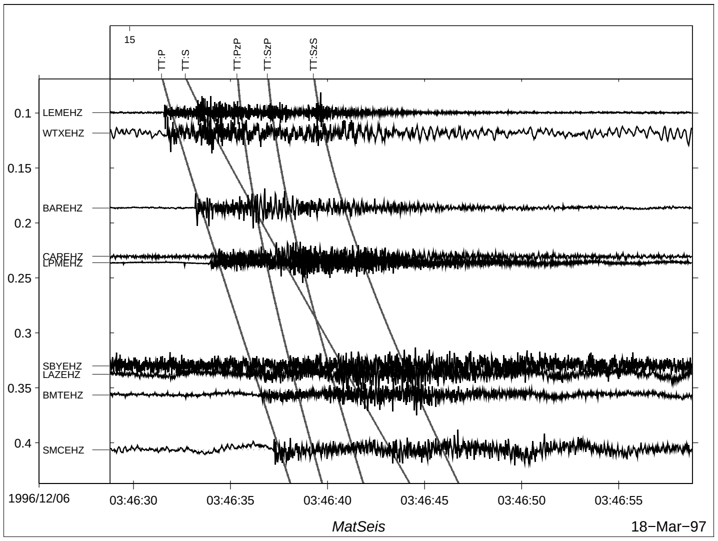
Primary Set Evid=16

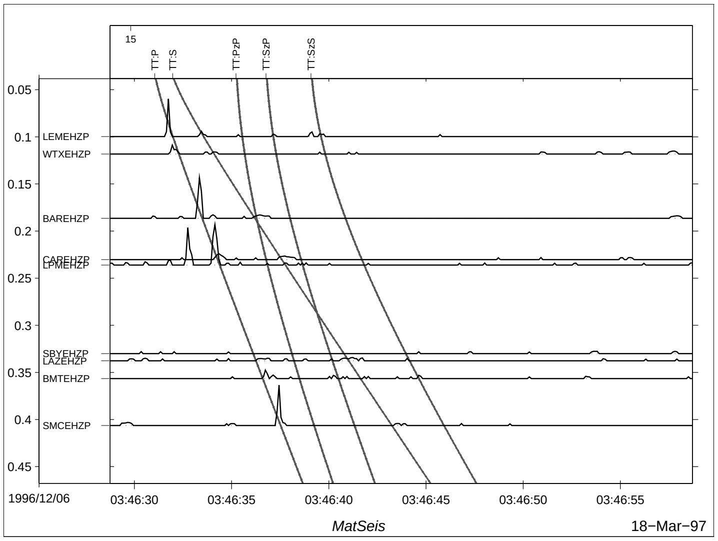


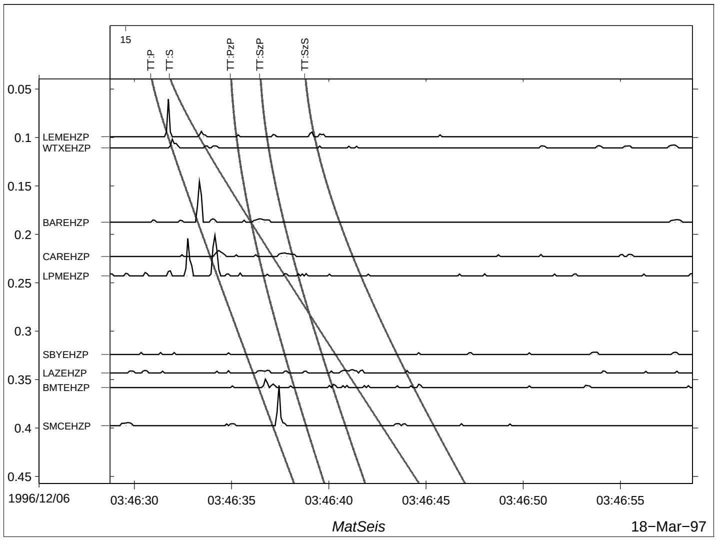




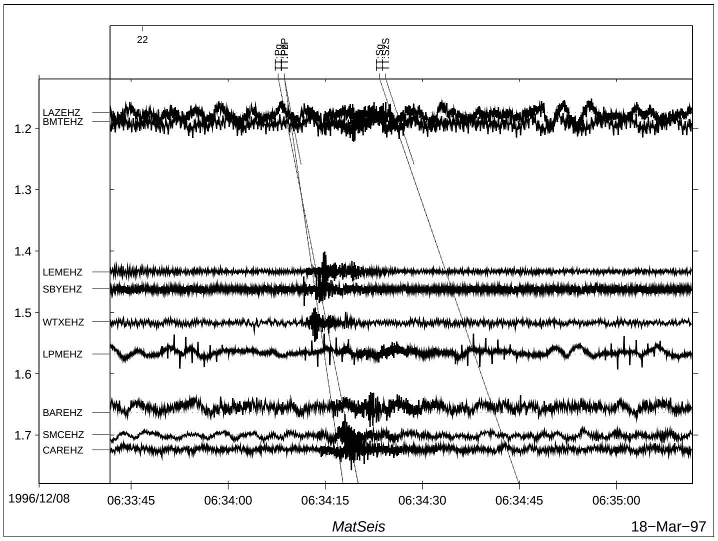
Expanded Set Evid=15

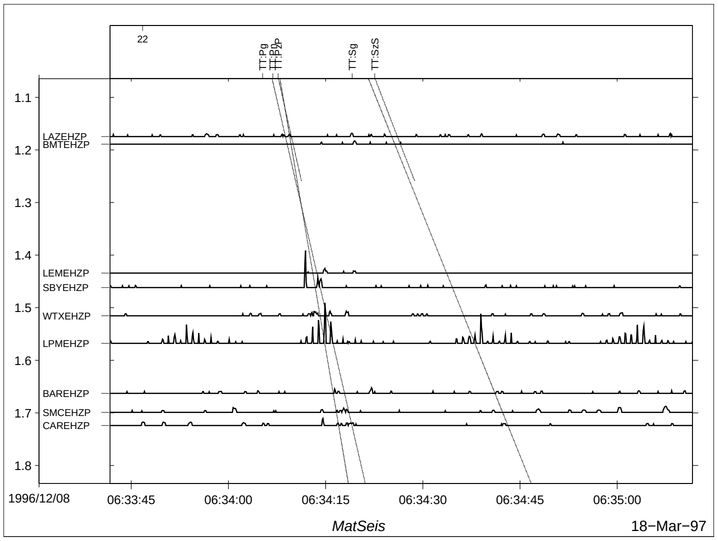


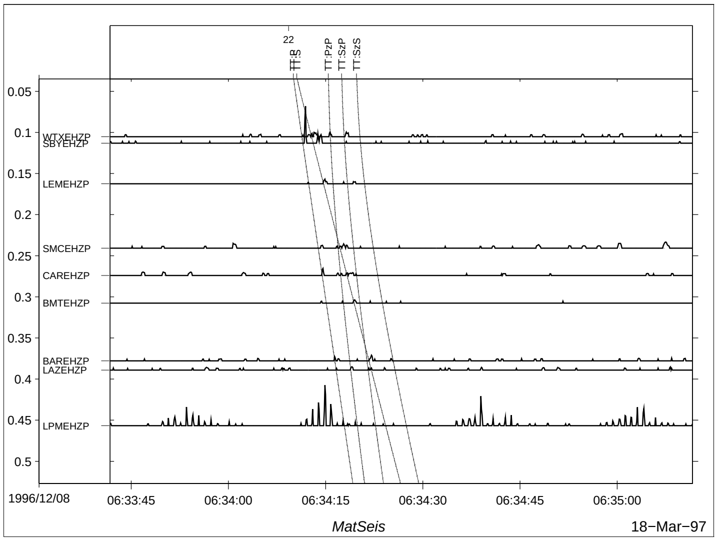




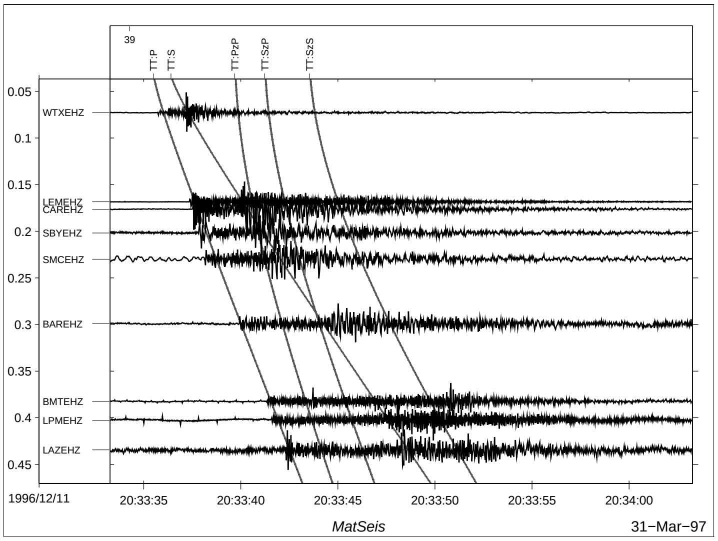
Expanded Set Evid=22

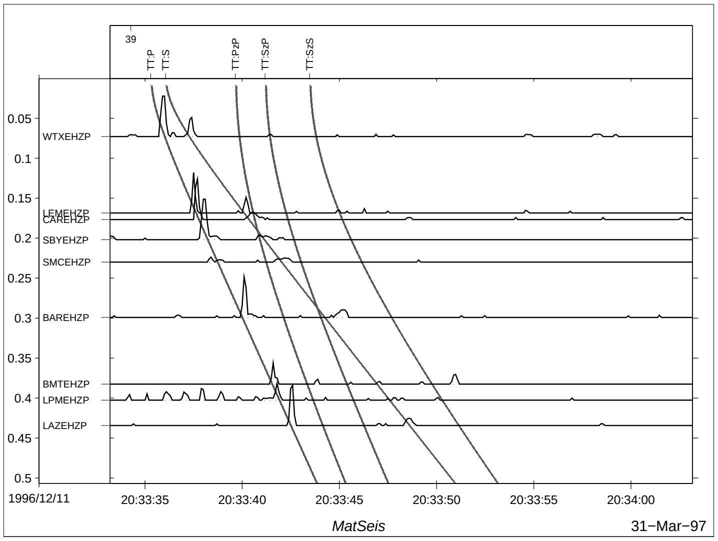


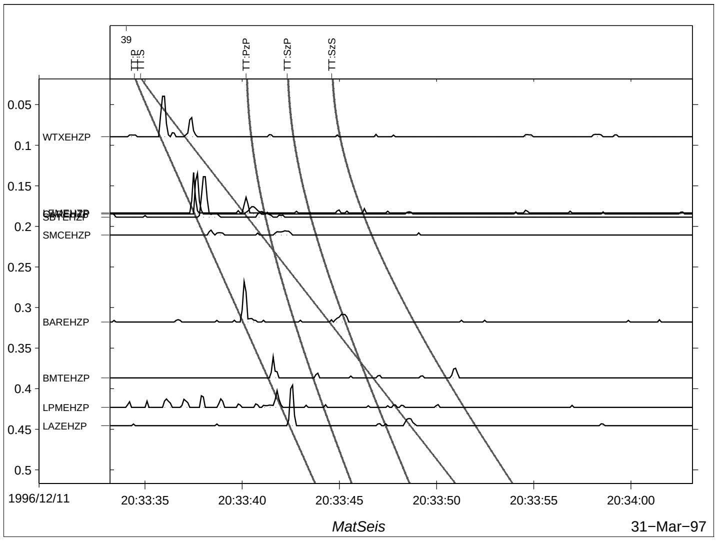




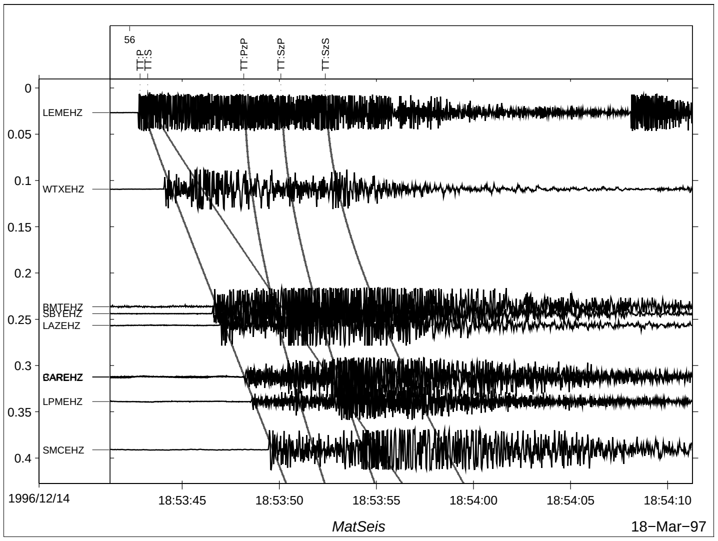
Expanded Set Evid=39

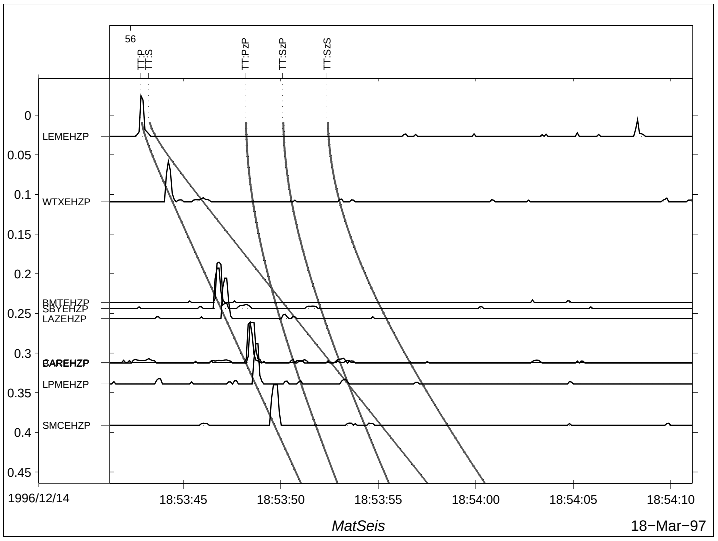


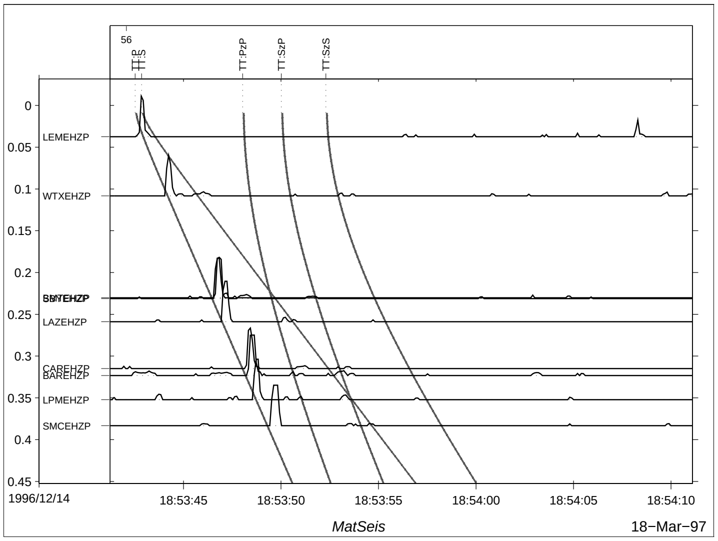




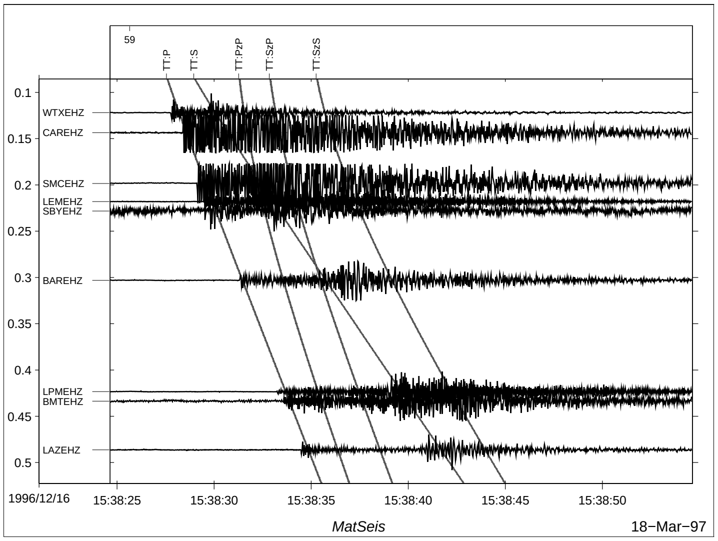
Expanded Set Evid=56

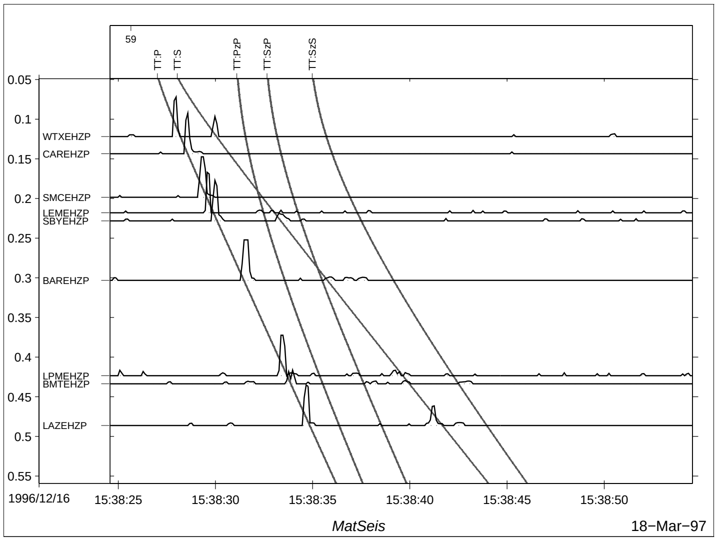


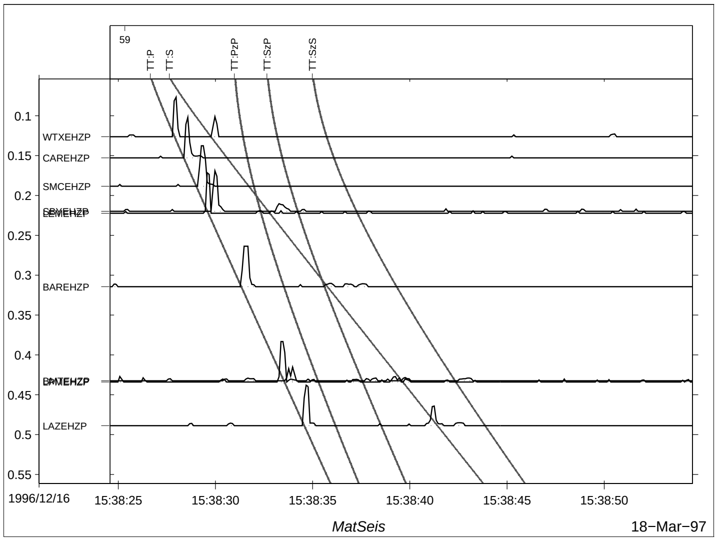




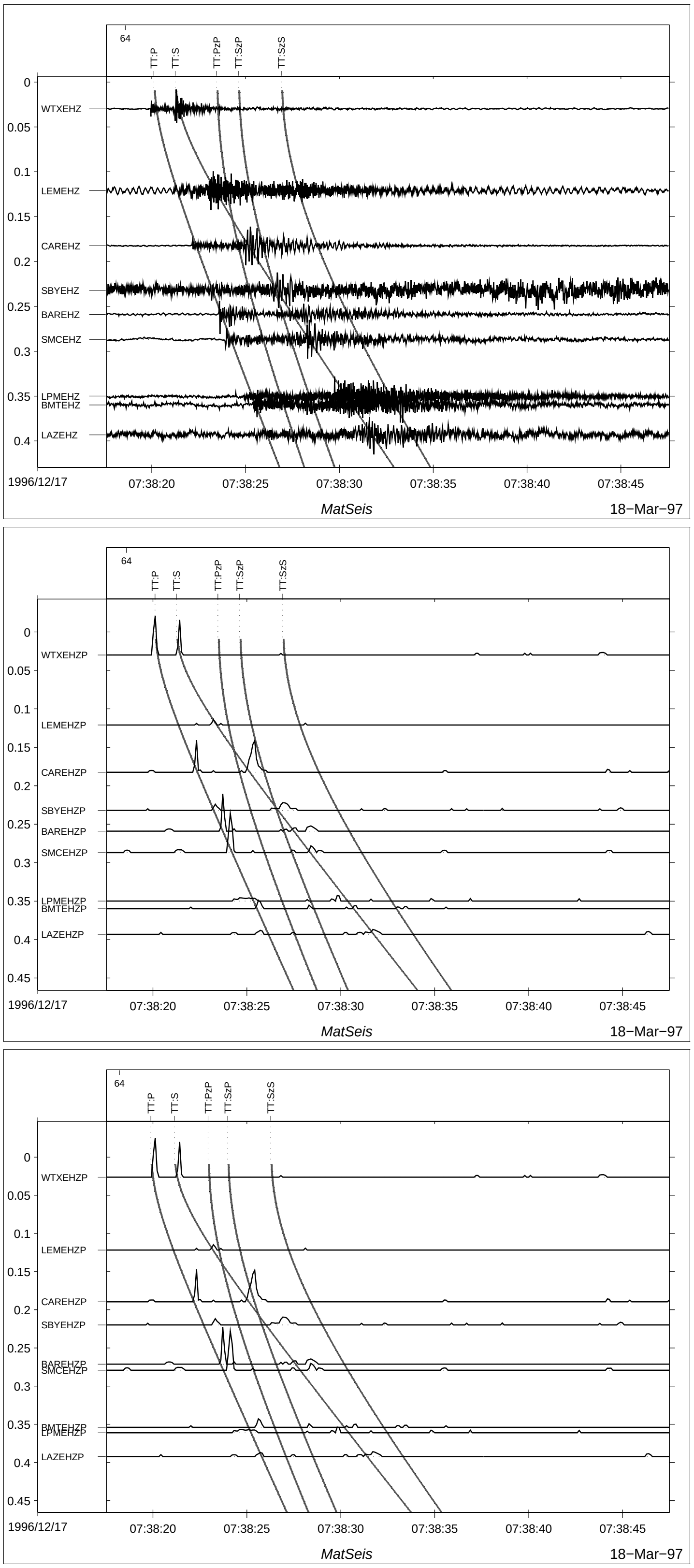
Expanded Set Evid=59



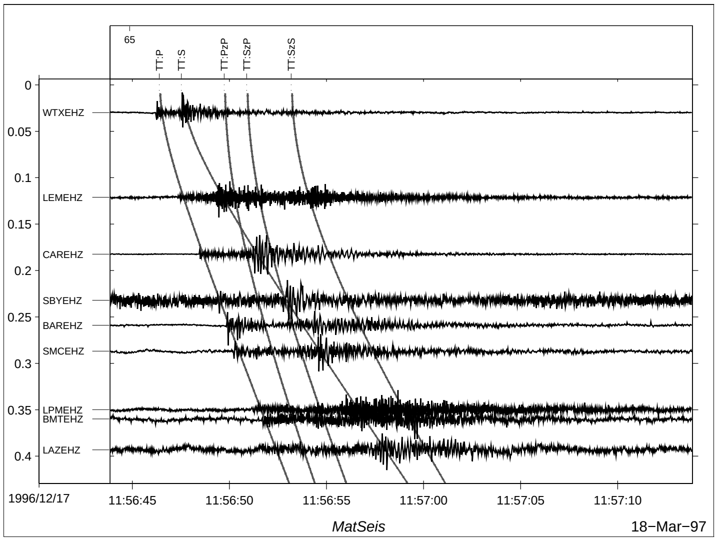


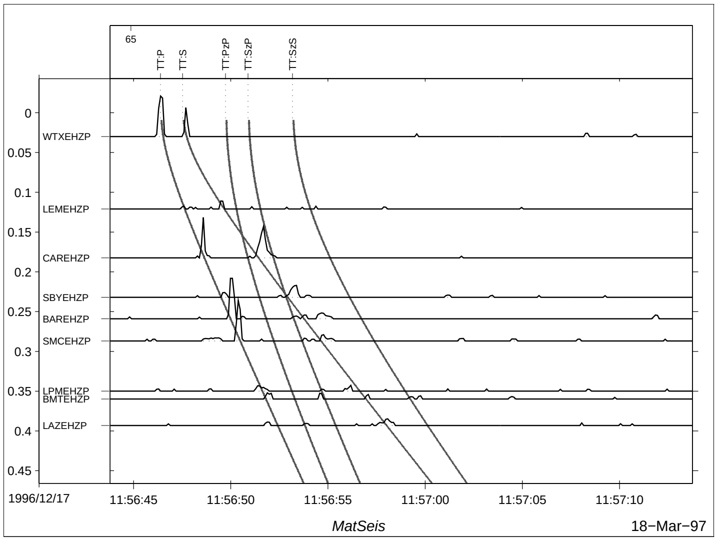


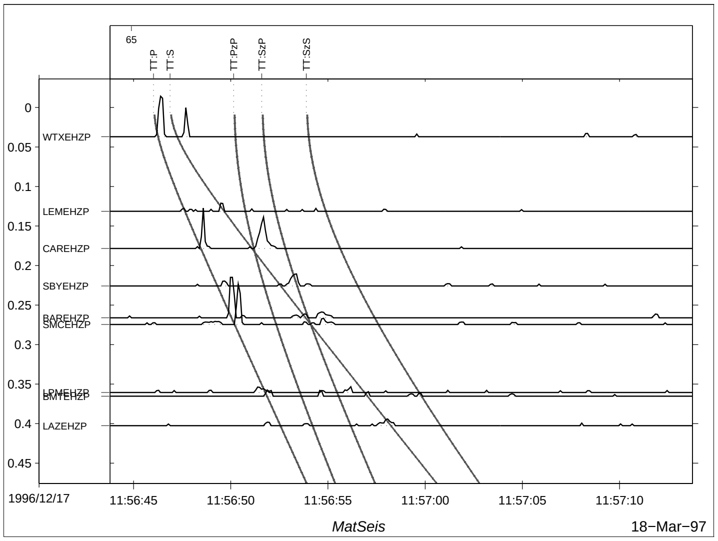
Expanded Set Evid=64



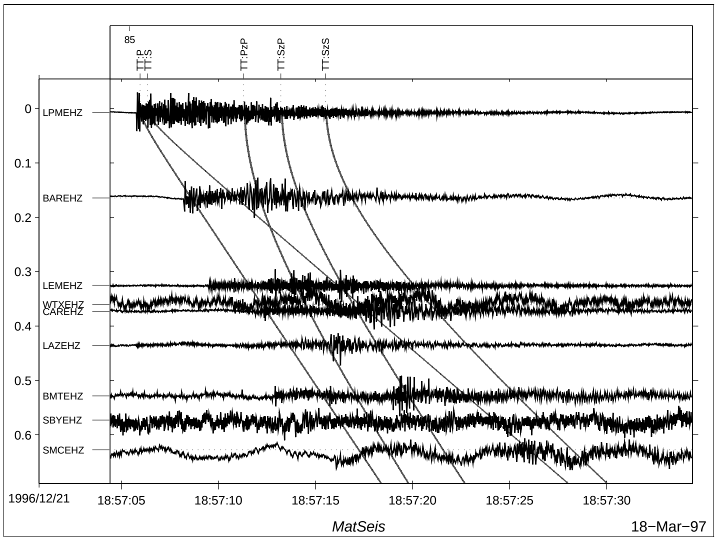
Expanded Set Evid=65

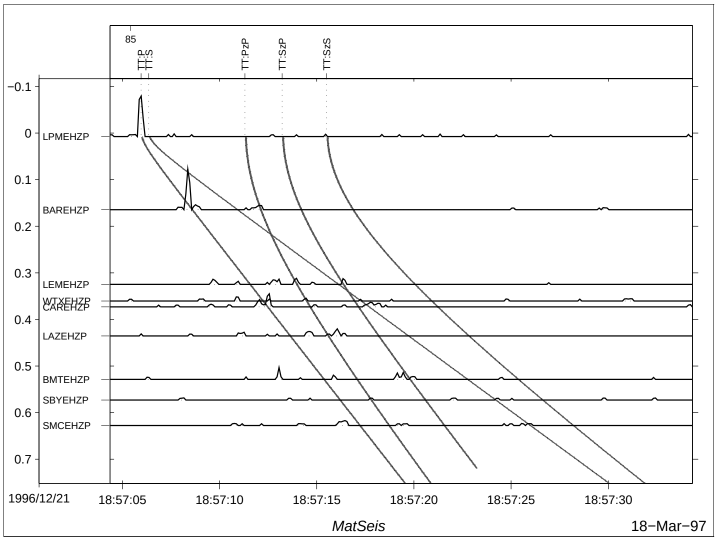


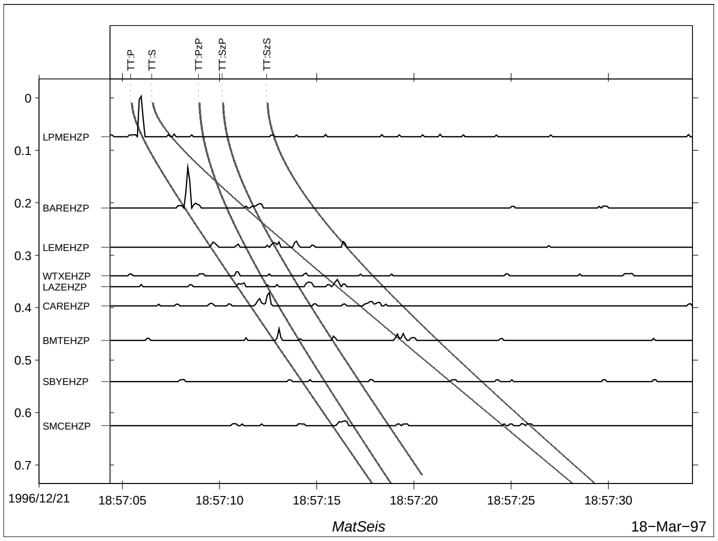




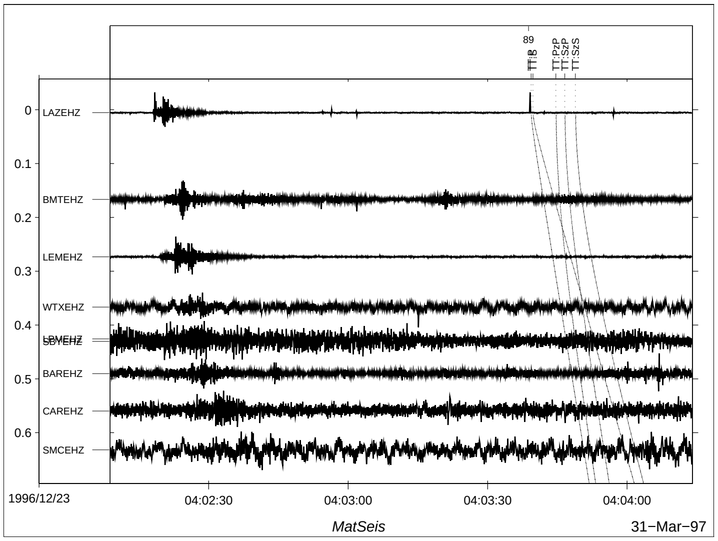
Expanded Set Evid=85

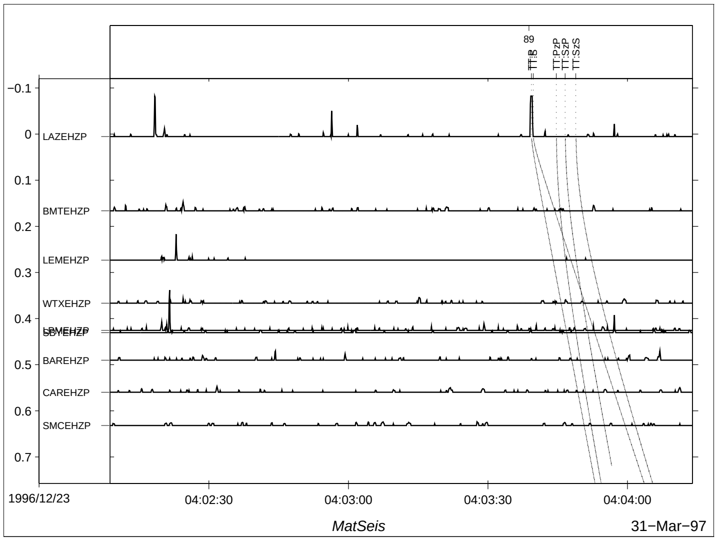


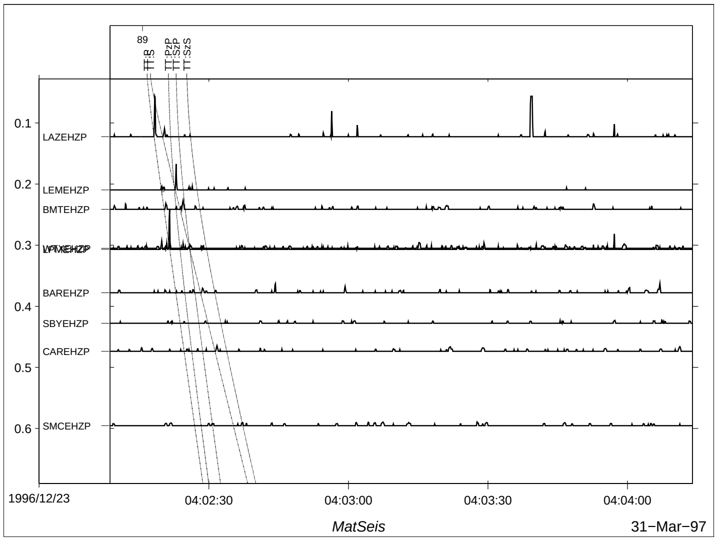




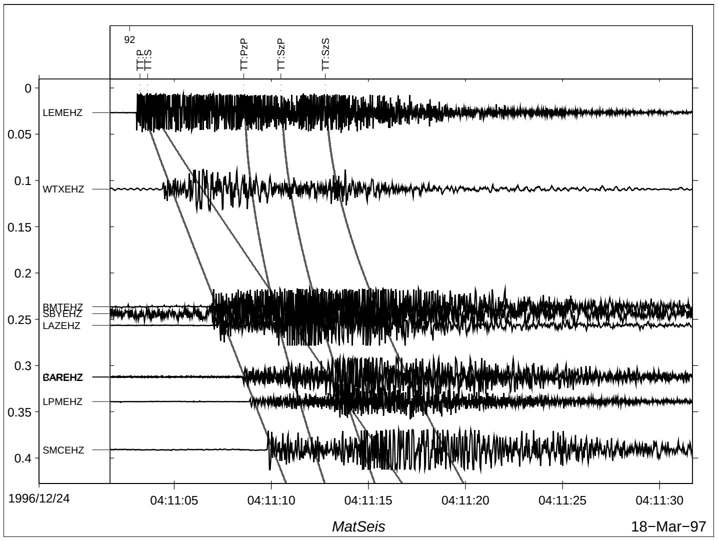
Expanded Set Evid=89

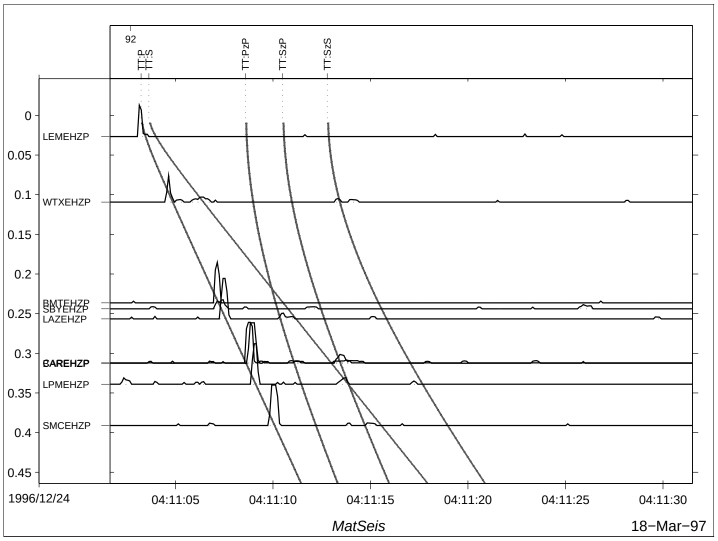


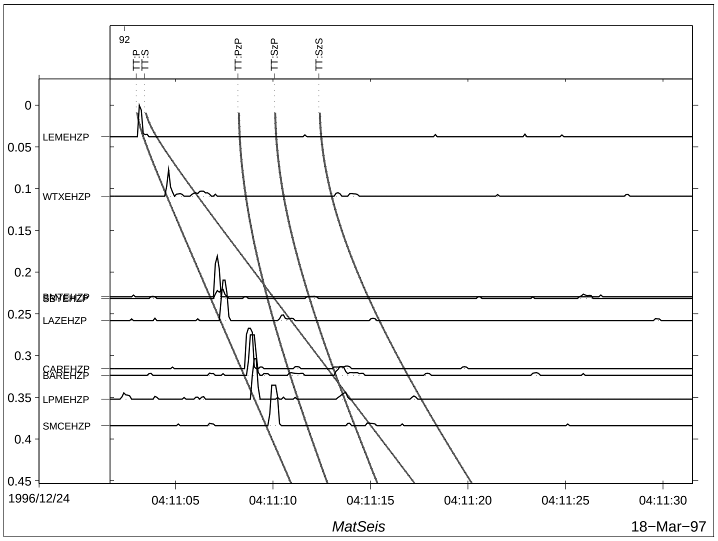




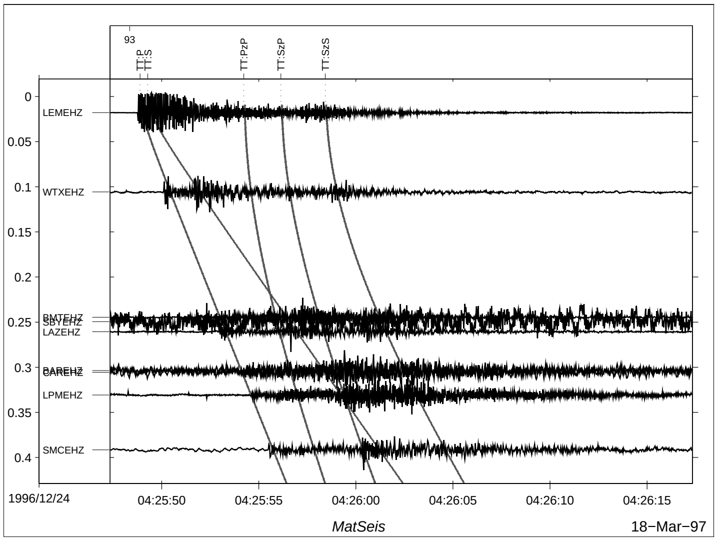
Expanded Set Evid=92

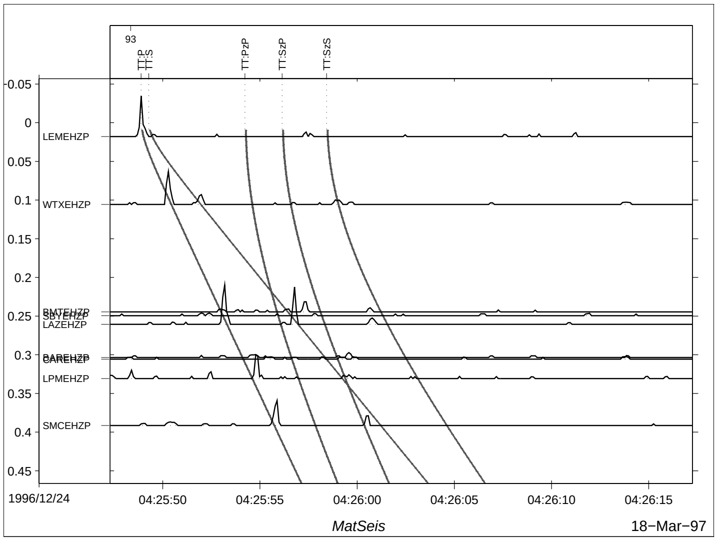


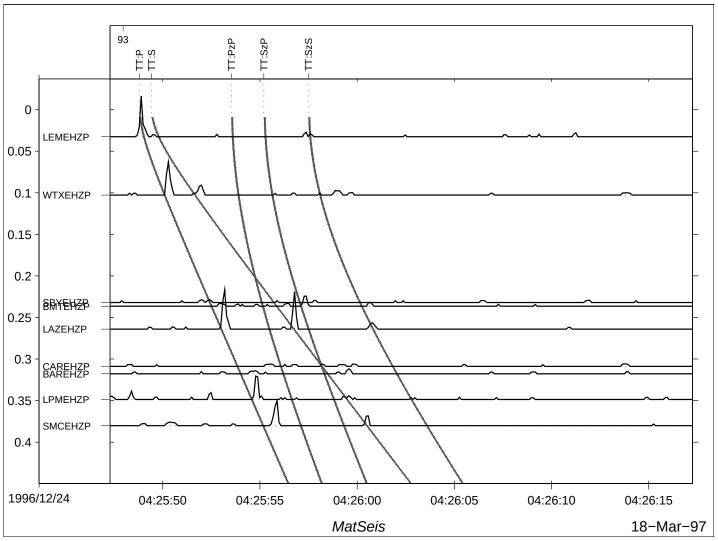




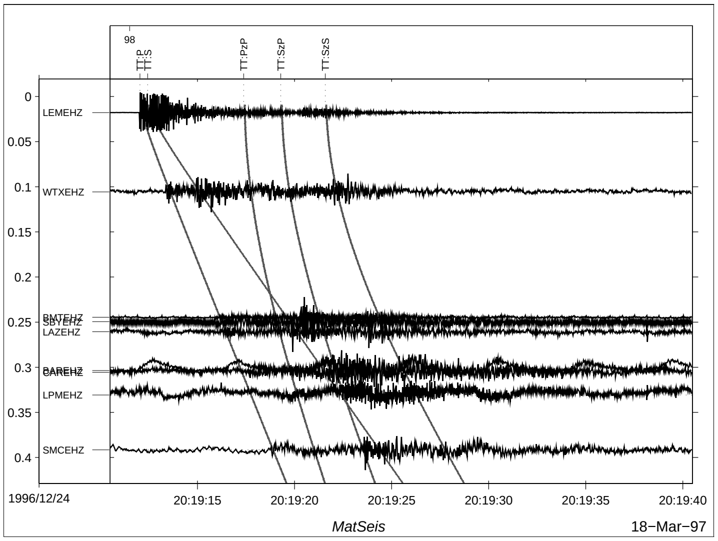
Expanded Set Evid=93

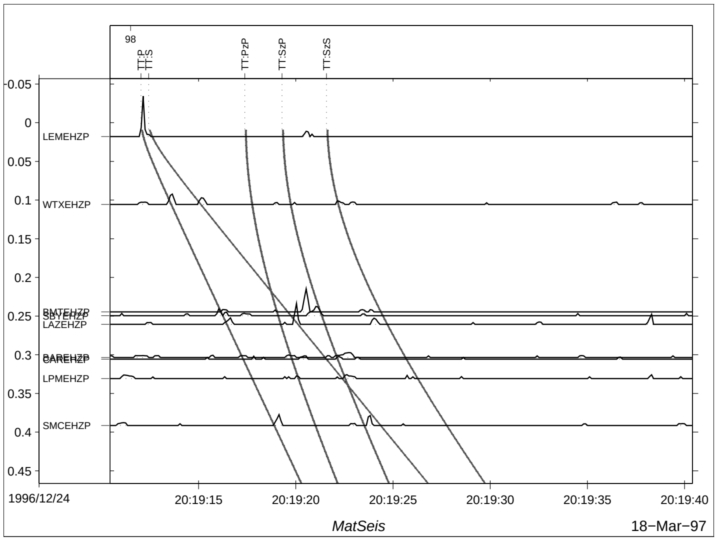


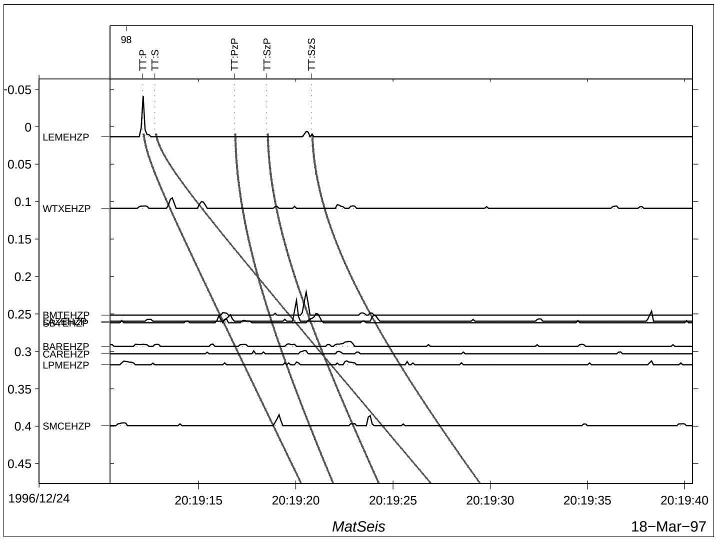




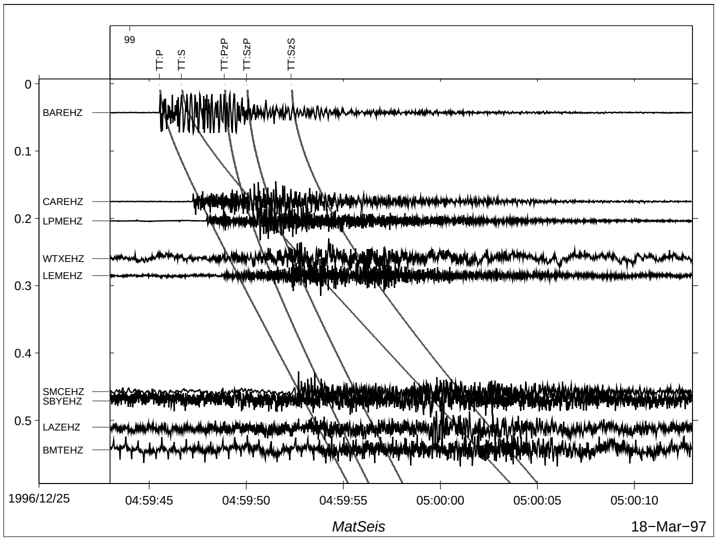
Expanded Set Evid=98

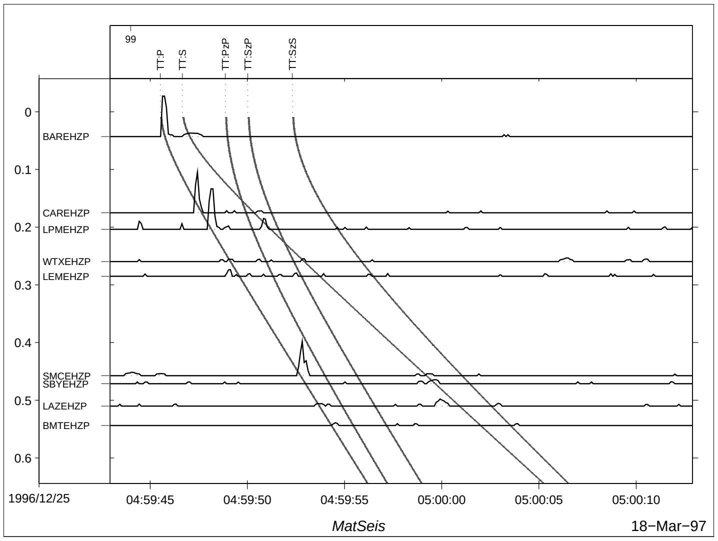


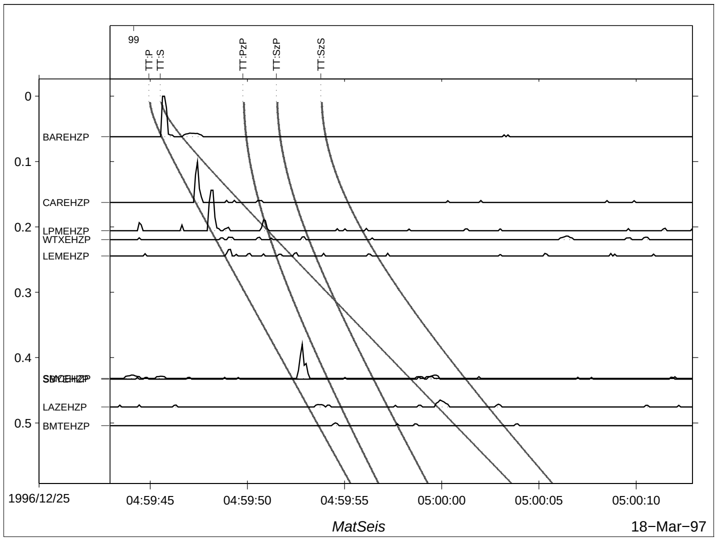




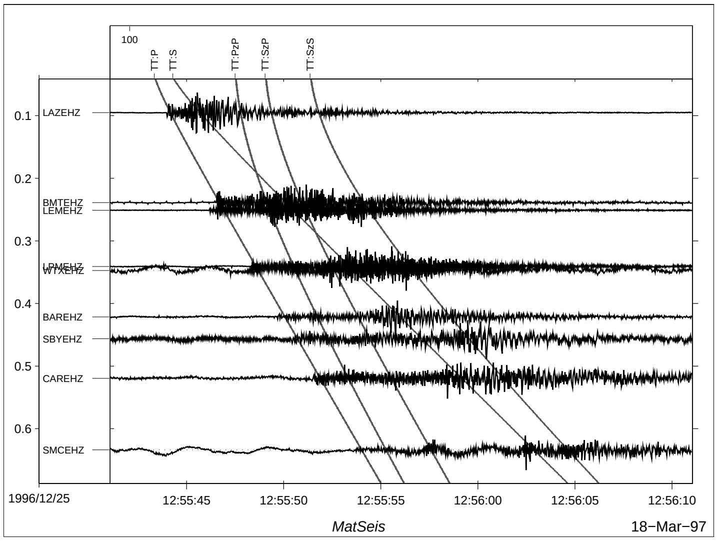
Expanded Set Evid=99

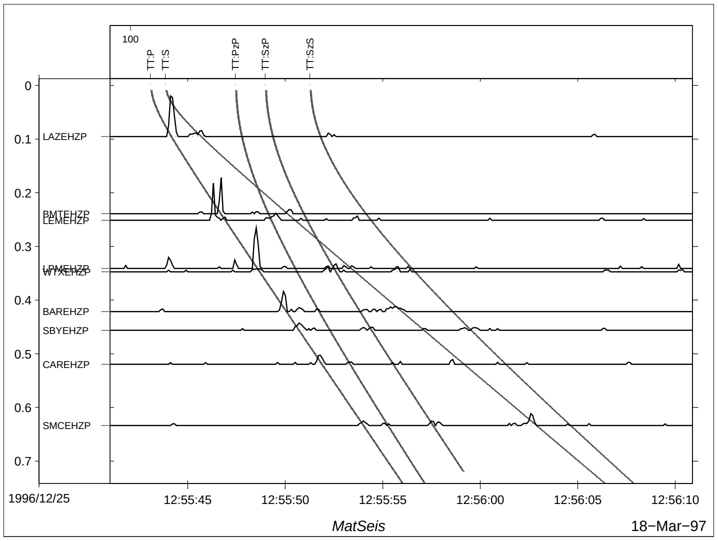


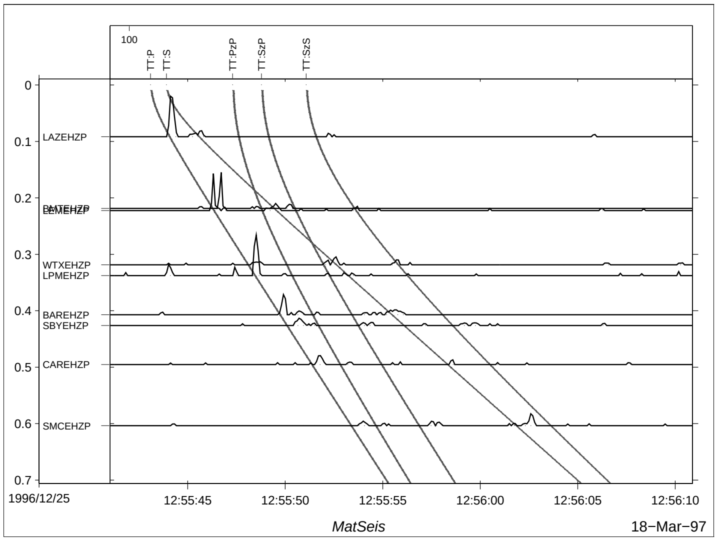




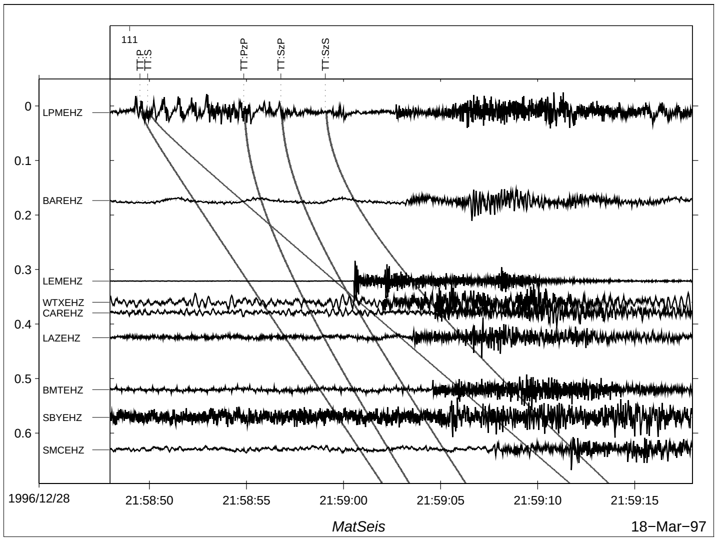
Expanded Set Evid=100

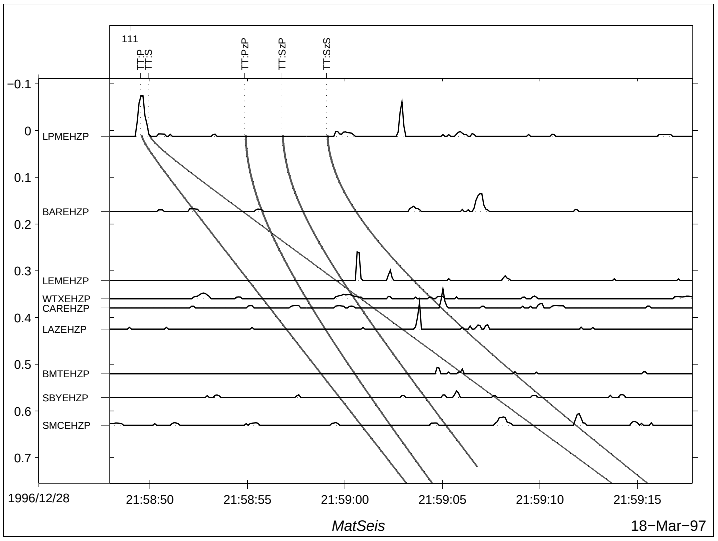


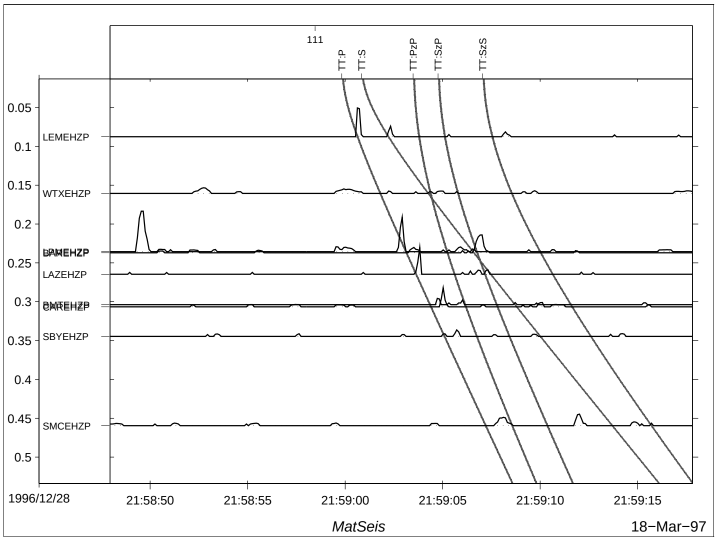




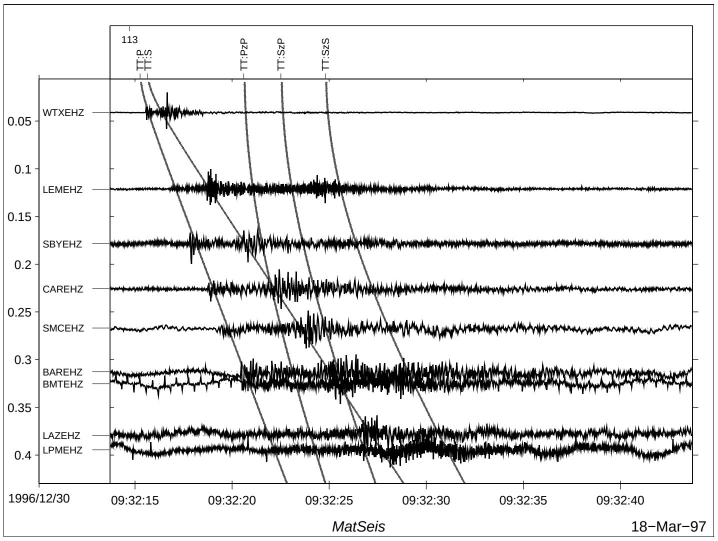
Expanded Set Evid=111

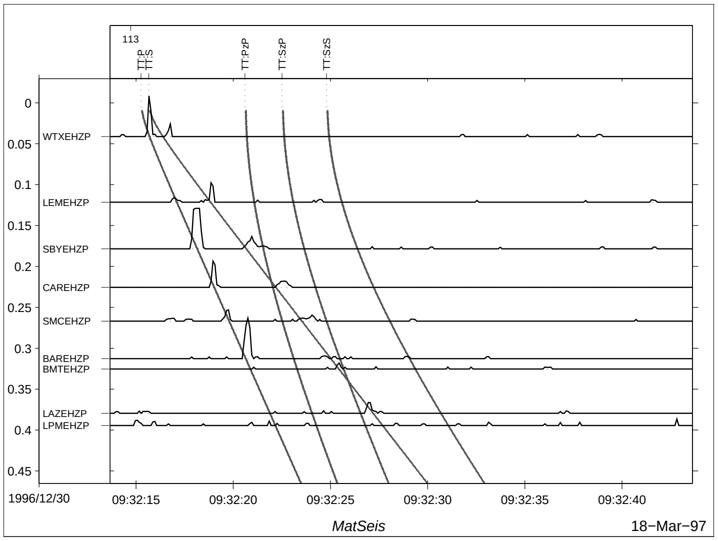


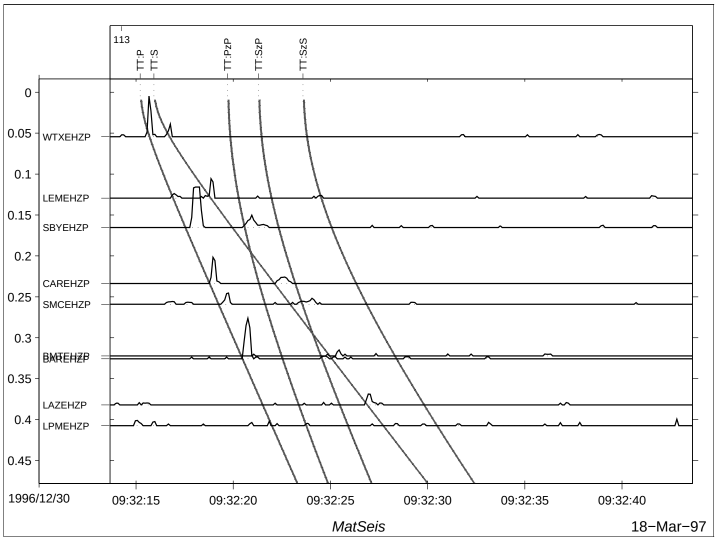




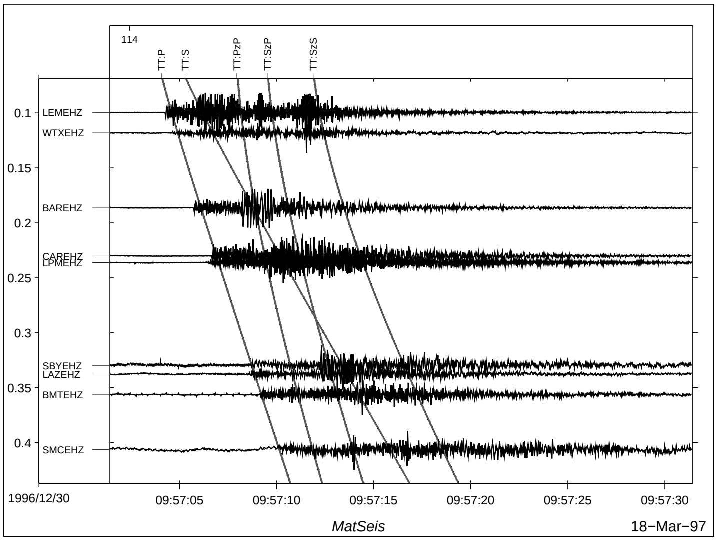
Expanded Set Evid=113

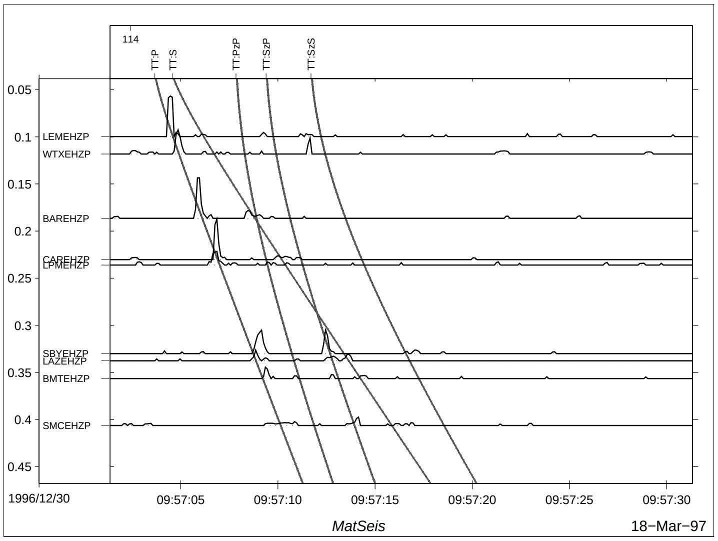


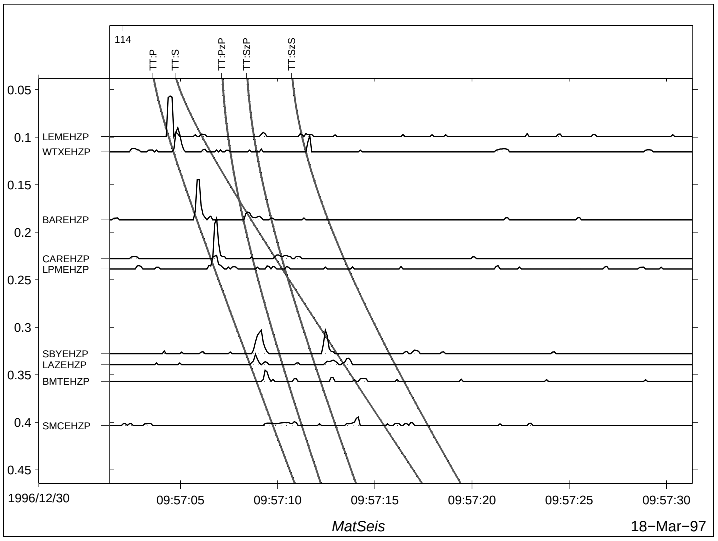




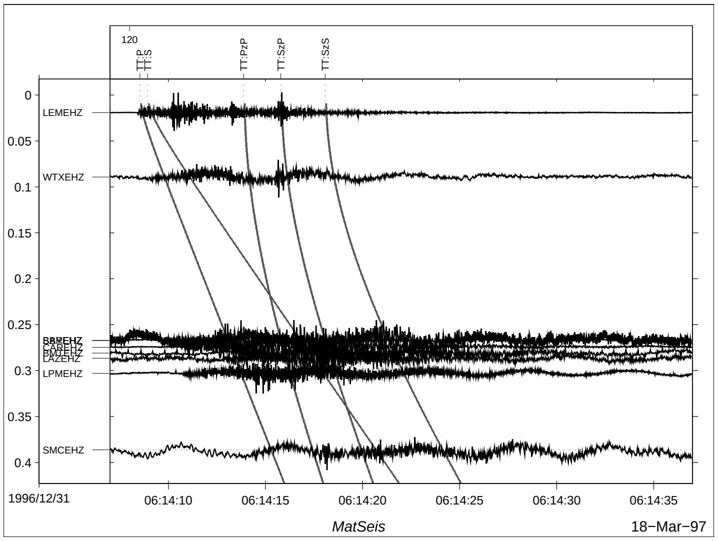
Expanded Set Evid=114

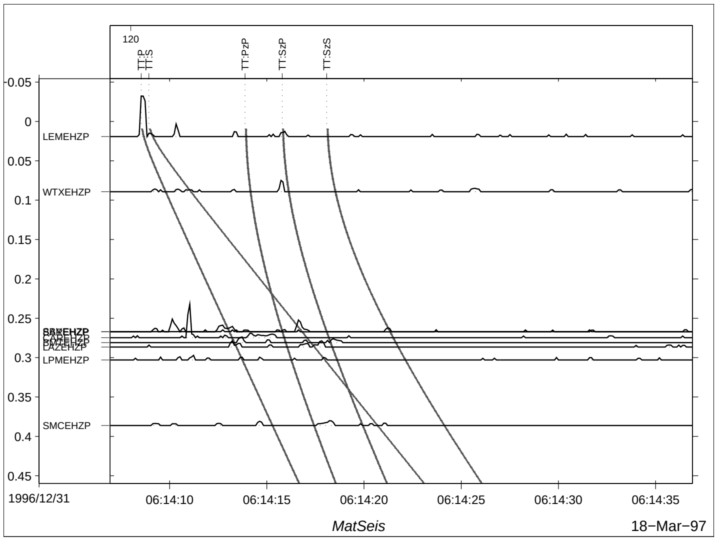


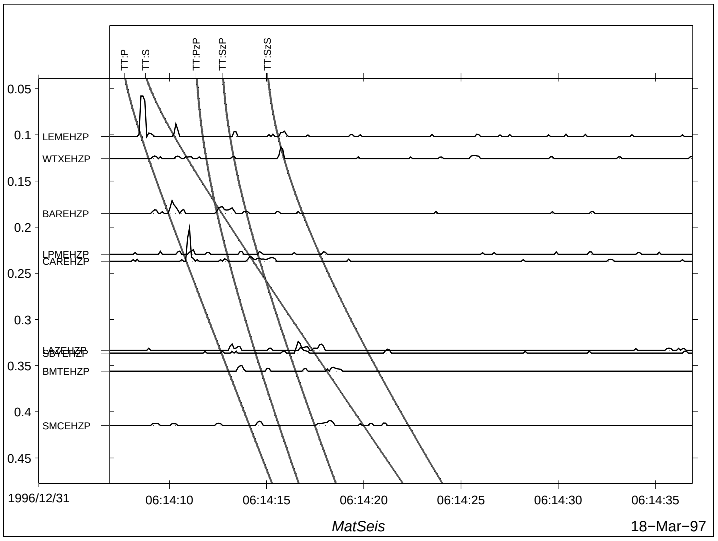




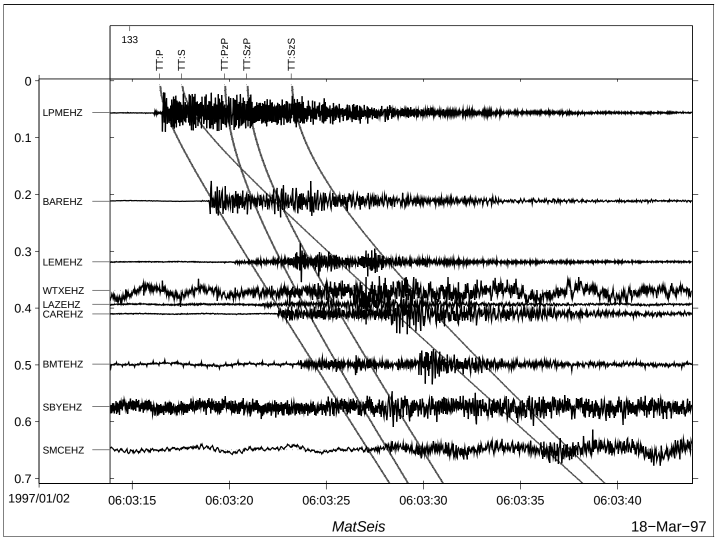
Expanded Set Evid=120

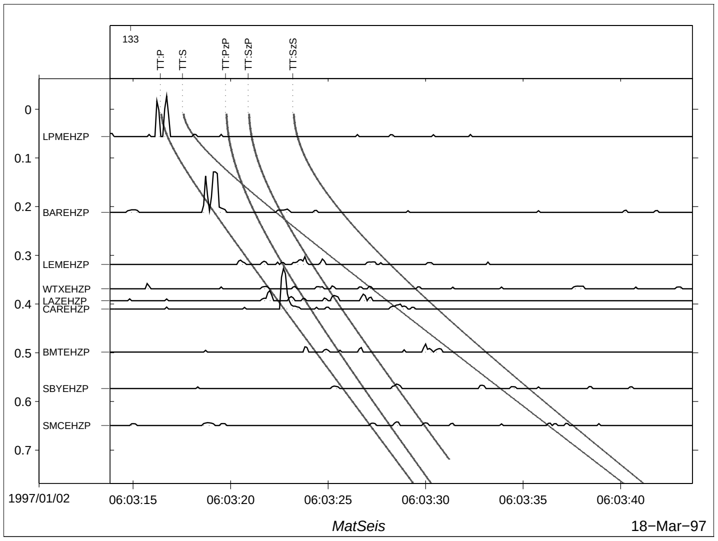


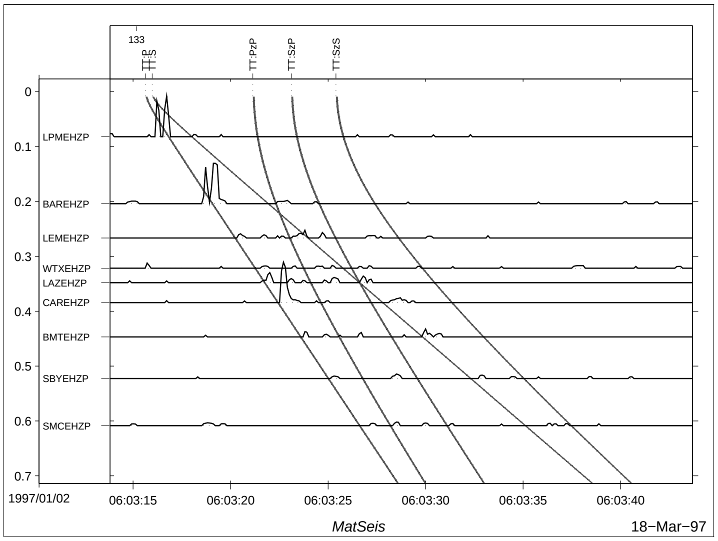




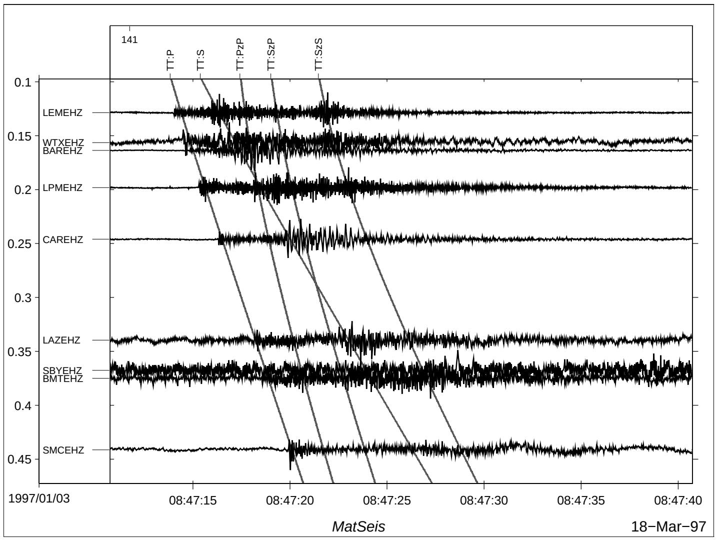
Expanded Set Evid=133

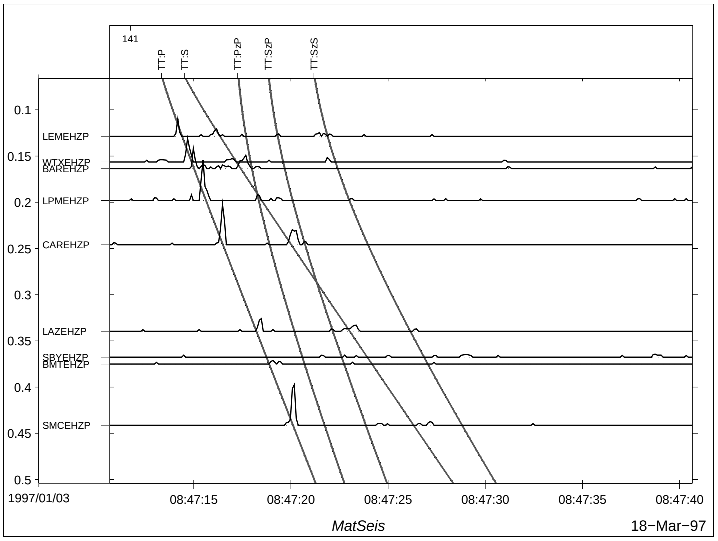


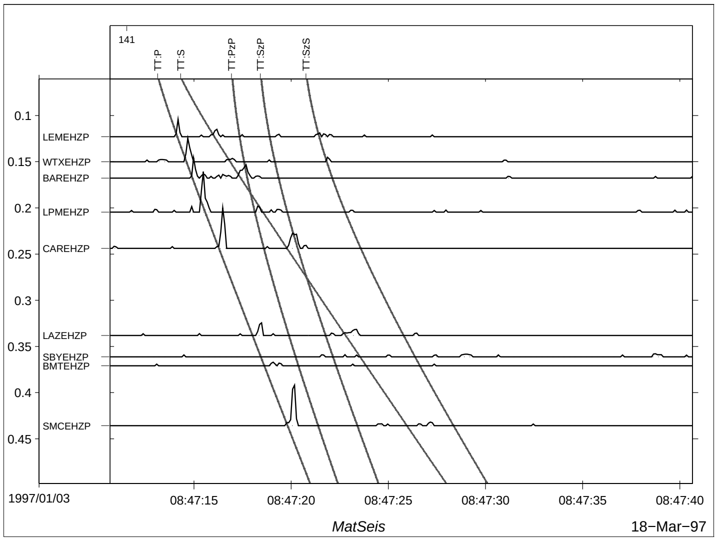




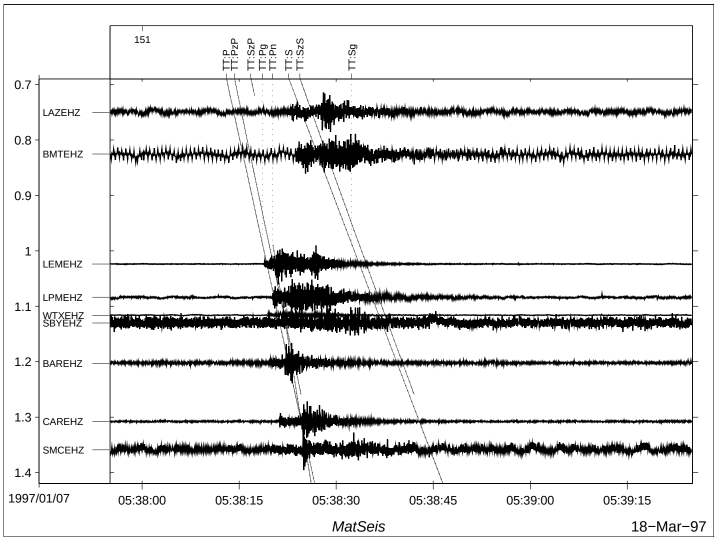
Expanded Set Evid=141

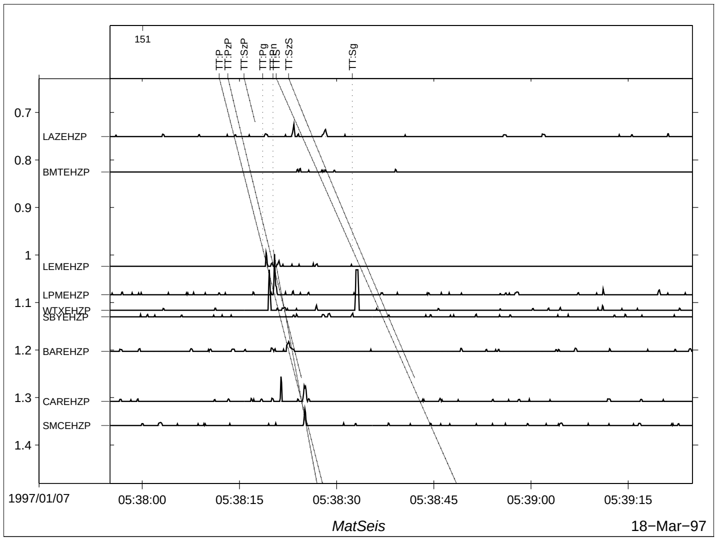


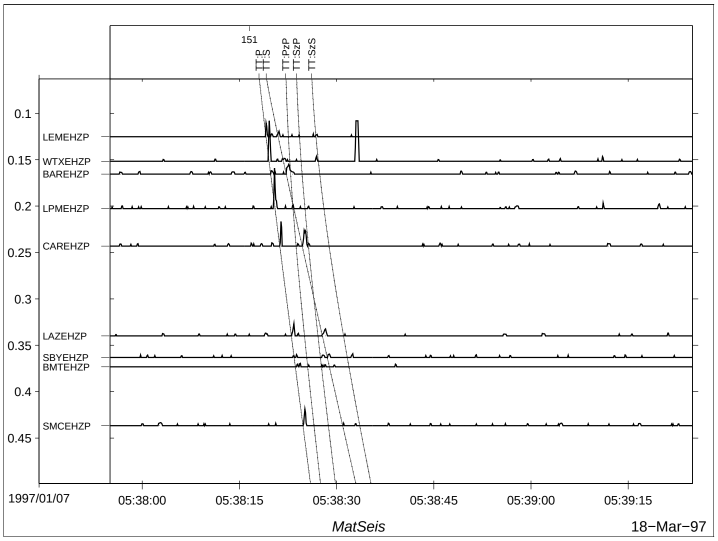




Expanded Set Evid=151







Appendix IV. Journal Article

A Comparison of Select Trigger Algorithms for Automated Global Seismic Phase and Event Detection

Mitchell Withers and Richard Aster
Department of Earth and Environmental Science
and Geophysical Research Center
New Mexico Institute of Mining and Technology
Socorro, NM 87801

Christopher Young, Judy Beiriger, Mark Harris,
Susan Moore, and Julian Trujillo
Sandia National Laboratories
Albuquerque, NM 87185

Abstract

Digital algorithms for robust detection of phase arrivals in the presence of stationary and nonstationary noise have a long history in seismology, and were exploited primarily to reduce the amount of data recorded by data logging systems to manageable levels. In the present era of inexpensive digital storage, however, such algorithms are increasingly being used to flag signal segments in continuously recorded digital data streams for subsequent processing by automatic and/or expert interpretation systems. In the course of our development of an automated, near-real-time, waveform correlation event detection and location system (WCEDS), we have surveyed the abilities of such algorithms to enhance seismic phase arrivals in teleseismic data streams. Specifically, we have considered envelopes generated by energy transient (STA/LTA), Z-statistic, frequency transient, and polarization algorithms. The WCEDS system requires a set of input data streams which have a smooth, low amplitude response to background noise and seismic coda, and which contain peaks at times corresponding to phase arrivals. The algorithm used to generate these input streams must perform well under a wide range of source, path, receiver, and noise scenarios. Present computational capabilities allow the application of considerably more robust algorithms than have been historically used in real-time. However, highly complex calculations can still be computationally prohibitive for current workstations when the number of data streams become large. While no algorithm was clearly optimal under all source, receiver, path, and noise conditions tested, an STA/LTA algorithm incorporating adaptive window lengths

controlled by nonstationary seismogram spectral characteristics of was found to provide an output which best met the requirements of a global correlation-based event detection and location system.

Introduction

Worldwide, approximately 50,000 earthquakes and mining blasts above $m_b \sim 3$ occur annually (e.g., Gutenberg and Richter, 1954), with many thousands exceeding the $m_b = 4.25$ 1990's detection threshold goal (e.g., van der Vink et al., 1996) of the International Monitoring System used by the 1996 Comprehensive Test Ban Treaty (CTBT). The considerable investment in human analyst resources necessary to meet the goals of the CTBT as well as the additional interests of the seismological community, can be considerably reduced through the application of robust automatic event detection systems. To this end, we are currently developing a global waveform correlation event detection system (WCEDS) at Sandia National Laboratories and New Mexico Tech (Young *et al.*, 1996). Conceptually, WCEDS is a pattern matching, or matched filter, algorithm that detects events by correlating processed seismic data streams with theoretical (e.g. Kennett and Engdahl, 1991) or empirical (e.g. Shearer, 1991) travel time envelopes. The output of the WCEDS system consists of a time-dependent correlation value for each gridpoint considered within the volume of the Earth. High correlation indicates that a seismic signal is present in the data and that the data streams have been properly organized in space and time for the given source and model. Poor correlation either indicates misalignment (improper hypocenter space or time coordinates) or the lack of a consistent pattern of seismic phase arrivals in the data. A significant advantage to this methodology is that a priori phase identification is not required to identify or locate an event. As correlation with raw waveforms presently would require a prohibitively detailed model and an unmanageable number of free parameters, the success of the present system is strongly dependent upon our ability to preprocess the data to enhance as many seismic phases as possible in near real-time. Many algorithms which can perform this function have previously been developed as triggers to initiate data acquisition. In this paper, we restrict our investigation to algorithms, or portions of them, which ideally produce peaks at all constituent phase arrivals and have a smooth and low amplitude response to seismic background noise and coda. We organize our investigation into four general categories: time-domain, frequency-domain, particle motion, and adaptive window length processing.

Previous Work

Time Domain Methods

Freiberger (1963) developed the theory for the maximum likelihood detector assuming Gaussian signal superimposed on Gaussian noise. Unfortunately, real data are not so statistically predictable. Vanderkulk and others (1965) used an STA/LTA algorithm operating on rectified data (e.g. absolute value) as at the time, this allowed significant computational savings over squaring the data.

Allen (1978) developed a detector based on an envelope that is equal to the square of the data plus a weighted square of the first derivative. This creates a time series that includes components of both the raw data and high-pass filtered data. The processed data stream is then passed to a set of logical and mathematical tests for phase identification and timing. Stewart (1977) used a modified data envelope (MDX) based on the derivative of the data where slope changes are emphasized. The MDX value is found by first estimating the derivative (DX) at the i^{th} point,

$$DX_i = X_i - X_{i-1}. \tag{1}$$

If the sign of DX has been constant for less than eight consecutive i, then

$$MDX_i = MDX_{i-1} + DX_i, \tag{2}$$

$$\text{otherwise } MDX_i = DX_i. \tag{3}$$

The resulting "feather-like" time series is a high-pass realization of the data where the first motion and oscillatory nature of the raw data have been preserved, and emergent signals have been enhanced. A set of detection criteria and thresholds must then be met to declare an event.

Frequency Domain Methods.

Shensa (1977) used the Fast Fourier Transform to develop a detector based on the Power Spectral Density (PSD). From the PSD he developed three algorithms: the average power detector, the maximum deflection detector, and the average deflection detector.

The average power detector finds the PSD, $P_i(k)$, at each time step, i. The output Y_i, is determined by removing the mean and normalizing by the standard deviation (i.e. "standard normalize") for some frequency index range, $n_1 \leq k \leq n_2$.

$$Y_i = \frac{\frac{1}{N}\sum_{k=n_1}^{n_2} P_i(k) - \mu}{\sigma} \quad N = n_2 - n_1, \tag{4}$$

where μ and σ are the mean and standard deviation, respectively, of $P_i(k)$. The average power detector is optimum for weak signals that exceed noise uniformly over a relatively broad range of k when both noise and signal are stable.

The maximum deflection detector "standard normalizes" the PSD at each frequency index, k. The maximum value across all frequencies, Z_i, is then output for each time step, i.

$$z_i(k) = \frac{P_i(k) - \mu(k)}{\sigma(k)}, \tag{5}$$

$$Z_i = \mathrm{MAX}[z_i(k=0), z_i(k=1), ..., z_i(k=N)]. \tag{6}$$

This detector is optimum for weak signals that exceed noise over at least one narrow frequency band.

The average deflection detector determines the "standard normalized" PSD for a given frequency index. The output for for the i^{th} time step, X_i, is the average of this normalized PSD (z_i from 5) across some index range, $n_1 \leq k \leq n_2$.

$$X_i = \frac{1}{N} \sum_{k=n_1}^{n_2} z_i(k) \quad N = n_2 - n_1. \tag{7}$$

This detector is optimum on weak signals that exceed background noise uniformly over a relatively wide range of k when both signal and noise are unstable.

Because of its speed, Goforth and Herrin (1980) used the Walsh transform (analogous to a Fourier decomposition, except square waves rather than sinusoids are used as basis functions) to generate filter weights in select pass-bands. Then, as an indicator of nonstationarity, the trigger used a short-term average of the Walsh weights compared to the median and 75th percentile of a long-term set of weights. Masso and others (1979) developed the MARS detector which first applies a suite of band-pass filters to the time series, then searches each pass-band for energy peaks. The temporal separation of these peaks is then used to narrow the time window over which to search for the highest energy non-dispersive value which is then used as the trigger. All of the previously mentioned detectors are reviewed by Berger and Sax (1980) and some are discussed by Allen (1982).

Other Methods.

Blandford (1974) used the F detector for array data which basically scans for signals which are coherent, in a delay and sum sense, across the array. Anderson (1978) looked for peaks between zero-crossings and compared them with a long-term average. An event is declared when the peak exceeds a multiple of the LTA and the associated zero crossings are sufficiently separated. Murdoch and Hutt (1983) developed an efficient algorithm which uses only peaks and troughs along with logical threshold comparisons. Joswig (1990) used a pattern-matching scheme based on the power spectral density as a function of time. The power of wavelet transforms applied to seismic data was alluded to by Donoho (1993) and was implemented as a trigger by Ebel (1996). Aster *et al.* (1990), Magotra and others (1989 and 1987), Montalbetti and Kanasewich (1974), and Flinn (1965), to name a few, constructed detectors and filters based on the eigenvalues and eigenvectors of the three component covariance matrix.

To summarize, previous work can be generally categorized into time domain, frequency domain, particle motion processing, or pattern matching. The algorithms discussed so far were often used as triggers for digital data acquisition or for phase identification and timing, ergo they were tuned to detect phase onset. None of the detectors were optimal under all situations and many were designed to operate under computing systems which are obsolete today. The Allen detector is currently used in many deployments of the widely used Earthworm (Johnson, 1995) near real-time network processing system and in the southern California CUSP system (Caltech and USGS Seismic Processing; Dollar, 1989). In combining the energy with the first derivative, the Allen detector incorporates a highpass filter making it inappropriate for teleseismic data. Indeed it is only used in local or regional settings. In the context of the WCEDS waveform correlation, we are primarily interested in generating a suitable data envelope for correlation rather than in detecting phase onset. Following the general time, frequency, particle motion, and adaptive scheme, the output of a few select preprocessing methods will be compared. The method used by WCEDS should enhance as many phases as possible with as few preset arbitrary parameters as possible; detection logic is embedded in the WCEDS correlation and need not be performed by the preprocessing algorithm. Consequently, processing should have smooth low amplitude response to background noise and coda (henceforth just background), and be computationally parsimonious.

Data

Because it is an automated global earthquake location system, any algorithm used in WCEDS must perform well in a variety of signal, path, and noise environments. We have tested the forthcoming algorithms on 7 different events with between 15 and 35 receivers per event, but only show results for one event at select receivers. The data used were 3-component high gain, broadband seismograms obtained from the Incorporated Research Institutions for Seismology (IRIS). The event shown in all figures is an $m_b \sim 6.2$ earthquake from southern Xinjiang, China on October 2, 1993 at 08:42:33 UT. For brevity, only vertical-component data recorded at station OBN ($\Delta = 38°$) are used to compare time-domain processing algorithms. For the remaining methods 3-component data are used from stations with distance $20° \leq \Delta \leq 40°$. A vertical-component network gather for this event is shown in Figure 1 where the time series are positioned vertically to reflect the distance between the source and receiver and the amplitudes have be normalized to enhance distant recordings. The travel time curves are IASPEI 91 (Kennett and Engdahl, 1991).

Time-domain Processing

The STA/LTA evaluates the ratio of short- to long-term energy density (squared data). To facilitate combining multi-channel data in a Pythagorean sense, the squared rather than rectified data are used; with current computing capabilities the additional processing time is trivial. Although optimal window lengths depend on the frequency content of the seismic signal, for IRIS broadband (20 sps) data recorded at teleseismic distances, reasonable body wave performance can be obtained with window lengths of 3 and 24 s (Nsta=60 points and Nlta=480 points respectively in equations 8 and 9). Note that the short window leads the long window and there is no overlap or delay; the long-term window begins one data point after the short window ends and the time index is set to the most recent point in the STA to preserve causality. We prefer this window placement because it gives better statistical independence between the short- and long-term averages than would overlapping windows. We operate on the vertical component seismogram shown at the top of Figure 2. The output of the non-recursive STA/LTA (rectangle impulse response) is shown in Figure 2A, where

$$STA_i = \frac{x_i^2 - x_{i-Nsta}^2}{Nsta} + STA_{i-1} \tag{8}$$

$$LTA_i = \frac{x_{i-Nsta}^2 - x_{i-Nsta-Nlta}^2}{Nlta} + LTA_{i-1}. \tag{9}$$

A better approximation to statistical independence between the two windows can be gained by separating them with a specified delay time (output shown in Figure 2B, for delay of 100 samples, Nsta=60, and Nlta=480). This method was used by Ruud and Husebye (1992), and is also part of a teleseismic trigger used by L. Powell (pers. commun., 1995) in an excellent portable data acquisition system developed at the University of Wisconsin, Madison. The statistical independence of the delayed windows allows shorter window lengths and consequently produces quicker recovery from transients. Shorter windows, however, cause greater variability in the STA/LTA.

It is common to use the recursive STA/LTA to avoid keeping a long data vector in memory. This is more efficient and allows a decaying exponential, which will recover more quickly from large energy transients, rather than a rectangular impulse response. The recursive scheme also yields smaller "shadow zones" where after a large transient passes the STA, the transient continues to dominate the output by causing a large LTA. The characteristic decay time, T, is the time required for the impulse response to decay to $\frac{1}{e}$ of its original value (Evans and Allen, 1983) and is embedded in the decay constant, C:

$$STA_i = Cx_i + (1 - C)STA_{i-1} \tag{10}$$

$$C = 1 - e^{S/T}$$

where S is the sample rate and T is the characteristic decay time (it is common to use $C = \frac{1}{Nsta}$ and $C = \frac{1}{Nlta}$, for the short and long-term decay constants). Output for this algorithm is shown in Figure 2C. The effects of the exponential decay are revealed in the smoother response to background. Even though the post-P "shadow zone" of 2C appears longer than the same zone in 2A, the resolution of this figure is insufficient to adequately display the post-P response of the recursive algorithm which is smooth and low amplitude due to the exponential decay in the STA rather than the "shadow zone".

Swindell and Snell (1977) developed the Z-detector where Z is the standardized variable (mean removed, and normalized by the standard deviation).

$$Z(x) = \frac{x - \mu}{\sigma} \tag{11}$$

Following the implementation of E. Chael (pers. comm. 1995), x is the STA, μ is the average of the STA, and σ is the standard deviation of the STA. The Z-detector estimates the distance of the data from the mean in units of the standard deviation. It has the advantage of automatic adjustment to variance in the background noise. If the background variance is small, a small change in input is required for a large change in output. If the background variance is large, a large input change is required for a significant output change. The envelope generated by the Z-detector is shown in Figure 2D, but for more direct comparison with other methods the output of the recursive STA/LTA operating on the Z-envelope is shown in Figure 2E. The peak at approximately 1 minute in Figure 2D is due to window initialization effects. In this implementation, σ is assumed zero for points prior to the initialization window (2 s for the STA and 40 s for the average of the STA) resulting in a very small denominator for the first few points after the 42 s initialization window. For plotting purposes, the processed data have been decimated with an acausal filter from 20 sps to 1 sps which tends to smear the discontinuity at 42 s into surrounding points resulting in the slight ”ramping up” to the initialization peak.

One can also use the analytic envelope (e.g., Earle and Shearer, 1994) rather than the energy as input into the STA/LTA. The analytic envelope is defined as $|x(t) + iH[x(t)]|$ where $H[x(t)]$ is the Hilbert transform and $i = \sqrt{-1}$. The output of the recursive STA/LTA operating on the analytic envelope is shown in Figure 2F.

The non-recursive and delayed STA/LTA provide peaks at more phases than the other time domain methods, but their response to background is not as smooth as the recursive STA/LTA and the analytic envelope methods. The Z-detector has a noisy response to background and does not enhance secondary arrivals. For the correlation used by WCEDS, the recursive STA/LTA is the best compromise between providing smooth low amplitude response to background, and enhancing as many phases as possible. Thus, this method is used as a basis for comparison of forthcoming methods and results for the 3-component (channels are squared and summed) network gather are shown in Figure 3.

Frequency-domain Processing

The preprocessors discussed previously are essentially detectors of time-domain energy transients. Phase arrivals also introduce transients in the frequency content of the seismogram. To enhance abrupt changes in spectral content, we have developed a frequency-domain non-stationarity filter (FDN). McGarr and others (1964) used a similar method on windowed P-arrivals to remove the pre-event noise spectrum from the P spectrum.

In this method, power spectral density estimates are obtained for moving short- and long-term windows positioned much like the STA/LTA windows. Welch's method with 50% overlapping Hann tapers is used to estimate the power spectra in the long-term window and a single Hann taper to obtain the short-term spectral estimate. The long-term spectral magnitude is then subtracted from the short-term (and any negative values are set to zero) to generate a difference spectrum.

Using the difference spectrum and the phase from the short-term spectral estimate, the data are then transformed back to the time domain to obtain an FDN filtered data segment. The process is repeated for successive time segments, advancing both windows by the length of the short-term window. To reduce splicing discontinuities two FDN filtered time series are generated, where the tapers of the second time series are offset from the first by 50% of the short-term window length. The two time series are then combined to obtain a composite output. After FDN filtering, the data are processed with the recursive STA/LTA for direct comparison to other methods. A network gather of these outputs is shown in Figure 4.

Some arrivals, such as S at NRIL and ARU, have been particularly enhanced by using the FDN filter. The response to background however, is unacceptably noisy which is possibly due to poor spectral estimates in the short term window (a single Hann taper is used in the short term Welch estimate). The short term power spectral density estimate may be improved by using multi-taper analysis but not without significant additional processing. One would expect the large frequency change in surface wave arrivals to produce large output in the FDN filter. Unfortunately, this filter suffers from the same limitations of many other processing methods, in that a window length must be selected and this length is usually optimized for body wave arrivals. In this implementation a 64 point window (3.2 s) was used for the short term spectral estimates. With the 20 Hz sample rate the first non-zero frequency bin is centered on 0.3125 Hz where, at the distances displayed in the figure, the period of the surface waves will be twenty to thirty seconds or more. Longer windows could be used but not without significant additional processing nor dilution in the enhancement of body wave arrivals.

Particle motion processing

A third type of transient that may indicate the arrival of a seismic phase is a change in particle motion. In particular, one is often interested in identifying periods of highly linear particle motion that may be associated with body wave arrivals. To emphasize this feature of the data, the 3-component data are prefiltered with a polarization filter (e.g., Montalbetti and Kanasewich, 1970; Aster et al., 1990). We calculate the data covariance matrix for a moving window (3 s in this case) and decompose it into matrices of eigenvectors ($\vec{U} = [\vec{u_1}, \vec{u_2}, \vec{u_3}]$) and sorted nonnegative eigenvalues ($\lambda_1 \geq \lambda_2 \geq \lambda_3$). The linearity of the particle motion can then be characterized by a function, r, of the eigenvalues and the direction of linear motion is given by the eigenvector, $\vec{u}_1$, associated with the largest eigenvalue, λ_1. The data are then scaled by the linearity vector, $\vec{r} = r\vec{u}_1$.

$$\vec{r} = (1 - \frac{\lambda_2}{\lambda_1})^2 \vec{u}_1 \tag{12}$$

$$(z', n', e') = (zr_z, nr_n, er_e) \tag{13}$$

After applying the polarization filter and a Pythagorean summation of the three components, the recursive STA/LTA is applied and an output gather is shown in Figure 5. The polarization filtered data have higher peaks at some secondary arrivals than were seen with the unfiltered recursive STA/LTA and the response to background using the polarization filter is not as noisy as the FDN filtered data.

This method is able to produce an output curve with characteristics that are somewhat better for correlation purposes than the unfiltered recursive STA/LTA but the additional processing involved in diagonalizing the covariance matrix at each time step is considerable.

Wagner *et al.* (1996) proposed using the frequency-domain principal eigenvalue (λ_1) as a detection statistic, where the covariance matrix is now the dot product of the complex frequency components of each channel. For array data, this provides the advantage of automatic beaming, even for a non-planar wave-front, but a narrow pass-band must be judiciously chosen, and ray parameter and azimuth information are lost in the dot product. In the limit as the phase lag goes to zero (e.g. 3-component single-station data under isotropic path effects) this method reduces to the time-domain eigenvalue decomposition. Care should be taken when comparing statistics that reflect rectilinearity with those that reflect the value of the principal eigenvalue alone. The principal eigenvalue is a measure of the maximum energy in the principal direction which may be quite different from the degree of linear particle motion.

Adaptive processing

The frequency content of a nonstationary signal will vary with time, so adaptively varying the window lengths in the STA/LTA algorithm was investigated. As a fast estimator of the dominant frequency content, the zero crossings are determined in the demeaned data to establish the window length. Requiring 6 zero-crossings in the short-term window (3 ”cycles”) and a long-term window 9 times the length of the short-term window provides a reasonable compromise between sensitivity and noise reduction. This strategy was applied to the non-recursive STA/LTA, operating on the energy. Tong (1995) developed an adaptive STA/LTA which selects zero-crossings by requiring that the first derivative be greater than some threshold value. While this is a more robust method and warrants further investigation, the method sections the data and has constant window length within each section. Adjusting the window length on a point by point basis is preferable for its smoothness. Tong uses a recursive STA/LTA and adjusts the window length by changing the decay constant. Implementing the recursive STA/LTA in an adaptive window environment will produce discontinuities in the output and in some pathological cases could be mathematically imprecise. Consider, the recursive algorithm requires keeping only the previous average and one data point in memory and it represents an infinite summation:

$$STA_i = Cx_i + (1 - C)STA_{i-1} \tag{14}$$

$$= Cx_i + (1 - C)x_{i-1} + (1 - C)^2 x_{i-2} + \dots \tag{15}$$

To change the decay constant C requires resumming the data. One might be tempted to apply the new decay constant beginning at the i^{th} point but this presents the possibility of generating a very poor estimate. This is particularly true for the LTA in an expanding window which can cause the LTA to be unjustifiably small because the appropriate new decay constant should be small (but won't be without resumming the data) relative to the old, shorter window, constant. Thus we have opted to sacrifice memory and the exponential decay by using the non-recursive STA/LTA. Because the windows are usually short, there is very little penalty in not using the exponential decay and processing time is not significantly affected. For large data sets, memory use could become critical.

Low-frequency signals (e.g. surface waves) may produce extremely broad peaks in the output of the adaptive STA/LTA, and this is not desirable for the WCEDS correlation because of the large residuals associated with the travel times of these waves and the potentially large correlation value resulting from the broad peaks. We mitigate this effect by capping the short window length at 60 s. Within the limitation imposed by this clamp, windows may expand within the seismic coda as they generally encounter longer period arrivals. When the LTA expands backwards into a previous large arrival (e.g., into P while operating on PcP) the previous arrival may significantly perturb the long-term average and thus adversely bias the desired output peak corresponding to the later arrival (i.e. create a "shadow zone"). To control this effect the additional data to be added into the back-expanding long-term window is examined. If the power in the new long-term window segment is greater than ten times that of a similar-length average piece of the previous long-term window (indicative of a large energy transient), the STA/LTA calculation instead uses the appropriate fraction of the previous LTA energy rather than the data itself (not unlike the recursive realization where $C = \frac{1}{Nlta}$). The output of the adaptive STA/LTA is shown in the gather in Figure 6.

This algorithm produces large peaks at most arrivals and has smooth, low amplitude response to background. Further, the method need not be tuned to a given frequency band and is able to enhance both short and long period arrivals with a single algorithm. The resulting processed waveform is a truer representation of the data envelope than other methods are able to generate and is thus more suitable for correlation. Only very small increases in memory use and processing time are required over the recursive STA/LTA. An additional data stream (zero-crossings) for each station is incorporated, but these may be determined "on the fly."

Discussion and Summary

Significant gains in computing capabilities have allowed much more sophisticated processing algorithms to operate in real-time. Even with these gains, however, the more processing intensive methods may still be overwhelmed by adding many data channels. Philosophically, we prefer simplicity and consequently demand that more complex methods provide greater return. Because the data are time series, time domain methods are, in general, least complex.

The non-recursive STA/LTA has a rectangular impulse response and performance can be enhanced by using squared data (energy) rather than rectified data. The difference in processing time between squaring and rectifying, although once a significant concern, is negligible on modern computers. To more closely approximate statistical independence between the STA and LTA windows, no overlap between the two windows is permitted. To maintain causality the time index is fixed to the first point of the STA. Statistical independence can be further improved by separating the windows by some delay time. This allows shorter windows and consequently shorter "shadow zones" caused by energy transient saturation of the LTA. We can further reduce transient effects and significantly reduce computer memory requirements by using the recursive algorithm which has an exponentially decaying impulse response.

A bank of bandpass filters has often been used to prefilter the data prior to generating the data envelope. This is particularly useful for separating phases that have significantly different path lengths and for triggering on high frequency local events superimposed on teleseismic coda, or long period noise. The real-time performance of a system with many stations, however, may be deleteriously affected by the additional channels. We have developed a "noise" subtracting FDN filter in an attempt to consider the wide range of frequency bands encompassed by the suite of possible phase arrivals without adding additional data streams, but the smoothness of this algorithm in the presence of background noise is less than satisfying (often non-Gaussian over the time scales typically used for phase detection). Better signal spectra may be estimated for the FDN filter if multi-taper analysis were used (currently the short-term spectra are estimated with a single Hann taper) but this is computationally expensive since spectral estimates must be generated at each time step.

One can also search for linear particle motion in detecting body-wave phases, but near-surface scattering and superimposed wind-generated seismic background noise can have a significant corroding influence on linearity (Withers et al., 1996). Further, for the window lengths used, some background noise sources may generate particle motion with significant linearity.

Because frequency domain and particle motion methods are generally more computationally intensive than time domain methods, processing requirements are significantly greater. These requirements may be mitigated with larger time steps at the expense of resolution. For the purposes of WCEDS, the added complexity is not justified by the relatively minor gains in the data envelope.

Nearly any algorithm may be enhanced by implementing an adaptive window. The time domain and particle motion methods suffer from tuning the windows to p-arrivals. Even the attempt to adaptively filter the time series using the frequency domain noise subtracting method of the FDN filter failed to enhance surface wave arrivals because the window lengths used for the spectral estimates were tuned to body wave arrivals. If the windows were expanded to include sufficient resolution to enhance surface waves, the number of points in the fast fourier transform would require prohibitive processing.

For the adaptive STA/LTA, we have based our window length on a crude estimate of the dominant frequency which may be improved by a more robust technique (e.g. Tong, 1995). Better window length adjustment may become particularly useful in the case where a local event is superimposed on teleseismic surface waves. Our current technique would ”miss” the local event since the window length would be dominated by the surface wave.

The rectangular impulse response STA/LTA was used with the adaptive window in order to maintain real-time processing and simultaneously prevent improper weighting and discontinuities in the output. Tong’s segmented adaptive scheme appears to work as well as the method we developed, but we are mathematically uncomfortable with the method of employing the recursive STA/LTA in this environment (see equation 15). By updating the adaptive window length point by point, neither the lack of the exponential decay nor the added memory usage with current hardware capabilities is a significant sacrifice given the capability to smoothly enhance signals covering a wide range of frequency content using a single algorithm.

Conclusions

As mentioned earlier, an optimal data envelope is required for input into the WCEDS correlation. Thus less importance is placed on triggers and phase identification than in producing high signal-to-noise peaks at as many phases as possible without producing an appreciable number of spurious peaks. The Z-detector, for example, may do well as a trigger, but the envelope it generates is not optimal for correlation with travel-time curves. We have also found that squaring the data is preferable to rectifying. The recursive STA/LTA provides reduced memory requirements and is smoother in the absence of signal. Gains from the FDN and Polarization filters were not worth the orders of magnitude greater processing, nor did they consistently perform as well as the adaptive STA/LTA. Not surprisingly, no specific algorithm and set of user-defined parameters is optimal for all scenarios of source, path, receiver, and background noise. Nearly any processing technique that requires windowing will benefit from adaptively updating the window length; the polarization filter, for example, would benefit from having a longer correlation window in background noise and a shorter window during seismic phase arrivals. We have used a crude zero-crossing technique to calculate window lengths for the adaptive STA/LTA but more sophisticated methods might be chosen for specific scenarios. Non-recursive windows are required for the adaptive STA/LTA.

The Waveform Correlation Event Detection System will use the adaptive STA/LTA for its ability to enhance phase arrivals over a wide range of source, path, receiver, and noise scenarios with a single algorithm. This will become more important as the magnitude threshold is reduced because surface wave arrivals are often the only discernible phases for small shallow events at regional and teleseismic distances.

Acknowledgments

We greatly appreciate the assistance of Bob Hutt and Jim Murdoch of the Albuquerque Seismic Lab and Eric Chael at Sandia. This paper was improved by a comprehensive review from Marianne Walck at Sandia Labs. This work was supported by the U.S. Dept. of Energy under Contract Number DE-AC04-94-AL85000. Sandia is a multi-program laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy.

References

Allen, R. (1978). Automatic earthquake recognition and timing from single traces, *Bull. Seis. Soc. Am.,* **68,** 1521-1532.

Allen, R. (1982). Automatic phase pickers: their present use and future prospects, *Bull. Seis. Soc. Am.,* **72,** S225-S242.

Anderson, K. (1978). Automatic analysis of microearthquake data, *Geoexploration,* **16,** 159-175.

Aster, R., P. Shearer, and J. Berger (1990). Quantitative measurements of shear wave polarizations at the Anza seismic network, southern California: implications for shear wave splitting and earthquake prediction, *J. Geophys. Res.,* **95,** 12,449-12,473.

Berger, J., and R. Sax (1980). Seismic detectors: the state of the art, *AFTAC unclassified report,* SSS-R-80-4588. 103 pp.

Blandford, R. (1974). An automatic event detector at the Tonto Forest seismic observatory, *Geophysics,* **39,** 633-643.

Dollar, R. (1989). Realtime CUSP: Automated Earthquake Detection System for Large Networks, *U.S. Geological Survey Open File Report,* **89-320,** 3 pages.

Donoho, D. (1993). Nonlinear wavelet methods for recovery of signals, densities, and spectra from indirect and noisy data, *Proc. of Symp. in Appl. Math.,* **47,** 173-205.

Ebel, J. (1996). Development of a seismic event detection and identification algorithm based on wavelet transforms, *Seism. Res. Lett.,* **67, #2,** 37.

Earle, P., and P. Shearer (1994). Characterization of global seismograms using an automatic-picking algorithm, *Bull. Seis. Soc. Am.,* **84,** 366-376.

Evans, J. and S. Allen (1983). A teleseismic-specific detection algorithm for single short period traces , *Bull. Seis. Soc. Am.,* **73,** 1173-1186.

Flinn, E. (1965). Signal analysis using rectilinearity and direction of particle motion, *Proc. of the IEEE,* **53,** 1874-1876.

Freiberger, W. (1963). An approximate method in signal detection, *J. Appl. Math.,* **20,** 373-378.

Goforth, T., and E. Herrin (1981). An automatic seismic signal detection algorithm based on the Walsh transform, *Bull. Seis. Soc. Am.,* **71,** 1351-1360.

Gutenberg, B., and C. Richter (1954). Seismicity of the earth and associated phenomena, Princeton Univ. Press, Princeton.

Johnson C., Bittenbinder, A., Bogaert, B., Dietz, L., and W. Kohler (199 5). Earthworm: A Flexible Approach to Seismic Network Processing, *IRIS Newsletter,* **14 #2,** 1-4.

Joswig, M. (1990). Pattern recognition for earthquake detection, *Bull. Seis. Soc. Am.,* **80,** 170-186.

Kennett B., and E. Engdahl (1991). Traveltimes for global earthquake location and phase identification, *Geophys. J. Int.,* **105,** 429-465.

Magotra, N., N. Ahmed, and E. Chael (1989). Single-station seismic event detection and location, *IEEE Trans. on Geosc. and Remote Sensing,* **27,** 15-23.

Magotra, N., N. Ahmed, and E. Chael (1987). Seismic event detection and source location using single-station (three-component) data, *Bull. Seis. Soc. Am.,* **77,** 958-971.

McGarr, A., R. Hofmann, and G. Hair (1964). A moving-time-window signal-spectra process, *Geophysics,* **29,** 212-220.

Montalbetti, J. and E. Kanasewich (1970). Enhancement of teleseismic body phases with a polarization filter, *Geophys. J. Roy. Astron. Soc.,* **21,** 119-129.

Masso, J., C. Archambeau, and J. Savino (1979). Implementation, testing, and specification of a seismic event detection and discrimination system, *Systems, Science, and Software Report,* SSS-R-79-3963.

Murdoch, J., and C. Hutt (1983). A new event detector designed for the seismic research observatories, *U.S. Geological Survey Open File Report,* 83-785. 37 pp.

Ruud, B., and E. Husebye (1992). A new three-component detector and automatic single-station bulletin production, *Bull. Seis. Soc. Am.,* **82,** 221-237.

Shearer, Peter M. (1991). Imaging global body wave phases by stacking long-period seismograms, *J. Geophys. Res.,* **96,** 20353-20364.

Shensa, M. (1977). The deflection detector, its theory and evaluation on short-period seismic data, TR-77-03, Texas Instruments, Alexandria, VA.

Stewart, S. (1977). Real-time detection and location of local seismic events in central California, *Bull. Seis. Soc. Am.,* **67,** 433-452.

Tong, Chen (1995). Characterization of seismic phases–an automatic analyser for seismograms, *Geophys. J. Int.*, **123,** 937-947.

Vanderkulk, W., F. Rosen, and S. Lorenz (1965). Large aperture seismic array signal processing study, *IBM Final Report*, ARPA contract # SD-296.

van der Vink, G., Simpson, D., Butler, R., Hennet, C., and T. Wallace(19 96). CTBT... At last!, *IRIS Newsletter,* **XV, #3,** 1-3.

Wagner, G., and T. Owens (1996). Signal detection using multi-channel seismic data, *Bull. Seis. Soc. Am.,* **86,** 221-231.

Withers, M., R. Aster, C. Young, and E. Chael (1996). High-frequency analysis of seismic background noise as a function of wind speed and shallow depth, *Bull. Seism. Soc. Am.,* **86 #5,** 1507-1515.

Young, C., J. Beiriger, M. Harris, S. Moore, J. Trujillo, M. Withers, and R. Aster (1996). The waveform correlation event detection system project phase I: Issues in prototype development and testing, *Internal Report, Department of Energy, Office of Research and Development, NN20,* 45pp.

Figure 1. Vertical component seismograms for an m_b 6.2 earthquake in southern Xinjiang, China on October 2, 1993 at 08:42:33 UT. Station codes are to the right of the figure. Data are courtesy of IRIS and are positioned vertically to reflect distance between the source and receiver. Traces are amplitude normalized to enhance distant recordings.

Figure 2. Select time domain processing methods are shown for the Xinjiang event recorded at GSN station OBN in Obninsk, Russia ($\Delta = 38°$). Algorithms are discussed in the text.

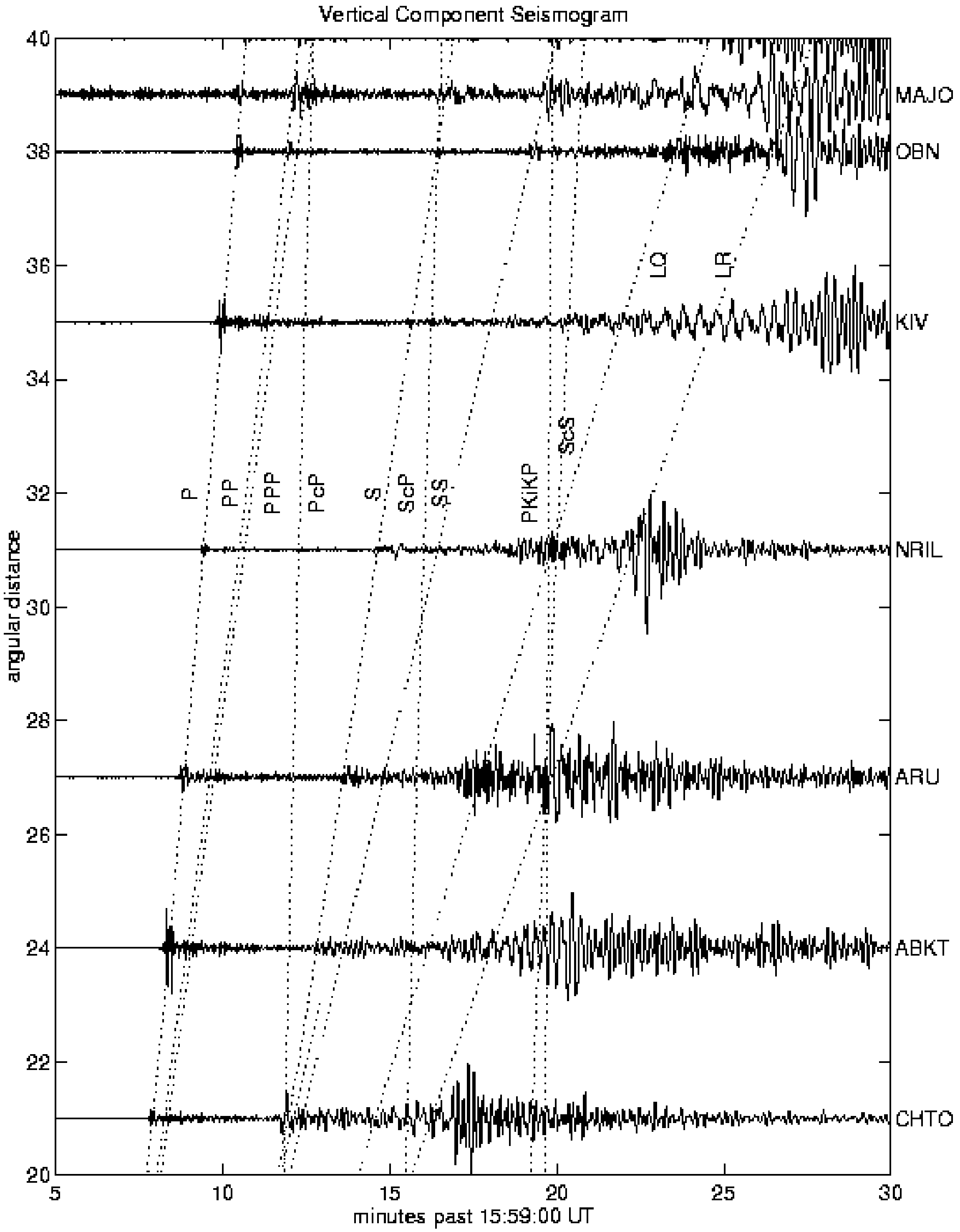
Figure 3. Recursive STA/LTA operating on squared summed three component data using a 3 s short-term window and a 24 s long-term window (see equation 10).

Figure 4. After filtering with the frequency-domain non-stationarity filter (FDN), using 3.2 s (64 points) and 25.6 s (512 points) for the short and long window lengths respectively, we process with the recursive STA/LTA. The FDN subtracts the long term power spectra (”noise”) from the short term power spectra (”signal”).

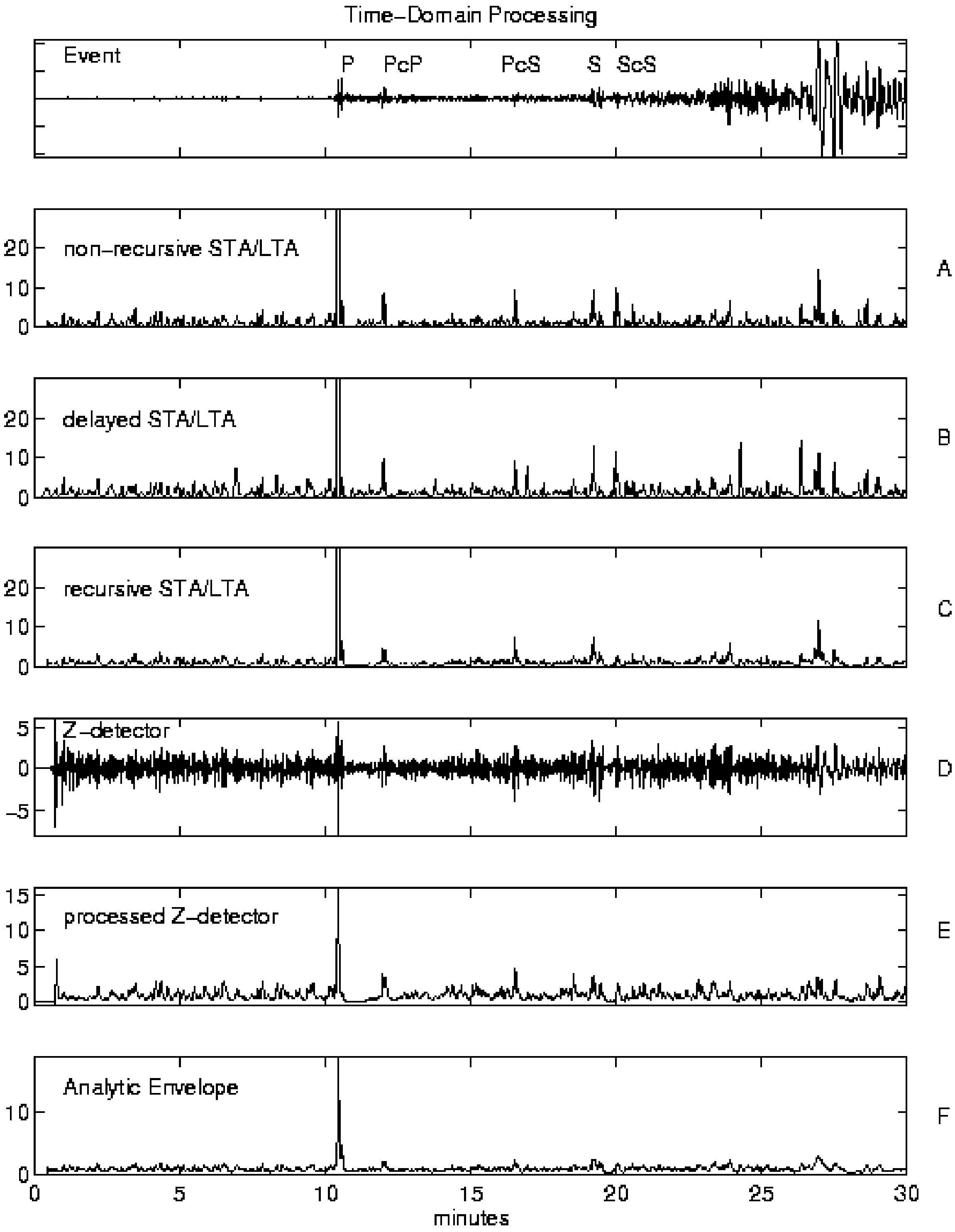
Figure 5. After filtering for linear particle motion using a 3 s moving window, the data are processed with the recursive STA/LTA. The filter applies weights to the time domain data, where the weights are related to the eigenvalues and eigenvectors of the decomposed covariance matrix.

Figure 6. The adaptive STA/LTA has window lengths which are updated at every point such that the short term window (within the bounds of minimum and max-

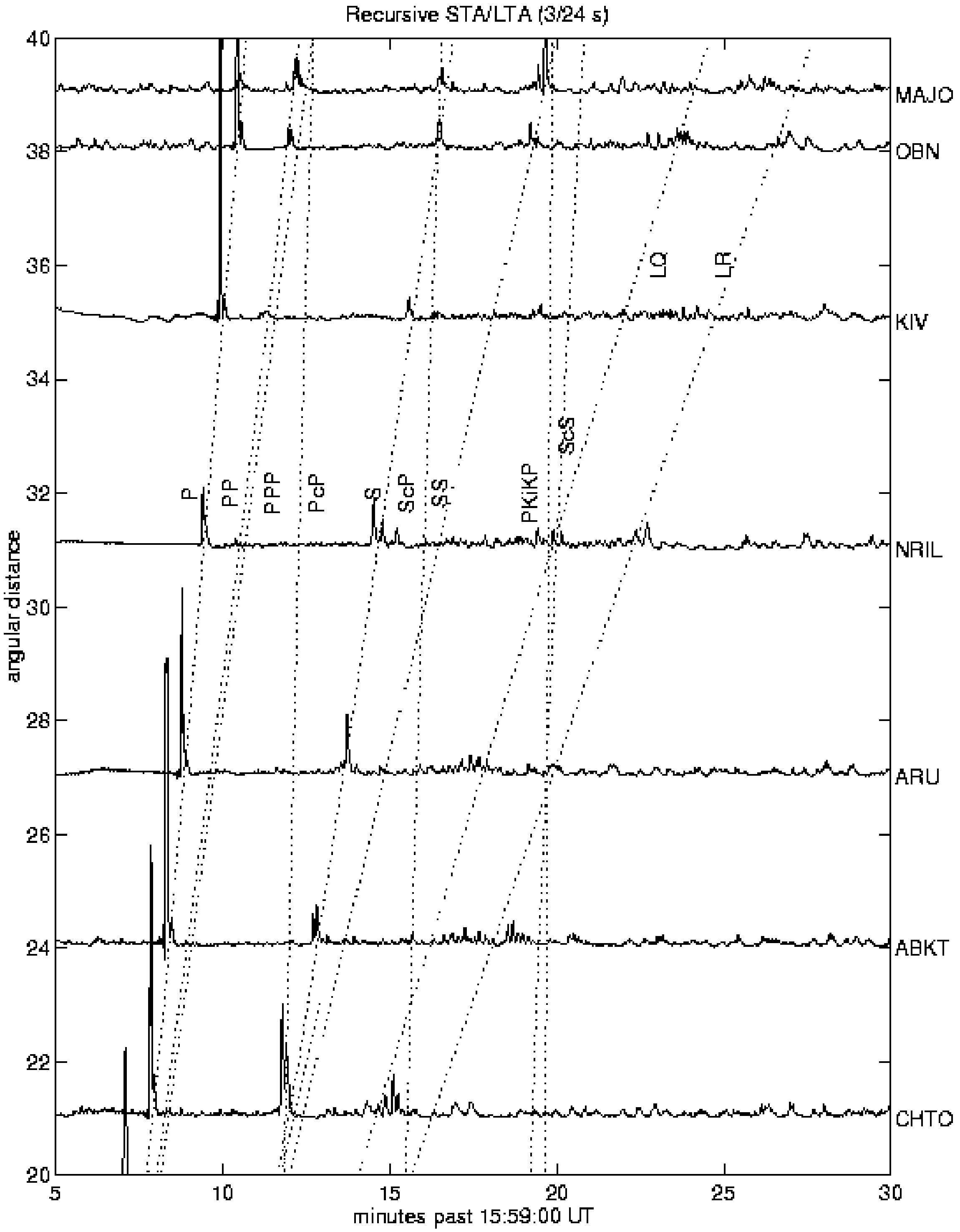
imum lengths of 0.6 s and 60 s respectively) contains six zero-crossings. The long term window is nine times the length of the short term window.



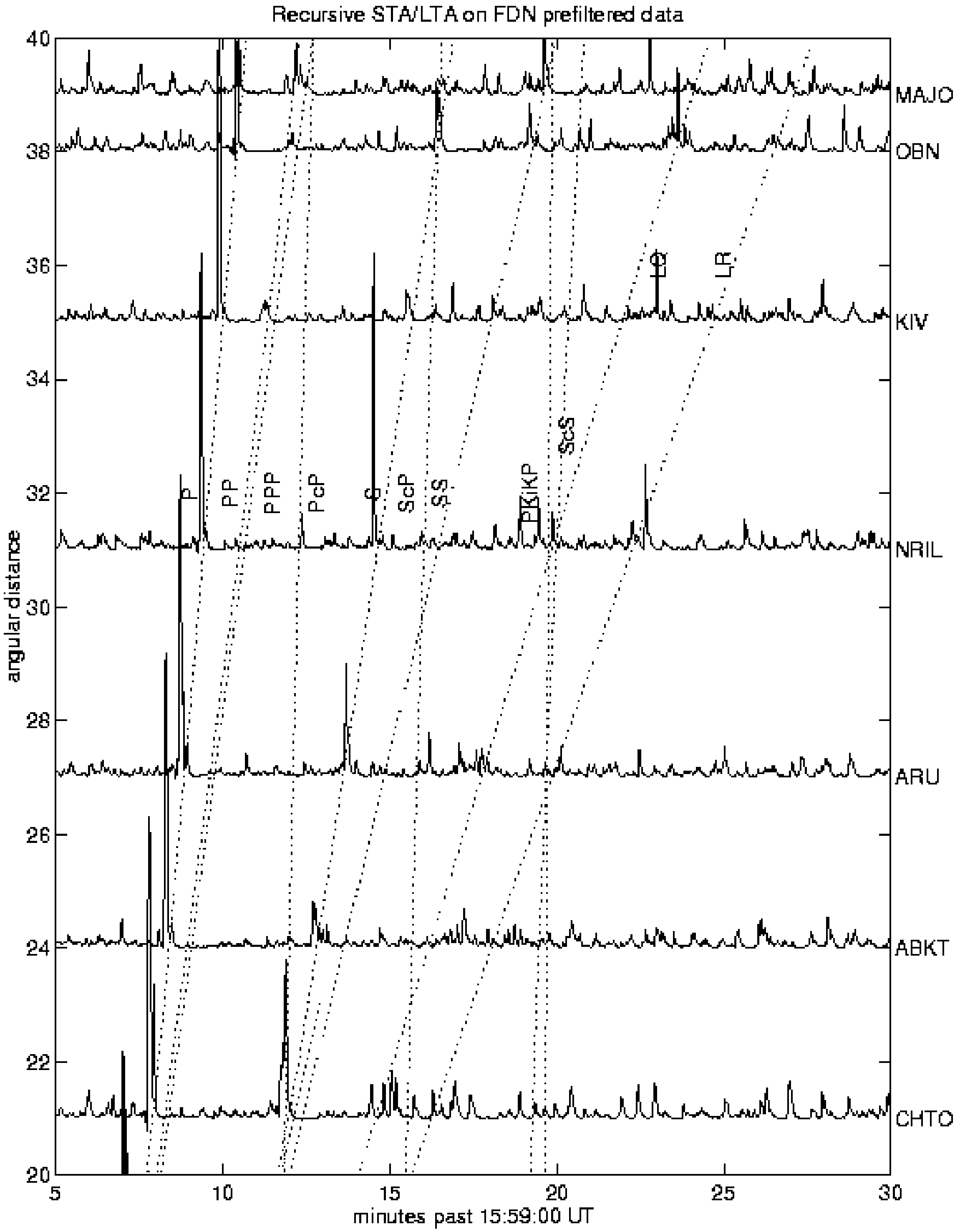
Journal Article Figure 1



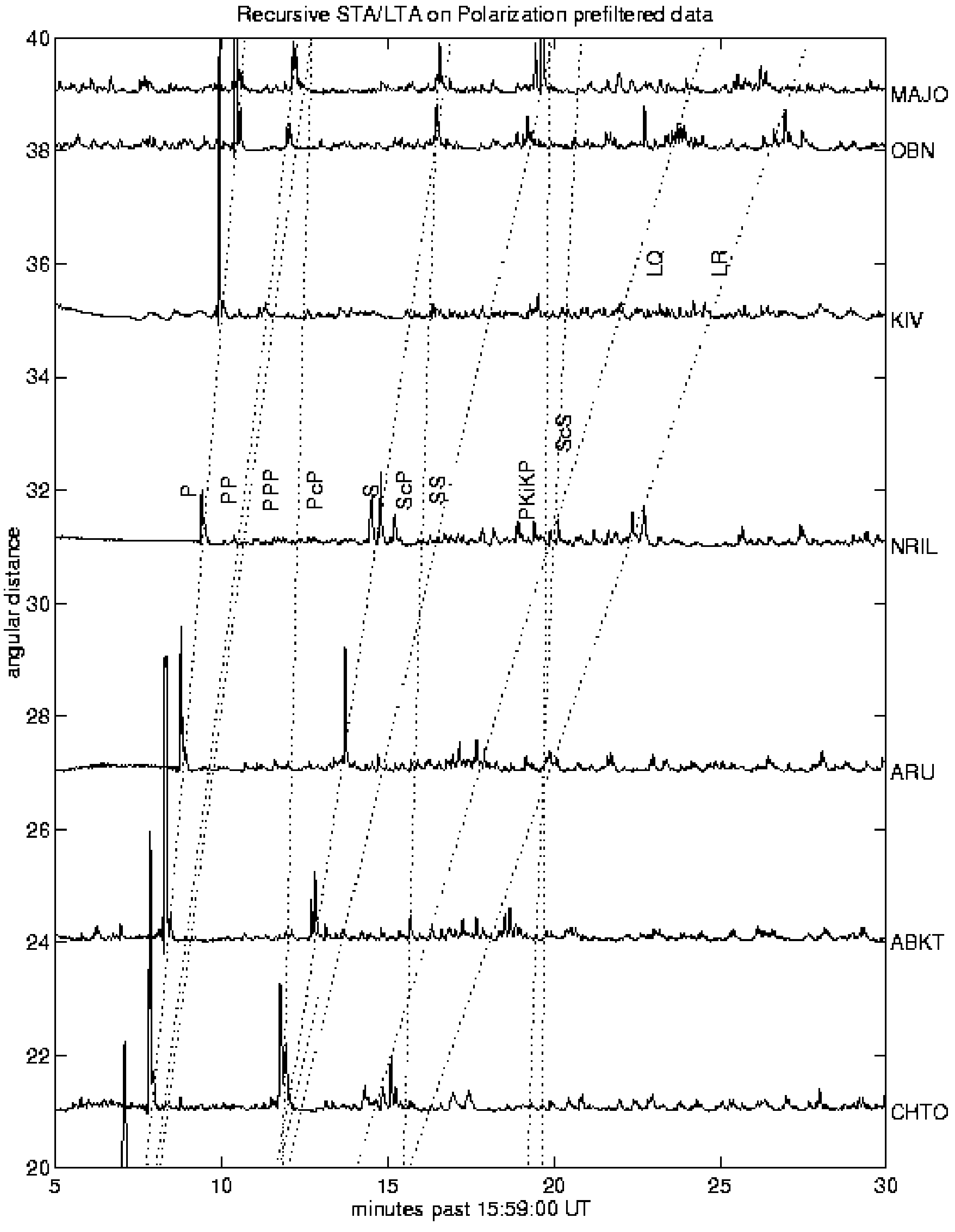
Journal Article Figure 2



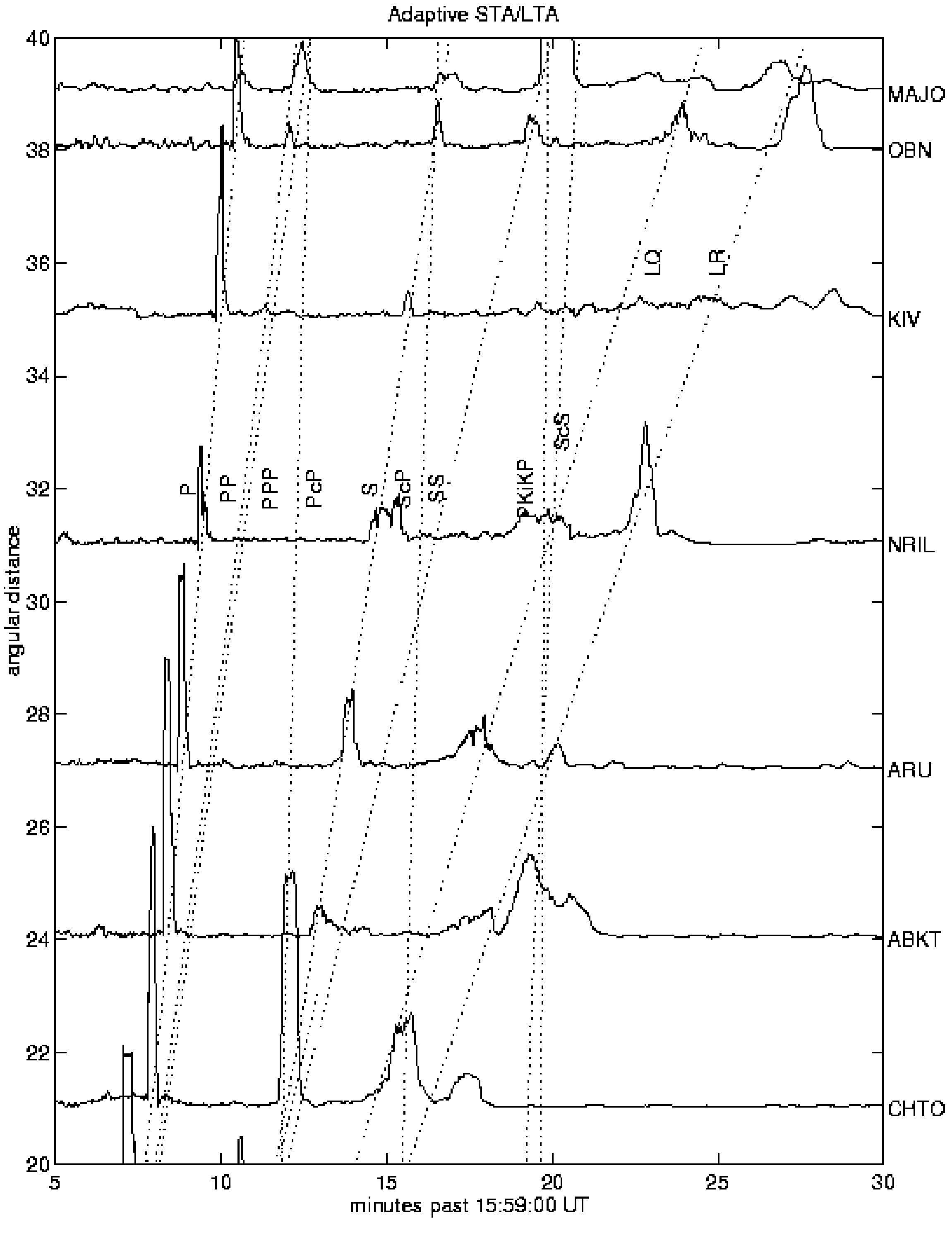
Journal Article Figure 3



Journal Article Figure 4



Journal Article Figure 5



Journal Article Figure 6