

**AN ADAPTIVE, AUTOMATIC PHASE-  
PICKING AND EPICENTER LOCATING  
PROGRAM BASED ON WAVEFORM CROSS-  
CORRELATION  
AND  
ITS APPLICATION FOR PROCESSING  
SEISMIC DATA IN NEW MEXICO**

by

Wenzheng Yang

An independent study submitted in partial  
fulfillment of the requirements for the  
degree of

Master of Science in Geophysics

New Mexico Institute of Mining and  
Technology

2004

Approved by \_\_\_\_\_  
Chairperson of Supervisory Committee  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

Program Authorized  
to Offer Degree \_\_\_\_\_

Date

## **ABSTRACT**

It is widely observed that seismic sources occurring very close to each other can share very similar waveforms. Using this property, an adaptive, automatic phase-picking and epicenter-locating program based on waveform cross-correlation has been developed and applied to process daily seismic data of New Mexico. The core of this program is a reference seismic event database, from which the phase and location information of a new event can be inferred. With a sufficiently complete reference seismic dataset, this program should produce robust initial seismic phase estimates while greatly improving the handling of false triggers due to telemetry or other transient noise. These automatic results are designed to be as precise as the manually processed results.

160 well-picked earthquakes and mining explosions in New Mexico during 1997-2003 were collected as an initial reference event set covering most of the historically active source regions in the state. Each waveform in the reference event set has a high signal-to-noise ratio and accurately-picked P and S phases. Two methods were developed to automatically calculate waveform cross-correlation phase-picking and the location of epicenter.

The first, *waveform-pair cross-correlation matching* method correlates the most similar waveform from the reference dataset to each waveform of a new event station by station and

assigns the corresponding seismic phases to the new waveform according to cross-correlation lag. Note that a new event may be correlated with different reference events at different stations.

The second, *event-pair cross-correlation matching* method correlates the most similar event from the reference dataset to the new event by stacking the cross-correlation curves of all common station-waveform-pairs. The second method provides more robust results than the first method.

Due to the concentration of seismic stations in the Socorro area, an “attracting” problem developed when the *event-pair cross-correlation matching* method was applied to process seismic data of New Mexico. This problem happened when the above method construed distant events to be local ones, because seismic events occurring more than 200 km from Socorro usually have weak S-phases and are therefore identified as seismic events from the Socorro area by the *event-pair cross-correlation matching* method. To nullify this problem, the *spectrogram event-pair cross-correlation matching* method was developed. This method matches event-pair in the frequency domain, decreasing the effect of “attracting” phenomenon. This works because within the frequency band of 6-10Hz, the energy of S-phase for local events is prominent and the background noise energy is very small.

The computer program (**PLRR**) incorporating these methods has been developed under **Matseis/Matlab** (developed at the Sandia National Laboratories) in the **UNIX** environment. It is used to detect initial seismic phases and identify locations of local seismic events in New Mexico from 2000 to 2004. Manual phase repicking and epicenter relocating are applied to the results given by this program to produce a final, reviewed catalogue. If there is a new, high signal-to-noise event that is incorrectly located by **PLRR** (as determined

by the review process), it will be added into the reference dataset. The accuracy of phase-picking and epicenter-locating thus improves as more data are processed.

## ACKNOWLEDGMENTS

I have been very privileged to have undoubtedly the most intuitive, smart and supportive research and academic advisor anyone could ask for: Dr. Richard C. Aster. I have been stimulated and excited by his constant flow of good ideas. I will always admire Rick's to cut through reams of algebra with a single visual explanation. In addition, he has taught me a great deal of new computer skill. Last but certainly not least, he has always known how to give me a little push forward whenever I needed it.

My most intense collaboration has been with Darren Hart. Darren gave me lots of help and ideas in my quest to understand **Matseis** and develop **PLRR**. I cannot remember how many weekends we spent working together in MSEC 345. Darren also proofread this independent study very carefully and provided many valuable suggestions.

Dr. Allan Sanford and Dr. John Schlue gave me valuable instruction on how to pick seismic phases, especially the reflected phases off the Socorro magma body. My discussions with them never failed to ignite my interest in using similar waveforms to study crustal structure above the mid-crustal magma body.

The graduate students of the New Mexico Tech Geophysics Department (both past and present) comprise such a superb research group. The opportunity to bounce ideas off so many excellent minds has been priceless. The upcoming Dr. Daniel Bowman helped me process seismic events in New Mexico in the last month of this project and proofread this independent study work.

My fascination with the physical world is undoubtedly due to the influence of my parents. When I was a young child, they often brought me books about famous scientists and the natural world. These stories had a lasting impression on me. The accumulation of

knowledge and inspiration I inherited from them has greatly contributed to my current fascination with earth science research.

This work has resulted from a project supported in part by the Defense Threat Reduction Agency and Air Force Research Laboratory (contract number DTRA01-01-C-0085), US Department of Defense; the content does not necessarily reflect the position or the policy of the US Government, and no official endorsement should be inferred. Substantial components of the New Mexico Tech Earthworm system were funded by the US Geological Survey.

## TABLE OF CONTENTS

Abstract.....	(1)
Acknowledgements.....	(3)
Table of Contents.....	(5)
Chapter 1. Introduction	
Background Seismicity in New Mexico.....	(7)
The SC seismic Network and nearby Seismic Stations.....	(10)
Major Seismic Source Regions in New Mexico.....	(13)
Characteristic Waveform Patterns in different source regions.....	(17)
Pervious research.....	(20)
Chapter 2. Waveform cross-correlation technique	
Waveform Cross-correlation.....	(23)
Cross-correlation coefficient Matrix.....	(26)
Chapter 3. The development of PLRR	
MatSeis.....	(29)
PLRR.....	(30)
Data Format conversion.....	(31)
Reference dataset.....	(33)
Strategy of phase detection and epicenter location.....	(35)
Waveform-pair cross-correlation matching.....	(36)
Event-pair cross-correlation matching.....	(37)
Spectrogram event-pair cross-correlation matching.....	(39)
Ground data test.....	(41)

Chapter 4. Earthquake location in New Mexico (1997~2004)

Epicenter distribution of seismic events in New Mexico (1997-2004)  
.....(46)

Chapter 5. Summary, Suggestion and Conclusions

Summary.....(62)  
Suggestions.....(63)  
Conclusions.....(64)

References (66)

Appendix I: THE USER'S MANUAL OF PLRR

Appendix II: PLRR source code list and related codes in automatically  
processing of the SC Earthworm system seismic data



## INTRODUCTION

### Background Seismicity in New Mexico

Due to its long distance from major tectonic plate boundaries and weakness of internal tectonic activity, the seismicity in the state of New Mexico is relatively quiet compared to very active seismically regions such as California (Figure 1-1). However, historic activity shows a moderate level of earthquake activity.

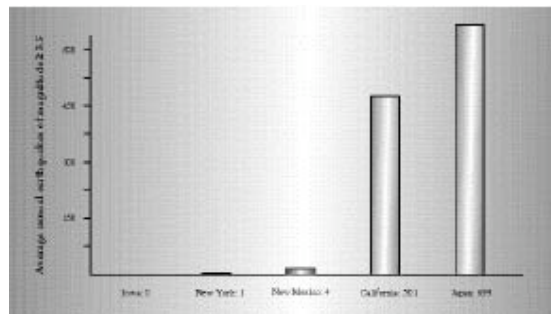


Figure 1-1. Relative seismicity rate of New Mexico as compared to some earthquake-prone and seismically quiet regions [1].

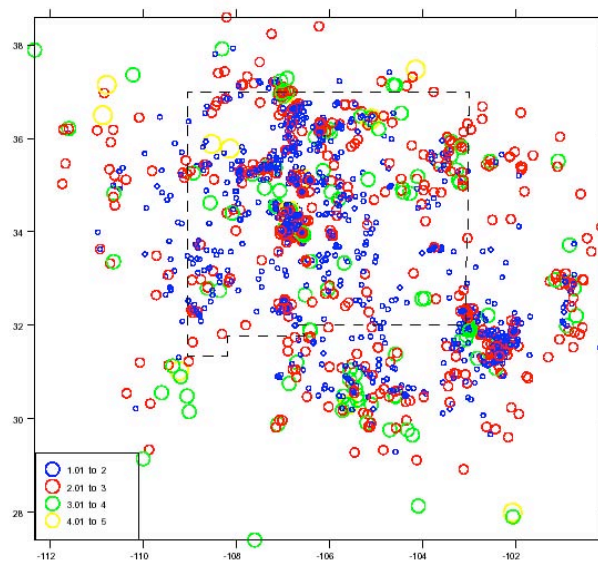


Figure 1-2. Earthquake epicenter distribution in New Mexico and surrounding regions (1962-1995), using the data from [3]. Dashed line outlines the boundary of the state of New Mexico.

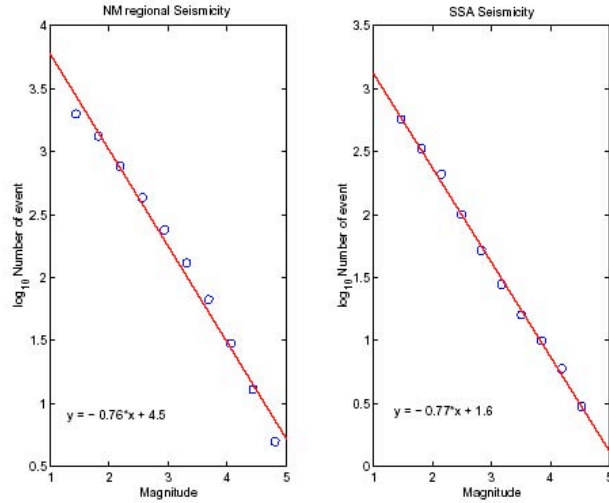


Figure 1-3(a,b).  $b$  value in for New Mexico regional seismicity and New Mexico seismicity associated to the Socorro Seismic Anomaly (SSA) (1962-1995, data from [3]).

Figure 1-2 illustrates the seismicity of natural earthquakes in New Mexico and bordering regions from 1962 to 1995. Most of them were recorded by the Socorro seismic network of New Mexico Tech (NMSN) [3], but some of the earthquakes that occurred in the 1970s were recorded by the Los Alamos Laboratory, the University of Texas at El Paso and the USGS. From Figure 1-2, we observe that the NMSN has documented numerous earthquakes with duration magnitude ( $M_d$ ) above 2.0 inside the state of New Mexico. However, the NMSN may not be sensitive enough to detect all earthquakes during this time with less than 2.0  $M_d$ , occurring in the NE, NW, and SW corners of New Mexico, hence the incomplete seismic record in those regions. According to the historical earthquake catalog ([3]), only 8 % of all 2000 earthquakes that occurred in New Mexico from 1962 to 1995 were larger than 3.0  $M_d$ . Most earthquakes occurring in New Mexico are small events. Their  $M_d$  typically falls within the range of 1.3 to 2.5, and they are usually located in isolated source regions. The magnitude-frequency curve is usually used to check the completeness of earthquake catalogue, and the slope of this curve ( $b$  value) usually ranges from 0.6 to 1.3

around the world. However, as shown in Figure 1-3a, for New Mexico and bordering regions, the  $b$  value of earthquakes was 0.76 and the fitting residuals were large. The best-fit line departs from the data around  $M_d$  1.5, which tells us that earthquakes with a magnitude less than 1.5 are incomplete in [3]. Figure 1-3b shows that the  $b$  value of the Socorro Seismic Anomaly (SSA) is 0.77. The fitting residuals were very small, which means the catalog of [3] is complete for SSA region down to approximately  $M_d$  1.0.

### The SC seismic Network and nearby Seismic Stations

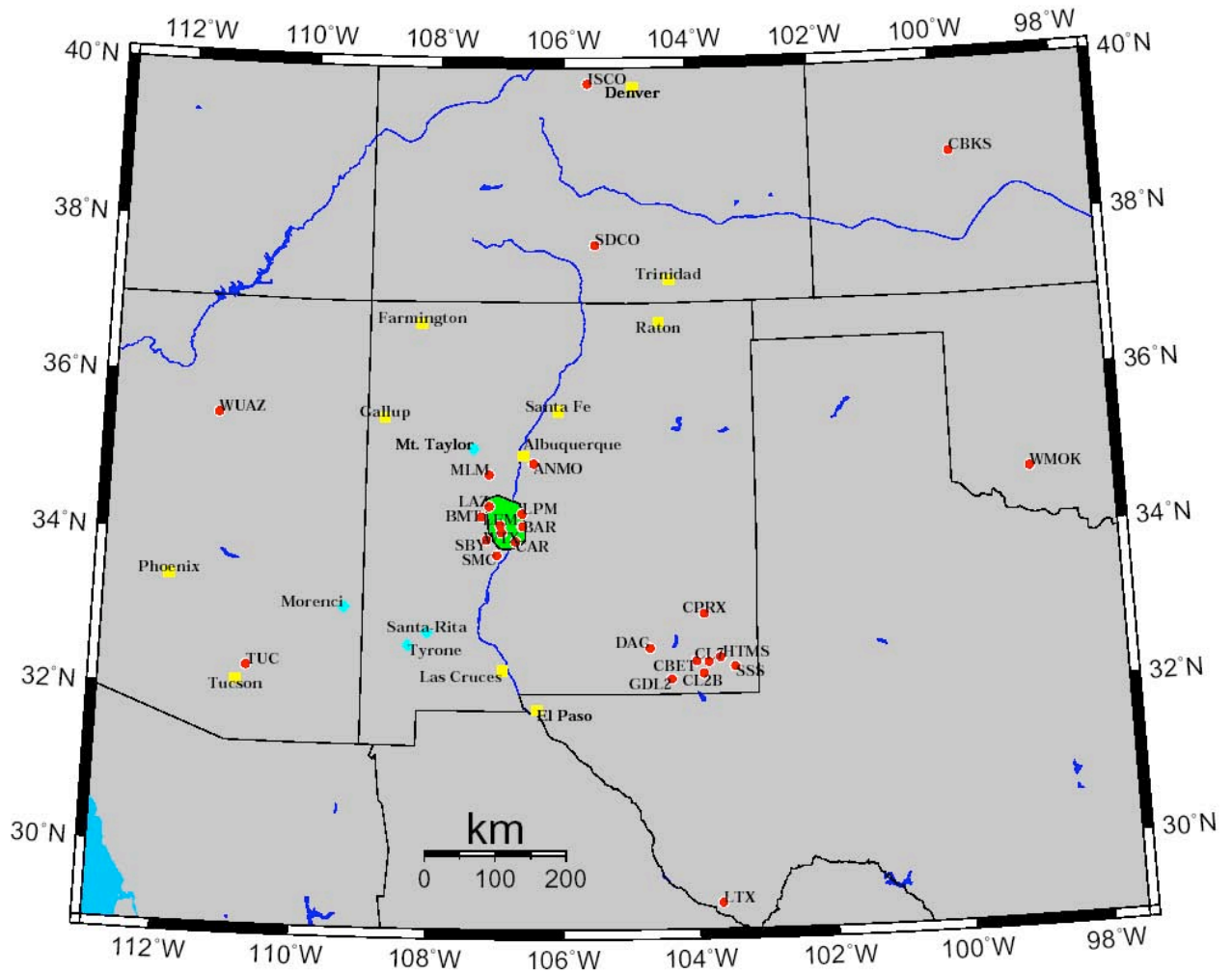


Figure 1-4. Seismic stations around New Mexico. The green area is the Socorro Seismic Anomaly (SSA), the red points are seismic stations, and the cyan diamonds are major mines.

Table 1-1. Seismic Station list (1997-March, 2004)

Station	Network	Component	Latitude	Longitude	Altitude(m)
BAR	SC	ZEHZ	34.150	-106.628	2121
BMT	SC	ZEHZ	34.275	-107.260	1987
CAR	SC	ZEHZ	33.953	-106.735	1658
CBET	SC	ZEHZ	32.421	-103.990	1042
CL2B	SC	ZEHZ	32.264	-103.879	1045
CL7	SC	ZEHZ	32.413	-103.807	1033
CPRX	SC	ZEHZ	33.031	-103.867	1356
GDL2	SC	ZEHZ	32.200	-104.364	1213
HTMS	SC	ZEHZ	32.473	-103.634	1192
LAZ	SC	ZEHZ	34.402	-107.139	1878
LEM	SC	ZEH(Z,N,E)	34.166	-106.974	1698
LPM	SC	ZEHZ	34.312	-106.632	1737
MLM	SC	ZEHZ	34.814	-107.145	2088
SBY	SC	ZEHZ	33.975	-107.181	3230
SMC	SC	ZEHZ	33.779	-107.019	1560
SSS	SC	ZEHZ	32.355	-103.397	1073
WTX	SC	ZEHZ	34.072	-106.946	1555
ANMO	IU	ZBHZ	34.946	-106.457	1853
SDCO	US	ZBHZ	37.746	-105.501	2569
WMOK	US	ZBHZ	34.738	-98.781	486
TUC	US	ZBHZ	32.310	-110.784	906
ISCO	US	ZBHZ	39.800	-105.613	2743
WUAZ	US	ZBHZ	35.517	-111.374	1592
LTX	US	ZBHZ	29.334	-103.667	1013
CBKS	US	ZBHZ	38.814	-99.737	677
DAG	SC	ZEHZ	32.591	-104.692	1396.7

(ps: After May, 2004, station HTMS was replaced by SRH)

Digitized data from 26 seismic stations in and outside New Mexico has been collected for processing at the New Mexico Seismological Observatory since 1992. In 1999, the Earthworm system was set up at NMSN to manage data flow and produce automatic triggered waveforms files from automatic triggers. Since the beginning of 2003, the digitized seismic data (i.e. waveform data) from the NMSC network has been archived at IRIRS-DMC

in Seattle. This was done so that researchers all over the world could download the data from the IRIS-DMC website and to provide an independent long-term backup archive.

Out of the 26 seismic stations in the region, 18 of them are equipped with short period seismometers maintained by the New Mexico Tech and are known as the SC (the associated Federation of Digital Seismic Networks code) network. The SC network is concentrated in two areas, Socorro area and Carlsbad. Ten seismic stations are located to study seismicity associated to the SSA, and 8 seismic stations are set to monitor the seismicity localized around the Waste Isolation Pilot Plant (WIPP) near Carlsbad. The sampling rates of these short period seismometers are 100 samples per second (sps). Station LEM is the only three-component short period station used in this study.

There are 8 broadband seismic stations operated by other organizations around New Mexico. All except ANMO (Albuquerque, New Mexico) are located outside of the state of New Mexico. These stations compensate for the short coverage of the SC network in the boundary area of New Mexico to some degree. All the broadband stations data used has a sampling rate of 40 sps, except for TUC (Tucson), which is 20 sps.

NMSN uses the Earthworm v6.2 system (<http://folkworm.ceri.memphis.edu/ew-doc/>), developed by the USGS, to digitize analog signals and store the continuous data for the last seven days period on a local hard disk. IRIGE signal from GPS is used to time stamp the digitized trace data. In Earthworm system, the module *picker\_ew* picks the P-arrivals using STA/LTA method, and the module *binder\_ew* associates these P-arrivals with events. The module *trig2disk* saves a user defined section of seismic data (usually 180 seconds) of to

disk as an event. Due to the existence of electric noise within the seismic signal transmission system (i.e. signal telemetry from field-site location to the data-storage facility via phone line, or microwave transmission), less than half of the triggered events generated by the *trig2disk* module each month are real seismic events. Because this false-positive event detection rate, we are looking into at a method to adaptively winnow out the noise events from the real seismic events.

### **Major Seismic Source Regions in New Mexico**

Although the northwestern area of the New Mexico is part of the Colorado Plateau, and the Rio Grande rift bisects New Mexico down the middle, there is no obvious tectonic province association connected with seismicity. Major seismic source regions in New Mexico are relatively independent from each other.

One major source region is the Socorro Seismic Anomaly (SSA), which is located in the central Rio Grande rift of New Mexico and overlies the Socorro Magma Body (SMB). The SSA has been estimated to have an area of about 5,000 km<sup>2</sup> ([4]) (green area in Figure 1-4). Nearly 36% of the naturally occurring earthquakes occurring in New Mexico are concentrated in this area, even though it covers only 1.6% of the total area of New Mexico ([4]). Some events occurring in the SSA region generate reflected P and S phases (PzP, SzP and SzS phase arrivals) allowing researchers ([4]) to invert and to estimate the depth and shape of the Socorro magma body and also to get more accurate estimations of event depth. Seismic swarms and repeat earthquakes often occur inside the SSA, possibly being generated by the injection of shallow magmatic materials into the crust at the depths above 12 km

([12]). The largest recent seismic swarms inside the SSA occurred during October 30 and November 2, 2002, with about 246 micro-earthquakes recorded automatically by the Earthworm real-time system. Most of them had obvious waveforms recorded at station WTX (Figure 1-5).

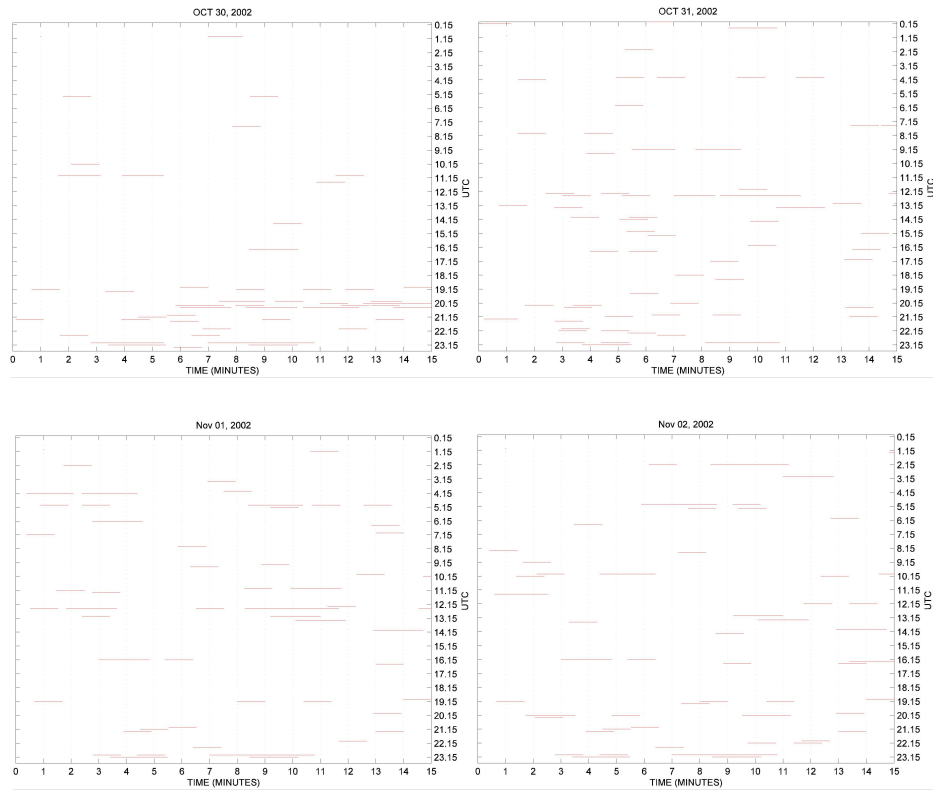


Figure 1-5 WTX pseudo-helicorder plots show the SSA swarm during Oct 30–Nov 2, 2002. Each line represents one event

The Raton area is another relatively active seismic region inside the New Mexico. In this case the seismicity is possibly enhanced by mining and associated injection activity. There have been several large earthquakes in the area between Raton and Trinidad Colorado in recent years. In the period from August 28 to September 21, 2001, twelve earthquakes were recorded by the USGS National Earthquake Information Center. The magnitude of the largest event was  $M_l$  4.6. Figure 1-6 shows three middle-size events that occurred in the

Raton area between June and August 2003. On March 22, 2004, another large earthquake with a magnitude of 4.4 occurred in the Raton area.

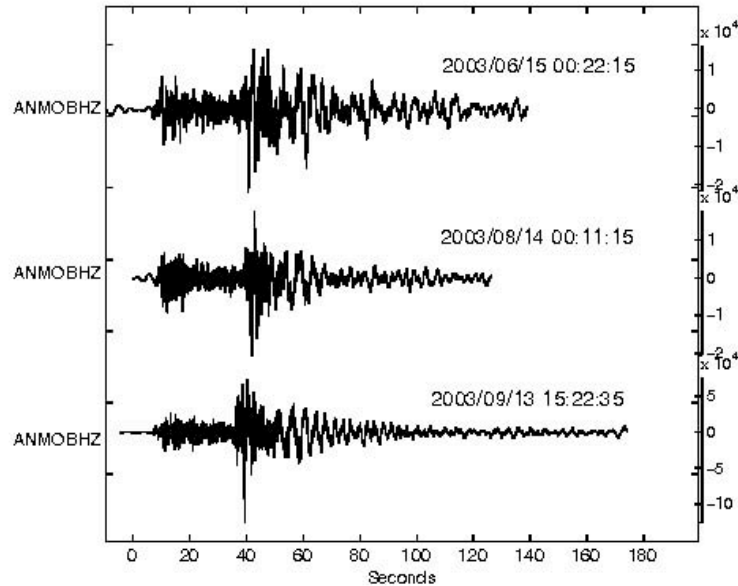


Figure 1-6. Similar earthquakes that occurred in the Raton area

Teleseismic events with large magnitudes often trigger the Earthworm system and may thus be initially improperly identified as local earthquakes. For this study of New Mexico seismicity, we needed to filter out these events. The earth itself acts as a low pass filter since teleseismic events have to travel much longer ray paths compared to local and regional events. It is therefore easy to differentiate teleseismic from regional seismic events by comparing spectrograms or by applying a high-pass filter to the signals. The main energy of teleseismic signal at short period stations is concentrated in 1~2 Hz band, whereas local and regional seismic signals contain high frequency energy up to 10s of Hz (Figure 1-7).



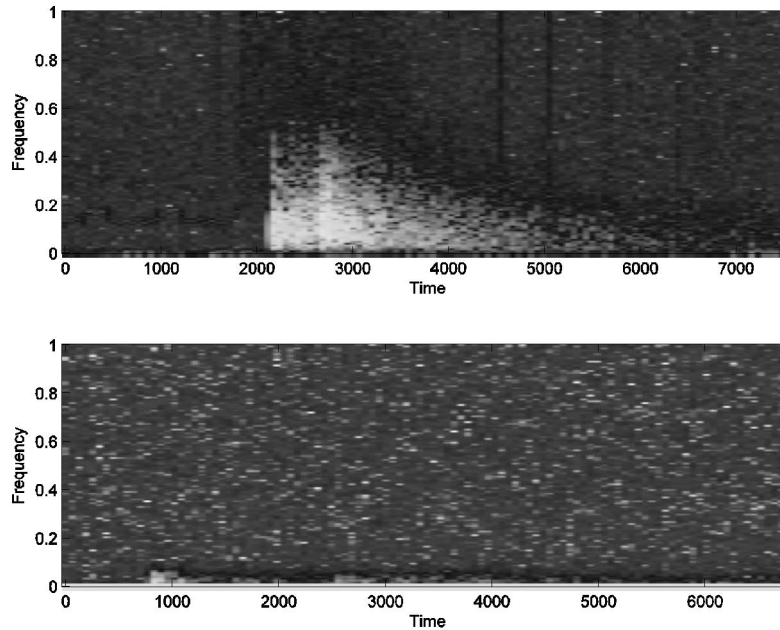


Figure 1-7. Spectrograms of a local (upper) and a teleseismic event (bottom) recorded by WTX. The vertical axis is labeled by normalized frequency (0 to 50 Hz).

Besides natural earthquakes, manmade seismic events (e.g., mining explosions) are often recorded by the SC network. Four major regions where routine mining blasting occurs are Morenci (southeastern corner of Arizona; cyan point in Figure 1-4), Tyrone (southwest corner of New Mexico), Santa-Rita (to the east of Tyrone) and Mt. Taylor (to the northwest of Albuquerque). Repeated earthquakes have often been observed around the station DAG since 2003, and they are likely induced seismic events ([2]).

### **Characteristic Waveform Patterns of different source regions**

When waveforms from different source regions are recorded at a signal station, they show different patterns or signatures due to differing Green's functions associated with their propagation paths. Generally, waveforms that come from the same source region share similar characteristics. In the following, I will describe the characteristic waveform characteristics for seismograms from different source regions by comparing waveforms recorded at station LEM.

The vertical component waveforms recorded by LEM from several major source regions are shown in Figure 1-8. Events from the SSA usually have very sharp P and S arrivals, and some of them have observable SzS, SzS or PzP (magma body upper surface reflected) phases. The S-P time is about 2 seconds at LEM for events from the SSA region (Figure 1-8(c)). The waveforms of SSA events decay very quickly into the background noise level, since most earthquakes observed from the SSA are microearthquakes with magnitudes of 1 or less.

Mining events from the Mt. Taylor region have relatively weak S phases and obvious Lg coda caused by the crustal resonance of the seismic waves (personal communication with Darren Hart). The S-P time at LAZ is about 17 seconds (Figure 1-8(b)). Blasting activity has also been observed in the region north of Gallup, New Mexico. Gallup is further to the northwest of Mt. Taylor from station LEM. The Pg phase arrivals from these events are as weak as the Sg phase arrivals. The S-P time is about 29 seconds for events in the Gallup region (Figure 1-8(a)). Pn phases are the first arrivals and can be clearly detected on the waveforms.

The waveforms from the Tyrone copper mine field have obvious Pn, Pg and Sg phases. The S-Pg time is about 29 seconds (Figure 1-8(d)). The waveforms from the Santa-Rita open-pit mining field only have obvious Pg and S phases. The S-P time is about 25 seconds (Figure 1-8(f)).

The waveforms from the Morenci open-pit mining field have obvious Pn, Pg and S phases. The S-Pg time is about 30 seconds (Figure 1-8(h)). There is more than one kind of shooting type in Morenic area that affect the waveforms; beside single-shot blasts, double-shot and ripple-fire blasts are sometimes used.

The waveforms from the Carlsbad (WIPP) area have obvious P and S phases. The S-P time is about 35 seconds (Figure 1-8(e)). The amplitudes of P and S phases are nearly equal to each other.

The waveforms from Raton area have P and S phases visible, but both of them are relatively weak. The S-P time is about 40 seconds.

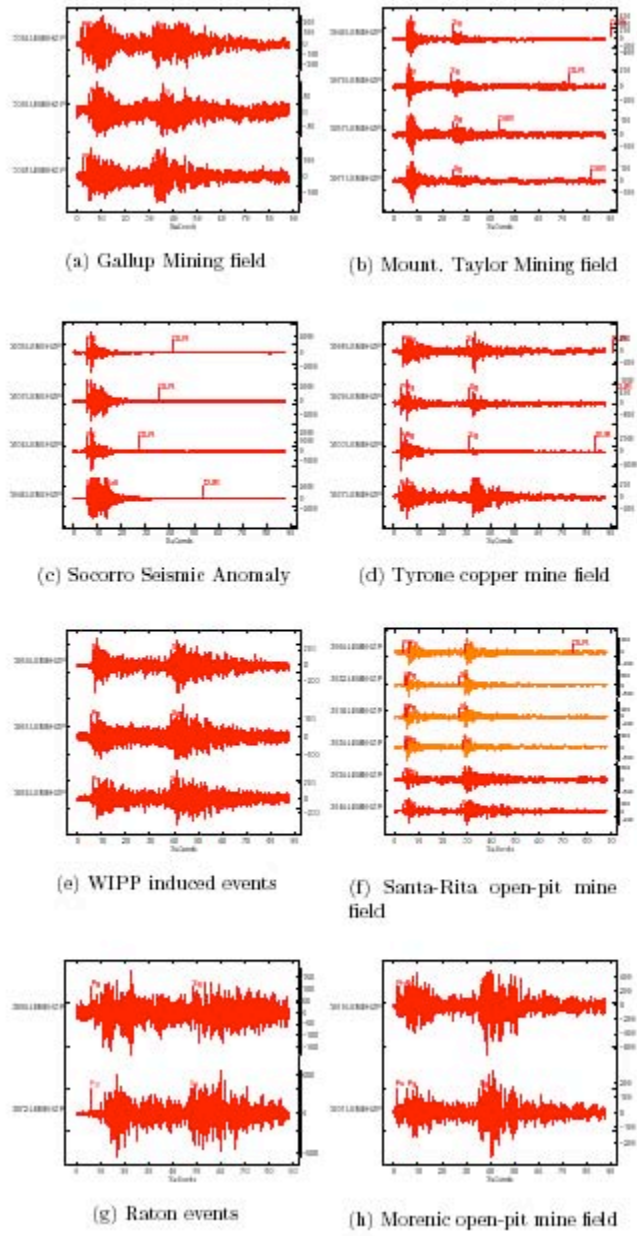


Figure 1-8. Some characteristic vertical component waveform patterns from major source regions recorded by station LEM, vertical component.

**Pervious research**

Dr. Allan Sanford began research on New Mexico earthquakes and seismicity based on instrumental data in the late of 1950's. Sanford and Holmes (1961) observed strong distinct

arrivals other than direct P and S on the seismograms in the Socorro area, leading to the discovery of Socorro Magma Body (SMB). Hans Hartse (1991) systematically summarized former progress on the structure research of the central New Mexico and used direct and reflected micro-earthquake phases to estimate hypocenters and to invert for a layered crustal velocity model. This velocity model (Figure 1-9) is the standard model used to locate earthquakes in New Mexico since the early 1990s.

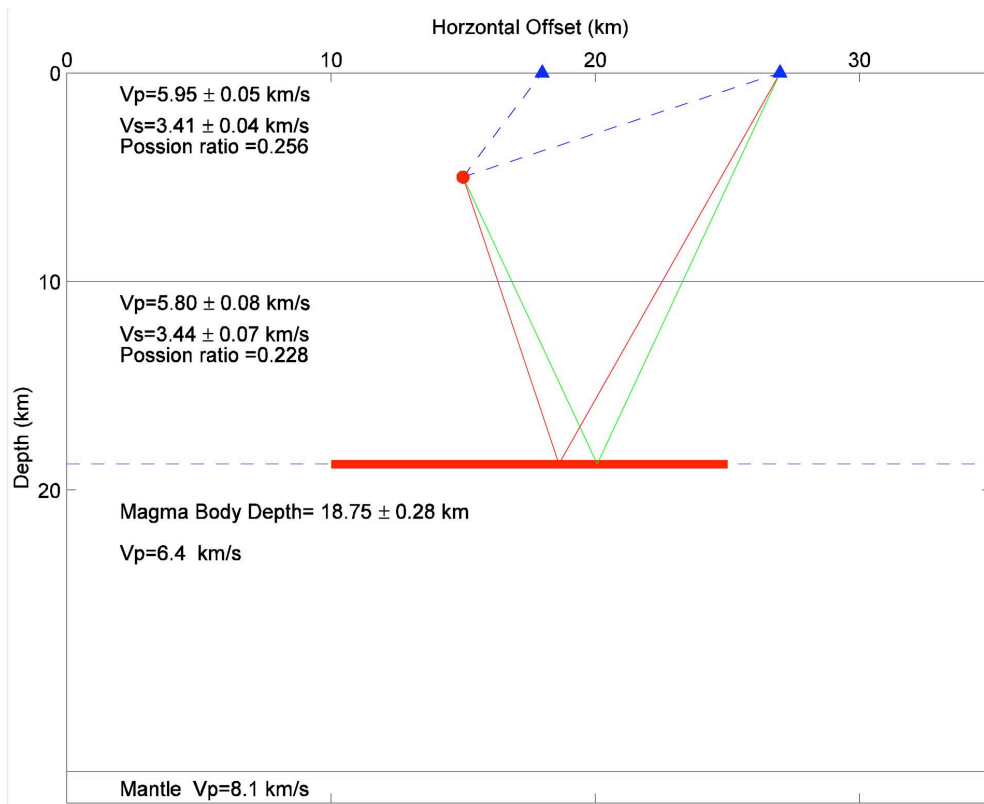


Figure 1-9. Velocity model in Socorro area (modified from Hartse (1991)). The solid circle is a hypothetical earthquake. Triangles indicate seismic stations. The reflecting layer at 19 km is the nearly-flat top of the Socorro Magma body. The dashed lines between the hypocenter and stations are a direct P and S waves, the reflected lines between the hypocenter and stations are PzP or SzS (green; incident angle is equal to reflect angle) and SzP (red; reflected angle is larger than the incident angle).

With this crustal velocity model, regional theoretical travel-time tables for seismic events in New Mexico can be built to help researchers to pick seismic phases. The travel-time curves for several often detected seismic phases are shown in Figure 1-10.

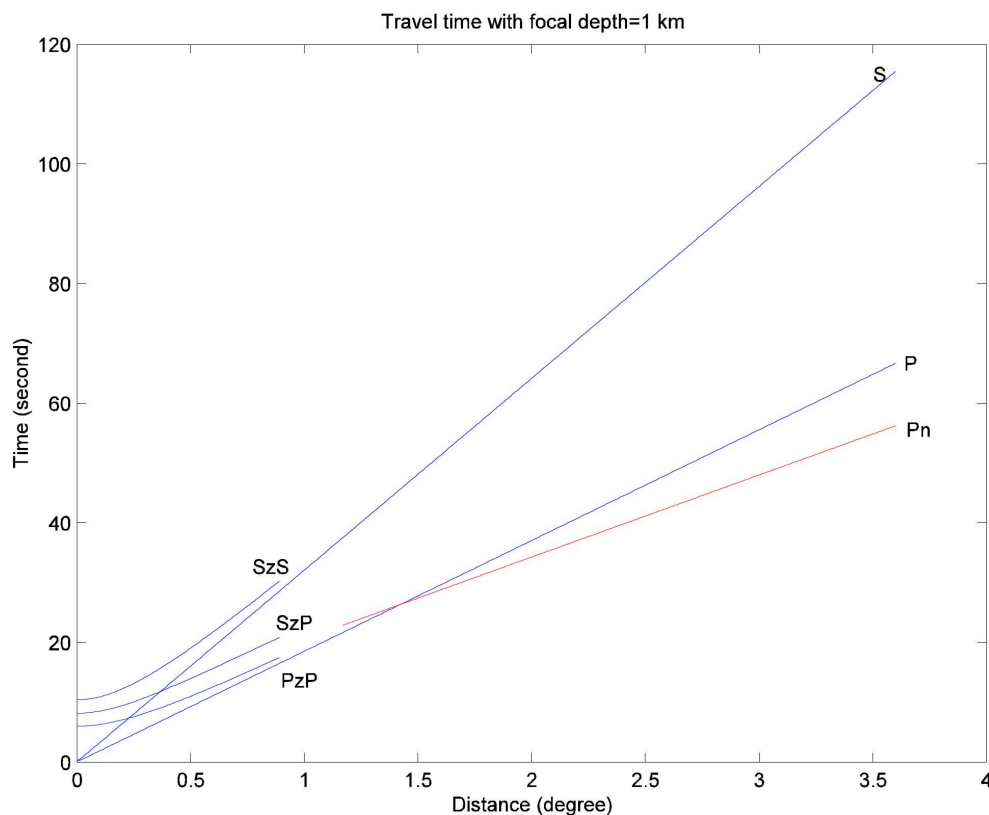


Figure 1-10. Theoretical travel-time curves for several often detectable seismic phases in New Mexico for earthquakes with depth=1km.

Before the digital data acquisition system was installed in 1991 (Zhang, 1997), seismic arrival times were commonly read and picked from analog paper records. The original digital data acquisition system used an algorithm distributed by IASPEI called **XDETECT** to detect seismic events. **XDETECT**'s algorithm was based on the common **STA/LTA** (the ratio between short term average (**STA**) and long term average (**LTA**)) seismic phase detection method. The **XPICK** picking program and the **Seismos** location program (Hartse, 1991) were used to locate earthquakes in New Mexico. Withers (1997) developed a method called

**LWCEDS** (Local Waveform Correlation Event-Detection System) and used it to detect earthquakes and also estimate coarse location, magnitude and original time based on processed waveform correlation. The seismic data was saved in **CSS-3.0** (Center for Seismic Studies Version 3) format. Zhang (1999) used **Matseis** to process seismic data of New Mexico from July 1997 to February 1998. Young et al (2001) used a dendrogram-based hierarchical tool to classify regional events of New Mexico into several sources regions. In 1999, the **Earthworm** system was setup to maintain seismic data from the SC network by with assistance from the US Geological Survey. From the middle of 2000, all possible seismic events automatically triggered by **Earthworm** were saved on hard disk in **AH** or **SAC** format. Only the raw waveforms were saved without any phase information available, and more than half of them were not earthquakes or mining blasts but were instead noise. Rowe et al (2002) developed an automatic, adaptive algorithm to refine the phase picking, but this method requires that all waveforms have initial phase picks produced by an analyst or autopicker. To use the method developed by Rowe et al. to relocate earthquakes that occurring in New Mexico, it would be necessary to reprocess the original triggered seismic events generated by **Earthworm**. The following chapters of this report will discuss an adaptive, automatic method which uses a cross-correlation technique to detect the location of seismic events and identify initial phases.

## WAVEFORM CROSS-CORRELATION TECHNIQUE

### Waveform Cross-correlation

In linear vector space, the distance between two vector spaces (e.g. two time series with a relative lag of  $t$ ) could be measured by  $L-2$  norm (*Euclidean* distance) as:

$$d^2(f, g) = \|f - g\|^2 = \int_0^{\infty} [f(\tau) - g(\tau - t)]^2 d\tau . \quad (1)$$

The definition of cross-correlation for two continuous time series  $f(t)$  and  $g(t)$  is:

$$c(f, g) = \int_0^{\infty} f(t)g(t - \tau)dt \quad (2)$$

$c(f, g)$  is the cross-correlation coefficient for the two time series at a particular time lag.

When  $f(t)$  and  $g(t)$  are discrete digitized data, the normalized cross-correlation expression is:

$$c_d = \frac{\sum_i [(f_i - m_f)(g_{i-d} - m_g)]}{\sqrt{\sum_i (f_i - m_f)^2} \sqrt{\sum_i (g_{i-d} - m_g)^2}} . \quad (3)$$

Where  $i=0, 1, 2 \dots N-1$  and  $N$  is the number of samples in  $f$  and  $g$ .  $c_d$  is the cross-correlation coefficient value at sample lag  $d$ , and it ranges from -1.0 to 1.0.  $m_f$  and  $m_g$  are the mean of time series  $f$  and  $g$  respectively.  $f_i$  and  $g_{i-d}$  are the  $i$ -th or  $i-d$ -th sample value of  $f$  and  $g$  respectively. The purpose of the summation in the denominator is to normalize the cross-correlation coefficient, so that similarities between several time series pairs can be compared irrespective of amplitude.



Figure 2-1 shows how cross-correlation works using two seismograms. When two equal-length waveforms are aligned and cross-correlated the possible sample lags range from  $-(N-1)$  to  $(N-1)$ .

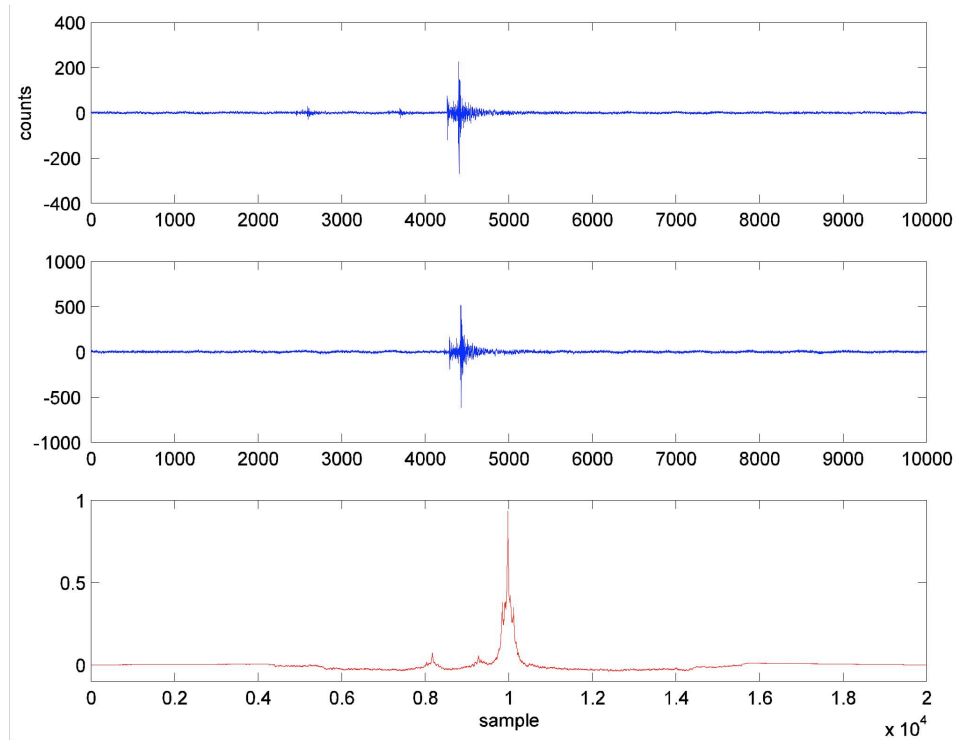


Figure 2-1. Cross-correlation coefficient curve (bottom plot) between waveforms of two similar earthquakes recorded by station WTX. The upper waveform is an earthquake that occurred at 1997-12-04 02:15, and the middle waveform is an earthquake that occurred at 1997-12-03 21:33. Both waveforms have 100 s of data with a sampling rate of 100 samples per second. The maximum cross-correlation coefficient is 0.9 at time lag equaling zero (the time segments were chosen to align on the P arrivals).

Since cross-correlation between two limited time series can be expressed as convolution between one limited time series and the time reverse of the second limited time series cross-correlation in seismology study has also been documented as (Papanicolaou, 2004):

$$I^{INT}(y^s) = \sum_{x_r, x_{r'}} P(x_r, \cdot) *_t P(x_{r'}, \cdot) | \tau(x_r, y^s) - \tau(x_{r'}, y^s). \quad (4)$$

Where  $y^s$  is the station, and  $x_r$  and  $x_{r'}$  are two sources.  $I^{INT}(y^s)$  is the maximum cross-correlation coefficient of two time series (waveform data:  $P(x_r, \cdot)$  and time reverse  $P(x_{r'}, \cdot)$ ) at time lag  $\tau(x_r, y^s) - \tau(x_{r'}, y^s)$ .

Cross-correlation techniques have been widely used in seismological research. Song (1996) used the cross-correlation method to compare seismic waveforms passing through the inner and outer core during the last 30 years and found a systematic time shift between the  $PkP(DF)$  and  $PkP(BC)$  phases. Aster et. al. (1991) used the waveform cross-correlation method to evaluate possible temporal anisotropic variations between similar waveforms and ruled out previous claims of using shear wave splitting for earthquake prediction.

When we are going to use the cross-correlation method to study seismic waveforms, there are several necessary conditions. First, the geological structure background of ray paths between several epicenters and the seismic station does not vary greatly (i.e. earth structure along commonly traveled ray paths are similar). Second, the source radiation patterns of closely spaced earthquakes are similar (i.e. source characteristics are comparable).

### **Cross-correlation coefficient Matrix**

The similarity of two waveforms is measured by the cross-correlation maximum, and is quantitatively expressed in the cross-correlation coefficient (CC<sub>c</sub>). When considering  $N$  waveforms from different sources and received at the same station,  $N(N-1)/2$  cross-correlation coefficients are calculated, which forms a symmetric cross-correlation coefficient

matrix with  $N \times N$  elements (e.g. Lower left plot in Figure 2-2). The  $CC_c$  matrix is a way to visualize the similarity between each waveform pair. By incorporating the  $CC_c$  matrix with cluster analysis techniques (e.g. dendrograms – upper left plot of Figure 2-2), waveforms from different source regions could be connected in such a way so as to graphically depict the degree to which the waveforms are similar.

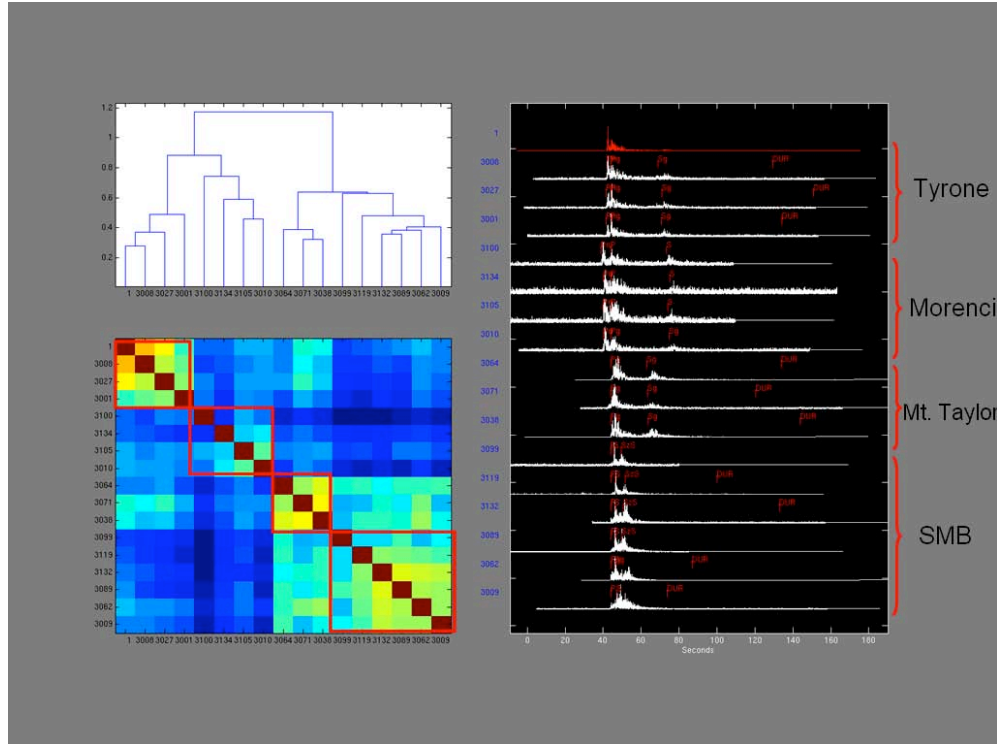


Figure 2-2. 19 reference P-phase aligned vertical-component waveforms recorded by station LEM (right), the  $CC_c$  matrix of each waveform pair (left, bottom) and the corresponding dendrogram (left, top). The index of the coefficient matrix and the waveform plot are sorted according to the dendrogram result. Waveforms are classified into four groups as marked by red squares in the matrix plot, and are labeled with the source region name in the waveform plot.

The  $CC_c$  matrix can be used to illustrate the waveforms grouped results calculated from the cluster analysis (e.g. Bergthora et al., 1987). One method proposed by Young. et al., (2001) is to group the  $CC_c$  matrix using a dendrogram cluster analysis technique (left top plot of Figure 2-2). The fundamental idea behind these clustering methods is to re-arrange the  $CC_c$  matrix

from random patterns into connected groups. In our application we would like each group to only contain waveforms generated from the same source region.

## **THE DEVELOPMENT OF PLRR**

### **MatSeis**

**Matlab** is not only a very popular tool for numerical computations, but it is also an advanced programming language comparable to **Visual Basic** and **Visual C++**. It provides many Graphic User Interface (GUI) functions that users can use to build interactive windows for analyzing and visualizing data. The **Matlab** package includes a signal processing toolbox (library of algorithms), and because of Matlab's design structure, users are easily able to construct new functions for use by writing Matlab code.

**MatSeis** was first developed as a seismic data processing package under **Matlab** by Mark Harris at Sandia Laboratory, and it relies on using Matlab's core signal processing libraries and programming environment to process seismic data. It is currently maintained and updated by a team at Sandia Laboratory in New Mexico [18]. **MatSeis** has a GUI window style interface, in which users execute algorithms by selecting pop-up menus or by clicking buttons. Since **Matlab** is a command line language, any operation inside Matlab can be executed by inputting command line statements. Therefore, users can write an M-script file to finish a batch of jobs automatically in **MatSeis** in addition to GUI (mouse) operation.

As a platform for seismic data processing, **Matseis** is good for data visualization and handling multiple events. It provides GUI windows that allow users to pick seismic phases, map the location of earthquakes and cluster waveforms into different seismic source regions.

Seismic data when loaded into **Matseis** are organized in the matrix form and managed by four basic functions. *Waveform* handles all the information related to the waveform data (e.g. station, channel, start time, sample rate). *Origin* handles the event origin information (e.g. location, original time, magnitude and location error of event). *Arrival* stores the phase information. *Traveltime* stores the theoretical traveling-time information.

## PLRR

Seismic data from the SC network is maintained by the **Earthworm** system. Every day the *trig2disk* module of **Earthworm** produces seismic event data files on disk based on the *STA/LTA* algorithm. Due to the layout of seismic stations and telemetric spike-like noise, at least half of these automatically generated files do not contain any seismic signals. Moreover, we could not find a good way to combine the seismic phase information generated by the **Earthworm** modules with the *SAC* format seismic event data files and ensure that the seismic phases generated by **Earthworm** modules were of good enough quality to be used as the initial seismic phases for earthquake relocation. We plan on answering the following two questions: How to select real seismic events from these automatic generated files automatically? How to find the more accurate initial phase information and locations for the real seismic events using an automated system?

To address these problems, we developed an adaptive, automatic phase-picking and epicenter-locating program. The program is built upon the **Matlab** environment (under UNIX) and is supported by **Matseis-1.7**. The package is named **PLRR** (**P**ick-**L**ocate-**R**epick-**R**elocate). The fundamental idea in **PLRR** is to use cross-correlation instead of the traditional *STA/LTA*

algorithm to identify seismic events, then associate phase arrivals for certain phases that we are interested in (e.g. P, S, Pn).

The most important function of **PLRR** is to identify the most similar event from a reference database, and then assign phases and location information using spectrogram cross-correlation.

No preliminary information is needed for the unknown event.

The diagrammatic data processing of **PLRR** is illustrated in figure 3-1.

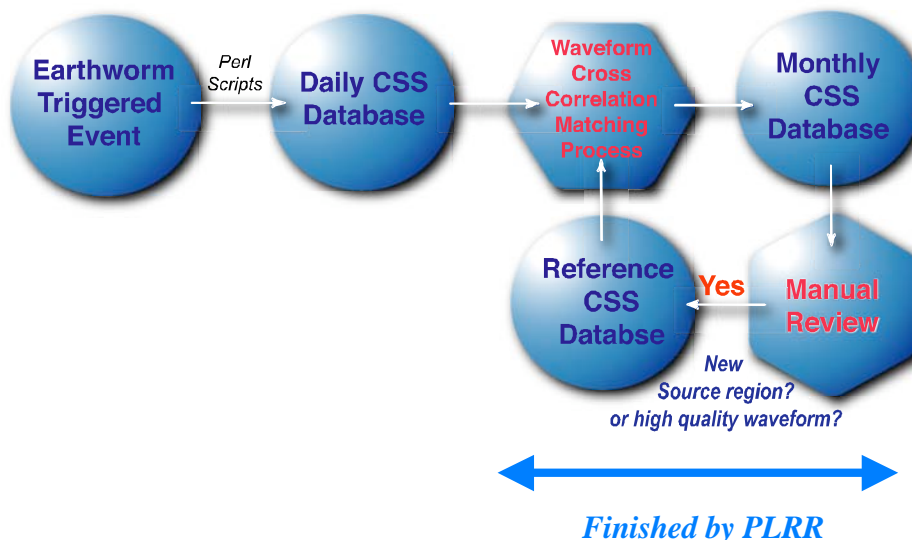


Figure 3-1. Flowchart of **PLRR**

The following sections will discuss the realization of **PLRR** step by step.

### Data Format conversion

The seismic event set format currently accepted in **MatSeis-1.7** is the *CSS3.0* format, which consists of waveform data and several tables containing origin time, arrival time, relationship between waveforms, and identifying numbers. Several of the tables necessary to form a **PLRR**

readable CSS database are: Origin, Wfdisc, Arrival, Assoc, Site and Sitechan. All of these tables are in ASCII format and can thus be edited using **vi** or **emacs**.

The seismic event files automatically generated by the Earthworm system are in *SAC* or *AH* format. Therefore, the first step of PLRR is to convert the data format of **Earthworm** output into the format which **Matseis** can read. A Perl script was written to perform this work daily. The source code follows with filename “**dailycss**” ( see reference II).

The output files from Earthworm are saved each month under the “/raid/data/Eworm/Triggers/incoming/yyyymm” directory and in **SAC** format. The file name has the format of “yyyyymmdd\_hhmmss\_quakeid”, in which the “quakeid” is assigned by **Earthworm**. The saved section duration of each waveform is set to be 180 seconds. All of the above sets could be modified in the “*trig2disk.d*” config file of **Earthworm**.

With the help of the **UNIX** command “**crontab**”, the dailycss could be executed once every day. The following sentence is set in crontab to run dailycss at 20:00 everyday:

```
03 20 * * * /raid/users/wyang/outgoing/dailycss
```

The outputs of “**dailycss**” are in CSS-3.0 format and saved under the “/raid/data/Eworm/Triggers/outgoing” directory. Each day, two new tables are built: “yyyyymmdd.origin” and “yyyyymmdd.wfdisc”, and the corresponding waveforms of one event are saved under the “yyyyymm” subdirectory with the name “yyyyymmddhhmmss”. We do not generate new “site” and “sitechan” tables every day, because they are typically stationary.



“**dailycss**” adjusts another **Perl** function named “**foreach.pl**” (programmed by Andreas Gohr: a.gohr@web.de ) and a shell script named “**sac2css\_sc**” with the following contents:

```
#!/bin/csh
set echo
set prefix = $2
sac << END
read $1/*
writecss $1/$prefix
END
```

### Reference seismic dataset

The reference dataset consists of many well-located seismic events, with complete seismic phases picked and high signal-to-noise ratio waveforms. It is essentially a knowledge base of historical seismic events for the New Mexico region with corresponding phase information. Since earthquakes tend to reoccur in the same source region over the years, the waveform similarity between a new event and old events (reference events) occurring in the same source region can be high. By applying the cross-correlation technique, and supposing the reference seismic dataset is completed, we can find the location and phases of an unknown seismic event far more precisely and robustly than the results given by traditional **STA/LTA** algorithm.

As shown in Figure 3-2, the reference seismic dataset currently used in **PLRR** has 153 events (by June, 2004) from all over the state of New Mexico representing typical seismicity. But the contribution from each station in this reference dataset varies considerably (Figure 3-3). For example, the Socorro subnet is located in the center of New Mexico, and these stations may receive higher quality waveforms for events occurred around the Socorro Magma Body. The subnet in Carlsbad in the southeastern corner of New Mexico is very sensitive to nearby events, but are nearly blind to microearthquakes occurring in distant source regions. The

stations located outside New Mexico are similarly most sensitive to small events occurring nearby. For example, station TUC typically receives high quality waveforms only from Morenci open-pit mine in our area of interest.

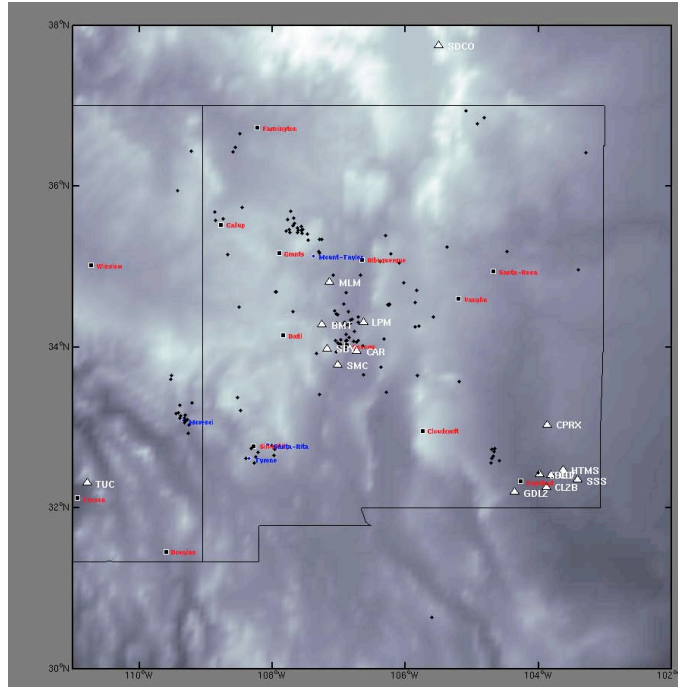


Figure 3-2. Epicenter distribution of reference seismic dataset for New Mexico (as of June, 2004; 153 events selected from the seismic data of 39 months)

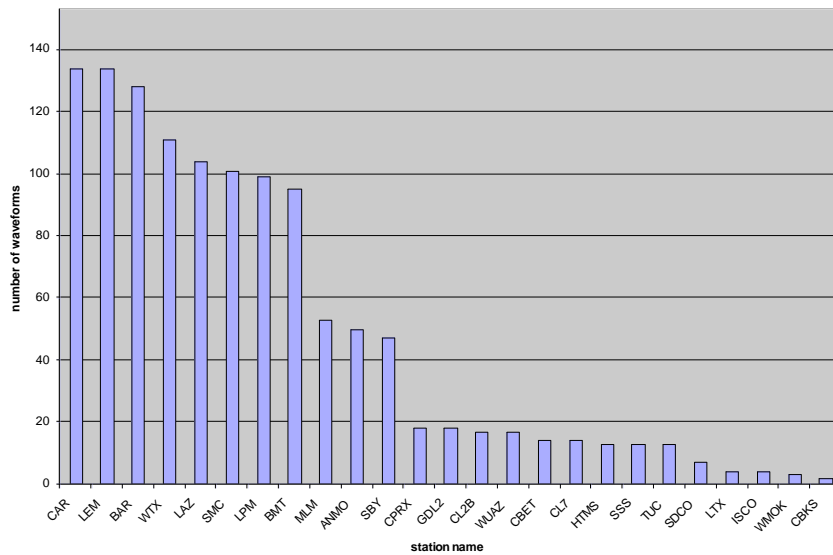


Figure 3-3. Station contributions to the reference seismic dataset (153 events).

As opposed to the traditional automatic seismic phase detection algorithm, the **PLRR** is an adaptive automatic seismic phase detection process. Its seismic phase detection capability will ultimately depend on the coverage and quality of the reference seismic dataset. When a new type of event occurs that does not correlate well with any reference event (with  $CC_c$  over a fixed threshold, e.g. 0.3), it will be added into the reference dataset if it has high quality waveforms, thus improving the reference dataset over time.

### **Strategy of phase detection and epicenter location**

For a new seismic event, there may exist several very similar events in the reference dataset, there are many cluster analysis methods available for use to address this situation based on different strategies. Here we prefer to use the **K-most-similar** algorithm proposed by Carmel et. al. (2003). It is stated as the follows:

1. Define a reference set, i.e. a set of representative measurements for which the class association is known in advance.
2. Given an unknown sample, calculate its similarity index with respect to each of the measurements in the reference set.
3. Find the K most similar measurements from within the reference set.
4. Associate the unknown sample with the cluster that contains the majority of the K most similar measurement.

In **PLRR**, we use the default  $K=1$ , which simplifies the strategy to simply be the most similar waveform.

## Waveform-pair cross-correlation matching

There are two waveform cross-correlation methods tested in **PLRR** to detect seismic phases and find locations for new events. The first one is *waveform-pair cross-correlation matching*. The schematic of this method is shown in Figure 3-4.

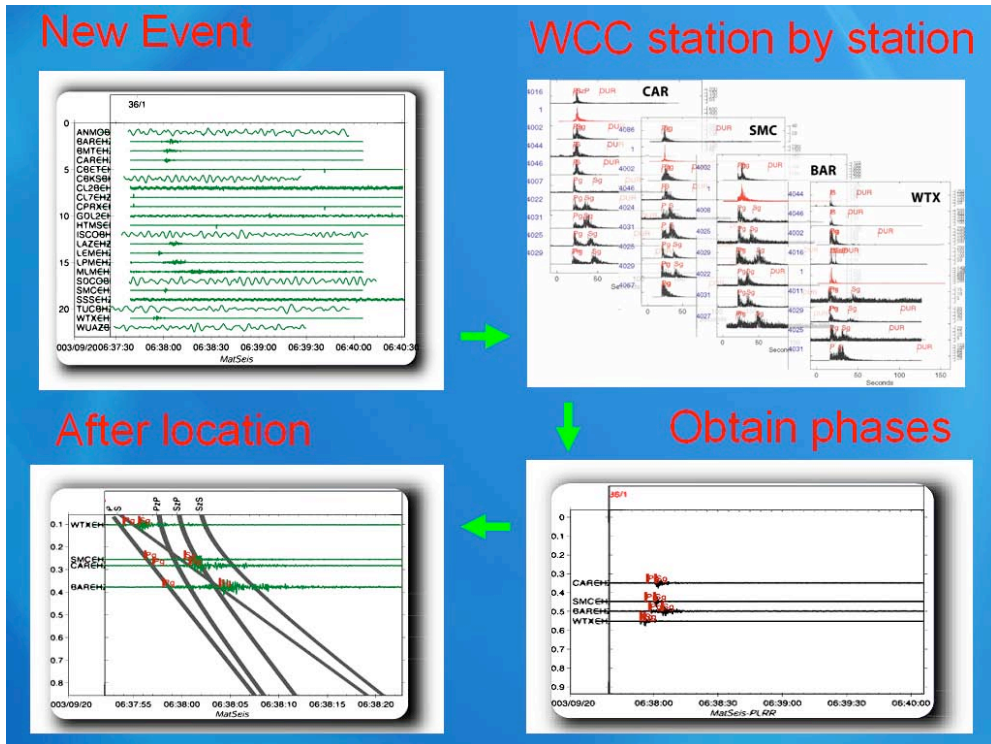


Figure 3-4. Illustration of waveform-pair cross-correlation matching method

This method calculates the  $CC_c$  of a waveform-pair (one from the new event and one from a reference event) recorded by the same station, and finds the waveform-pair with the highest  $CC_c$  value. Then it assigns the phase information of the reference waveform to the new waveform to provide initial picks. After finding phases for each waveform of the new event (each station's waveforms may possibly be correlated with a different reference event), a **Matseis** locating function (*MS Locator Tool*) is used to find the initial epicenter.

This method preprocesses the waveforms in two steps:

(1) All waveforms are high pass filtered with the frequency cutoff equaling 1.0 Hz. The waveforms of teleseismic events will become significantly attenuated after this step and will not match any waveforms in the reference dataset with a high  $CC_c$  value (e.g. 0.2), thus making them easily discounted.

(2) Filtered waveforms are converted into envelope based on the Hilbert transform. The Hilbert transform performs a 90 degree phase shift (of opposite sign at positive and negative frequencies) to a time series without changing its amplitude, and is defined as the following convolution:

$$y(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x(s)}{t-s} ds \quad (3-1)$$

The envelope is given by:

$$E(t) = \sqrt{x^2(t) + y^2(t)} \quad (3-2)$$

By taking the envelope of waveforms prior to cross-correlation, the signal-noise-ratio ( $SNR$ ) and associated correlation robustness is improved relative to what it would be if just using the waveforms had been used. However a drawback of the Waveform-pair cross-correlation matching method is its instability. This technique produced too many false associations, leading to the development of the following method.

### **Event-pair cross-correlation matching**

The second cross-correlation method tested in **PLRR** is *event-pair cross-correlation matching*. The schematic of this method is shown in Figure 3-5. This method stacks the  $CC_c$  curves of waveform-pairs between the unknown event and waveforms grouped from each event in the reference database. Then it finds the reference event that has the highest stacking  $CC_c$  value

(by taking the mean/median of stacking  $CC_c$  curves), and defines it to be the most similar event to the unknown event.

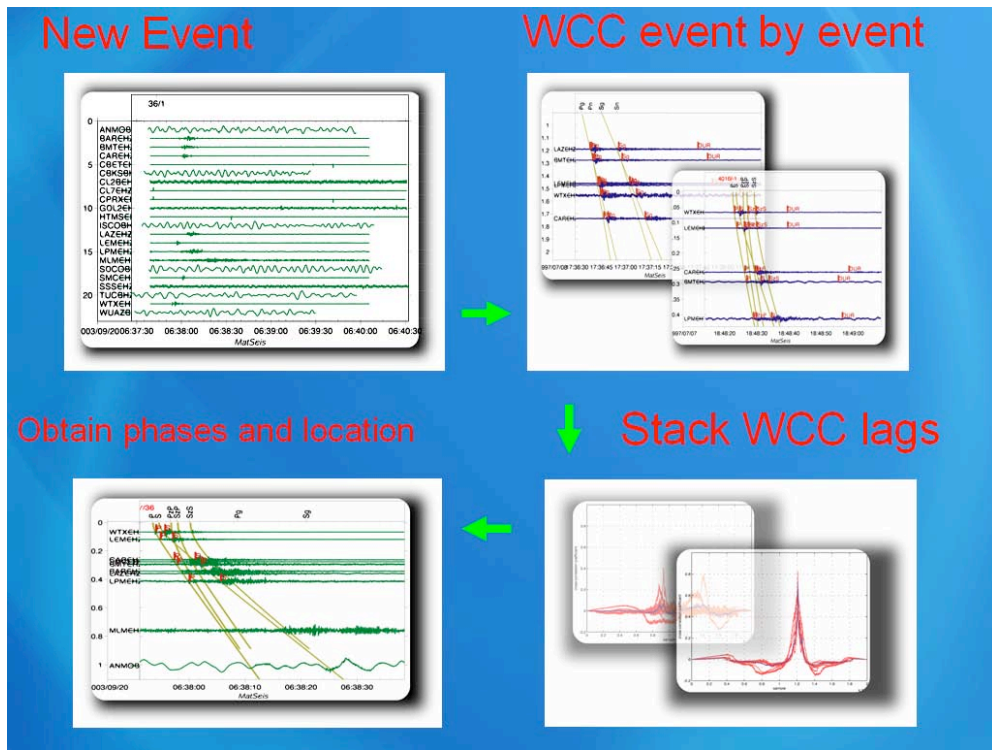


Figure 3-5. Illustration of event-pair cross-correlation matching method

In this method, waveforms with different sampling rates (for example, the sampling rate of CAR is 100 samples per second, and the sampling rate of ANMO is 40 samples per second) have to be converted to the same sampling rate, because the stacking  $CC_c$  curve needs to stack the  $CC_c$  curves from different stations point by point. I resample the low sampling rate waveform to be 100 samples per second. Also in this method, data window lengths for the waveform cross-correlation calculation have to be taken as the same, because waveforms recorded by different stations may not be of the same length in a given event.

Event-pair cross-correlation matching is more robust than the waveform-pair cross-correlation matching. It considers the cross-correlation curves of all waveform-pairs of events as a whole

instead of treating them separately (requiring the waveforms to be similar at all stations for a good match). It must match the time order of seismic phase arrival times of an event across the network. The contribution of each waveform pair ideally enhances the stacking waveform cross-correlation curve.

### **Spectrogram event-pair cross-correlation matching**

*Spectrogram event-pair cross-correlation matching* method is an enhanced method of the *event-pair cross-correlation matching*. When I first processed the seismic data of New Mexico using the *event-pair cross-correlation matching* method, I found that there exists an “attracting” phenomenon: seismic events occurring further away from the center of the SC network are sometimes identified as earthquakes occurring around the center of SC network. One fundamental reason for this phenomenon is that the reference dataset is incomplete and more events should be added to it. Another reason is that the waveforms of earthquakes occurring further away from the station usually have a strong P and weak S phase or vice versa (e.g., seismic events in the Mt. Taylor and Tyrone areas, Figure 1-8). From the point of view of waveform cross-correlation, this type of waveform (only one strong phase) is very similar to waveform of earthquake occurring very close to the station, because local earthquakes have a very short S-P time.

In the time domain, the amplitude of S-phases might be nearly as low as the background noise, but in certain frequency bands, the amplitude of S-phases stands out from the background noise. So one way to solve the “attracting” problem is to convert the waveform data from the time domain into the frequency domain, and only keep the frequency bands with both strong P and S energy for the cross-correlation. Using this knowledge, I developed the *spectrogram event-pair cross-correlation matching* method to calculate the stacking cross-correlation curve in several frequency bands.

The distance from major mining fields around New Mexico to the center of Socorro sub-network is more than 200 km. Waveforms coming from such sources usually have very weak S-phases. As shown in Figure 3-6, a waveform from Tyrone recorded by station LAZ has distinct Pn and P phases but a weak S phase. The normal waveform cross-correlation technique has a strong tendency to detect Pn as P and P as S, thereby potentially misidentifying this event as an earthquake occurring around the SMB.

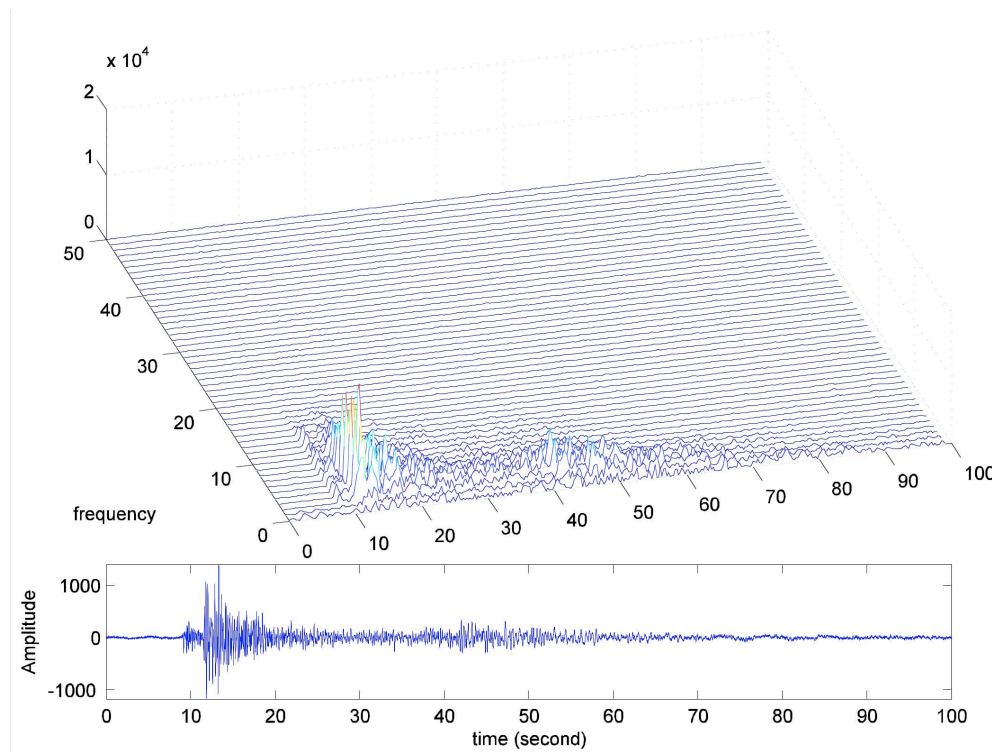


Figure 3-6. Waveform (bottom) of a mining blast from the Tyrone mining field and recorded by station LAZ (222 km to NE of the epicenter) and corresponding sonogram (upper plot). The energy is concentrated 15 Hz. In the band between 0 to 5 Hz, the background noise dominates the signal, making it difficult to apply for the cross-correlation technique. In the band between 6 and 10 Hz, the signal energy is very clean and both P and S phase arrival energy is very strong and easily distinguishable. In the band between 11-15 Hz, the energy again decays rapidly because of attenuation and the limited frequency response of the short period station telemetry.



In the frequency domain, the method selects a suitable frequency band (6-10Hz), and calculates the  $CC_c$  curve of each frequency. Then it stacks these curves and takes the mean of them as the spectrogram  $CC_c$  curve.

Compared with waveform  $CC_c$  curves in the time domain, the spectrogram  $CC_c$  curve has less sample points, sacrificing the accuracy for better similarity robustness. But for the initial phase-picking (in this case, the new event is initially collocated with the most similar reference event), the accuracy given by the *spectrogram cross-correlation matching* method is sufficient.

### **Ground truth test**

Starting at the end of 2003, I began to use the above methods to process the monthly seismic data of New Mexico and compare the location results given by PLRR with the results of careful manual location.

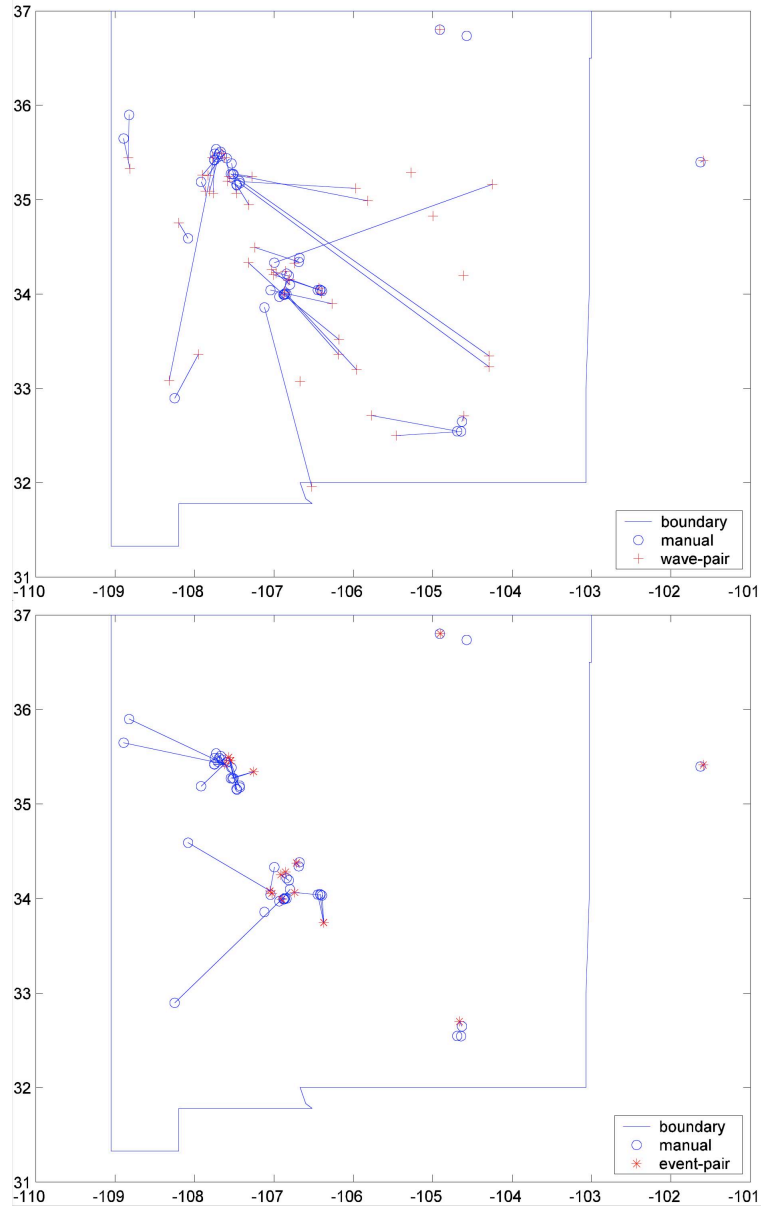


Figure 3-7. Comparison of waveform-pair cross-correlation matching method (upper) and event-pair cross-correlation matching method (bottom). Both methods were used to process the same data in the month of September, 2003. The circle symbols are epicenters of manual location and the star symbols are epicenters given by automatic method. . The line between each star-circle symbol pair is the discrepancy between manual and automatic locations.

In Figure 3-7, seismic data from the same month (September, 2003) were processed using the waveform-pair cross-correlation method and the event-pair cross-correlation method with the same initial condition. The comparison between results given by manual location result and the

results given by the two automatic methods shows that the event-pair cross-correlation matching method is more robust than the waveform-pair cross-correlation matching method (the mislocation vectors are smaller and fewer).

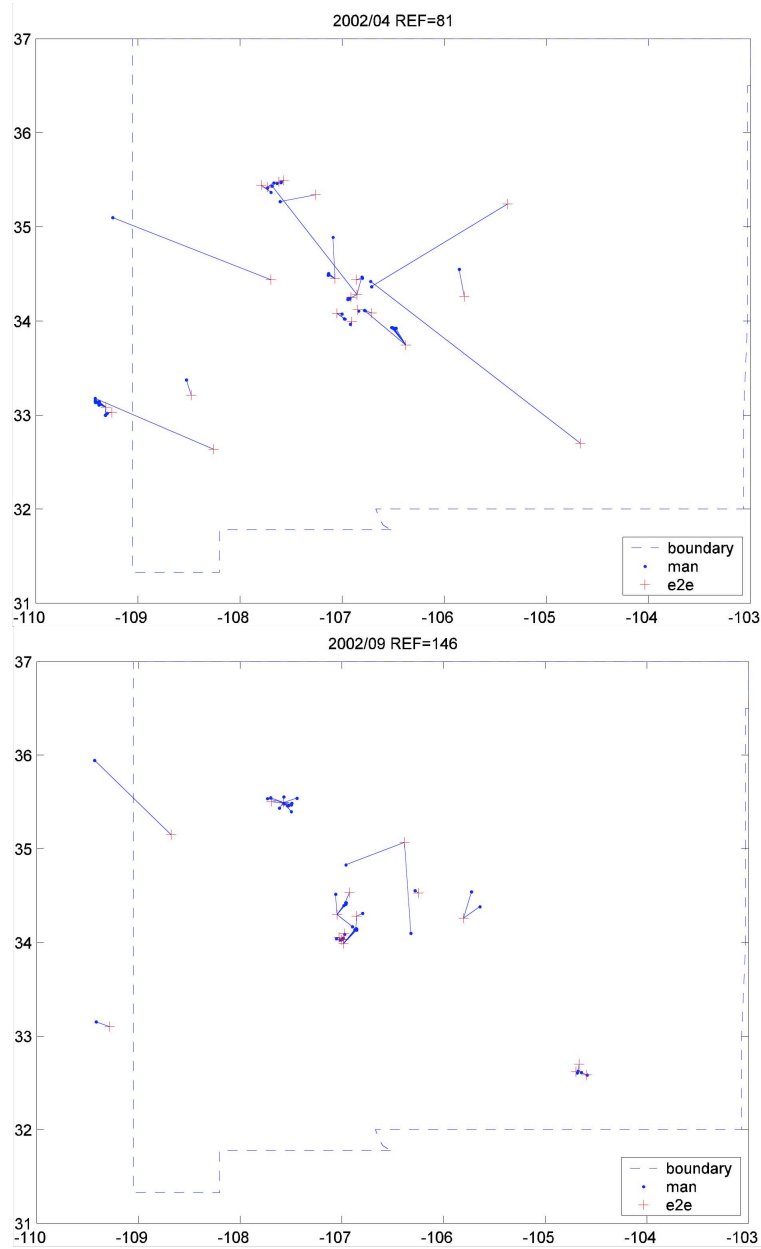


Figure 3-8. The comparison of adaptive automatic location results of two months. The upper plot is the result of April, 2002 and the plot at the bottom is the result of September, 2002. The cross symbols are location results given by PLRR, and the star symbols are manual location results. The line between each star-cross symbol pair is the discrepancy between manual and automatic location.

In Figure 3-8, the seismic event location results by **PLRR** over two months are shown. The left plot is the earlier result processed by *event-pair cross-correlation matching* method with 81 reference events, and the right plot is a recent result processed by *spectrogram event-pair cross-correlation matching* method with 146 reference events. It is very clear that increasing the number of reference events and the use of the spectrogram cross-correlation technique greatly improves the accuracy and stability of local seismic event automatic location.

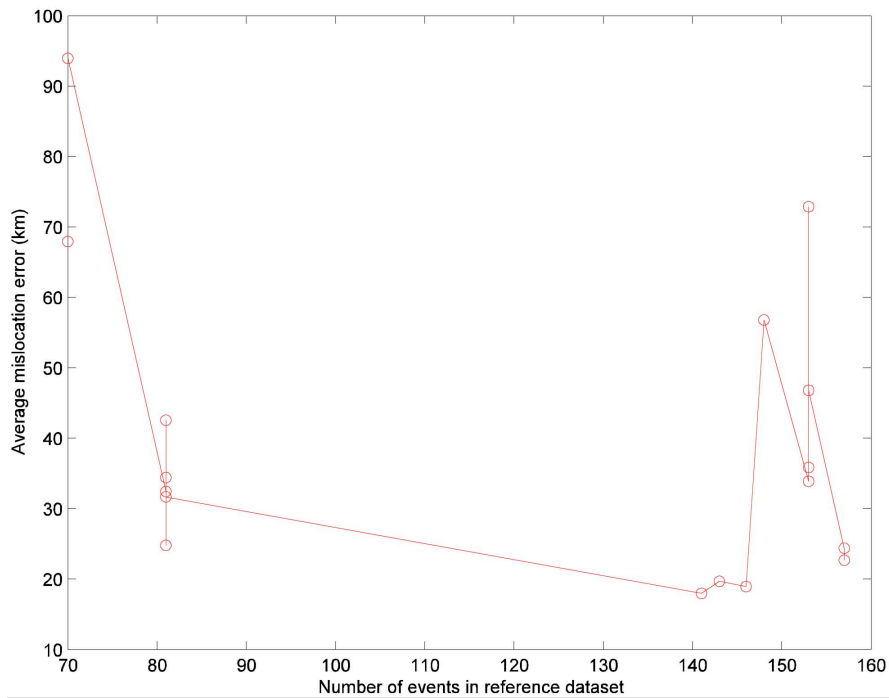


Figure 3-9. The relationship between the average mis-location error and the number of events in the reference dataset.

After processing the monthly seismic data for three years, I can get a plot of the average location error of **PLRR** results to manual results against the increase of events in the reference dataset (figure 3-9). As the mislocated seismic events of the previous month are absorbed by the reference dataset, the mislocation error of the next month will decrease according to the trend in figure 3-9. However, this will not apply when the waveforms of the new events do not

match any in the reference data set. Several point around 150 in the x-axis have higher location errors even through the reference dataset has more events. This happened because, in that months (e.g. October and November, 2003), a large earthquake swarm occurred in SMB, and there were many cases of double-events on one waveform (Figure 3-10), making it more difficult for the automatic method to find the correct location since it uses the whole waveform window to carry out cross-correlation.

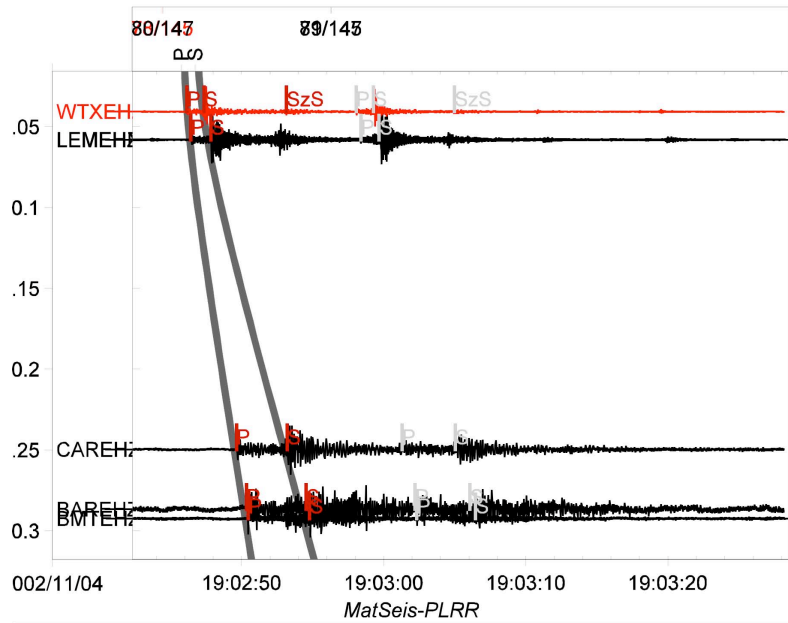


Figure 3-10. An example of double-event that occurred in the Socorro area in November, 2002. This type of earthquake together with ripple fires (several mining blasts in a short period) makes it difficult for **PLRR** to find the true location because matching reference events may not be available or appropriate.

## **EARTHQUAKE LOCATION IN NEW MEXICO (2000-2004)**

As discussed in chapter one, local seismic activity in New Mexico comprises natural earthquakes and man-made mining explorations concentrated in several source regions. These two types of seismic activity form the bulk of monthly seismicity of New Mexico. In this project, the PLRR program had been used to pick seismic phases and locate seismic events (2000-2004) in New Mexico automatically, and then we repick the phases and relocate the epicenters manually. As the products of this research, the distribution of seismic event in this period is demonstrated in this chapter.

In this chapter, the epicenter distribution of seismic events around the state of New Mexico will be presented every 4 months in a year from 1997 to 2004. The seismic event database from July 1997 to February 1998 was collected by Xiaobing Zhang ([8]) and repicked and relocated by Darren Hart.

In each of the following maps, the epicenters of earthquakes and mining blasts are plotted in different colors. The criteria we used to differentiate between mining blast from earthquakes were: (1) Mining blasts almost always occur during daytime working hours (8:00 am - 6:00 pm). (2) The epicenters of mining blasts are associated with known major mining fields.

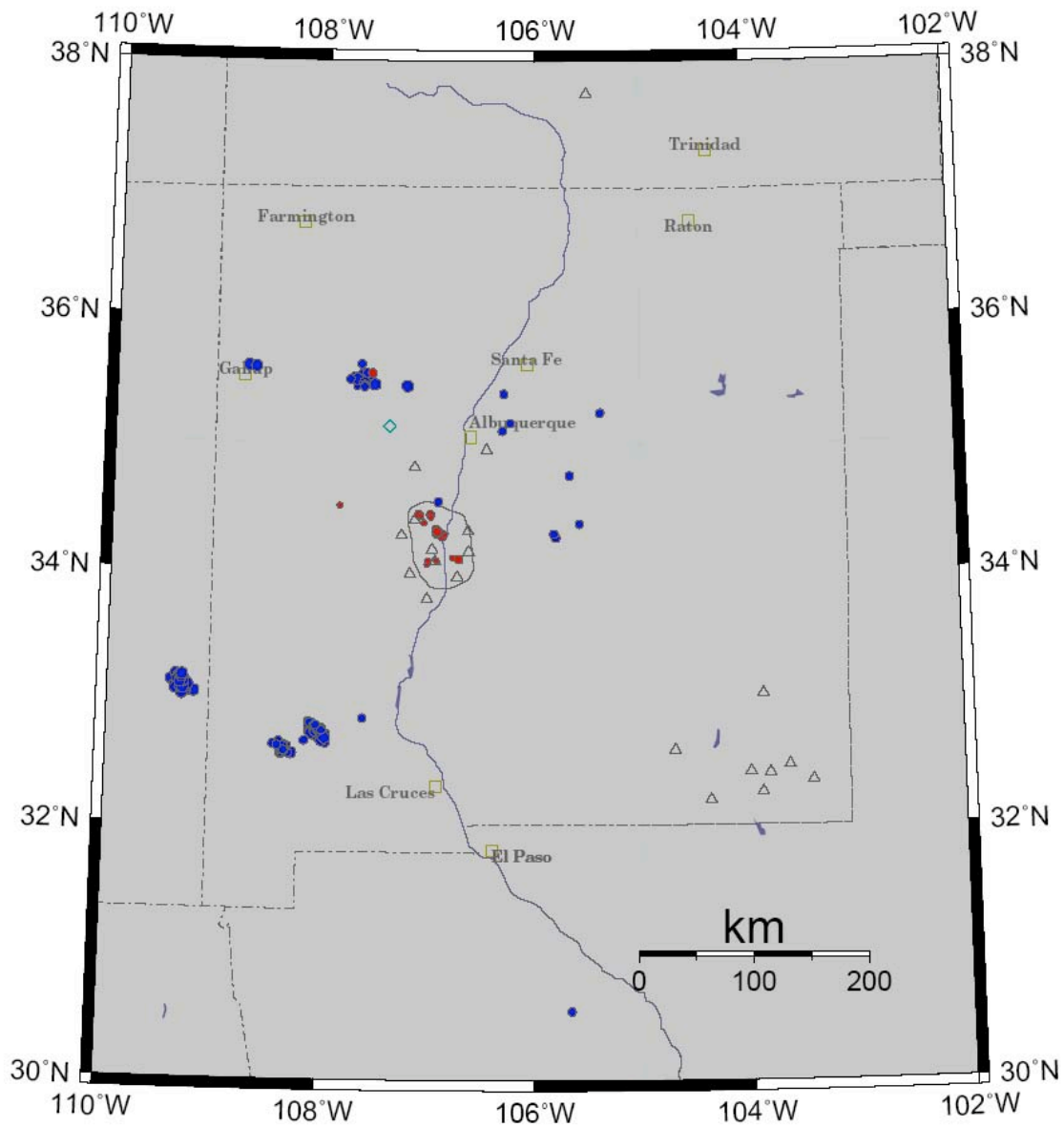


Figure 4-1. Epicenter distribution of seismic events around New Mexico (July, 1997 - August, 1997). Earthquakes are in red circles (dark circle in gray plot) and mining blasts are in blue circles (gray circles in gray plot). Triangles symbols are seismic stations. Diamond symbols are mining fields and square symbols are cities.

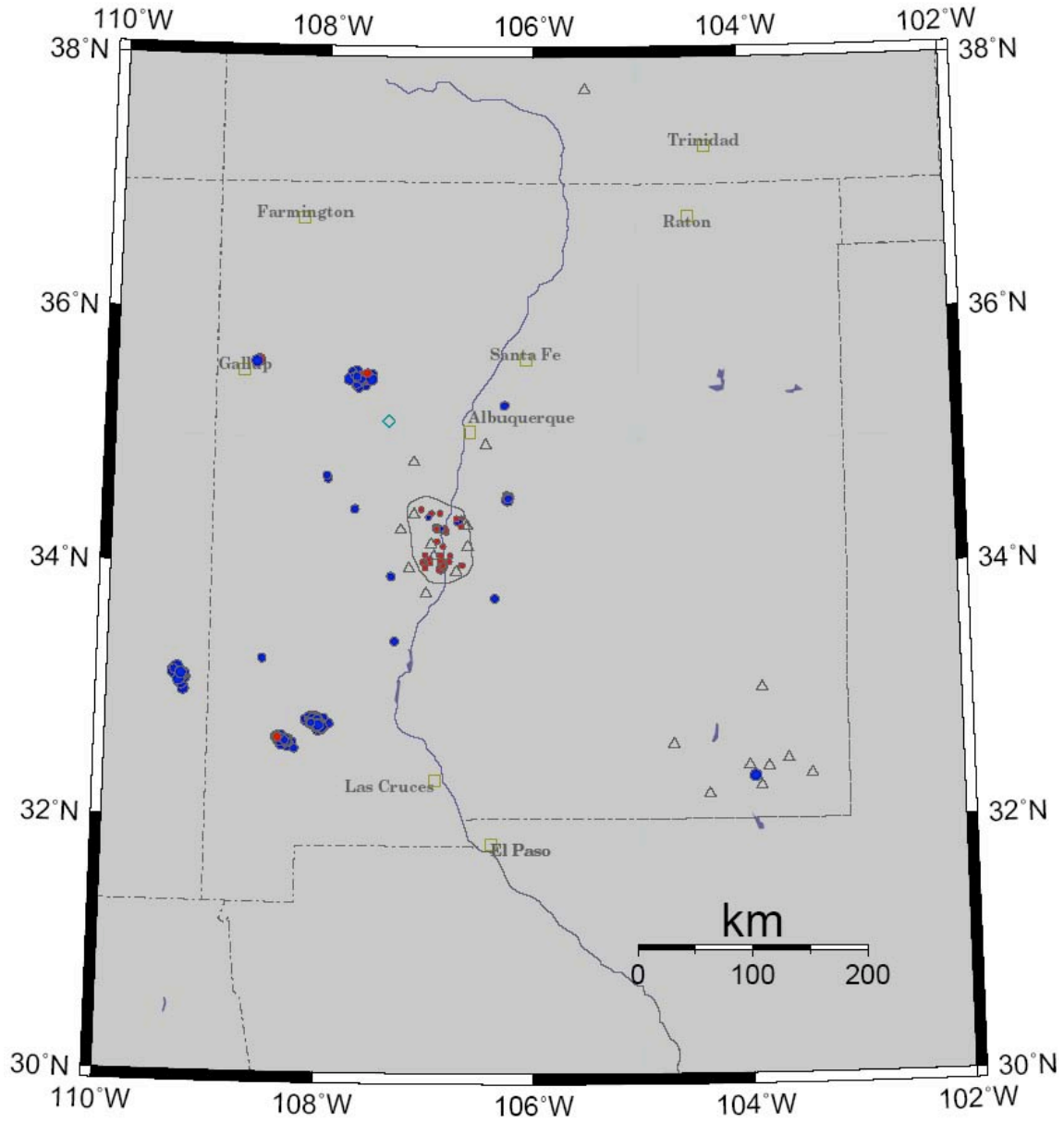


Figure 4-2. Epicenter distribution of seismic events around New Mexico (September, 1997 - December, 1997). Earthquakes are in red circles (dark circle in gray plot) and mining blasts are in blue circles (gray circles in gray plot). Triangles symbols are seismic stations. Diamond symbols are mining fields and square symbols are cities.



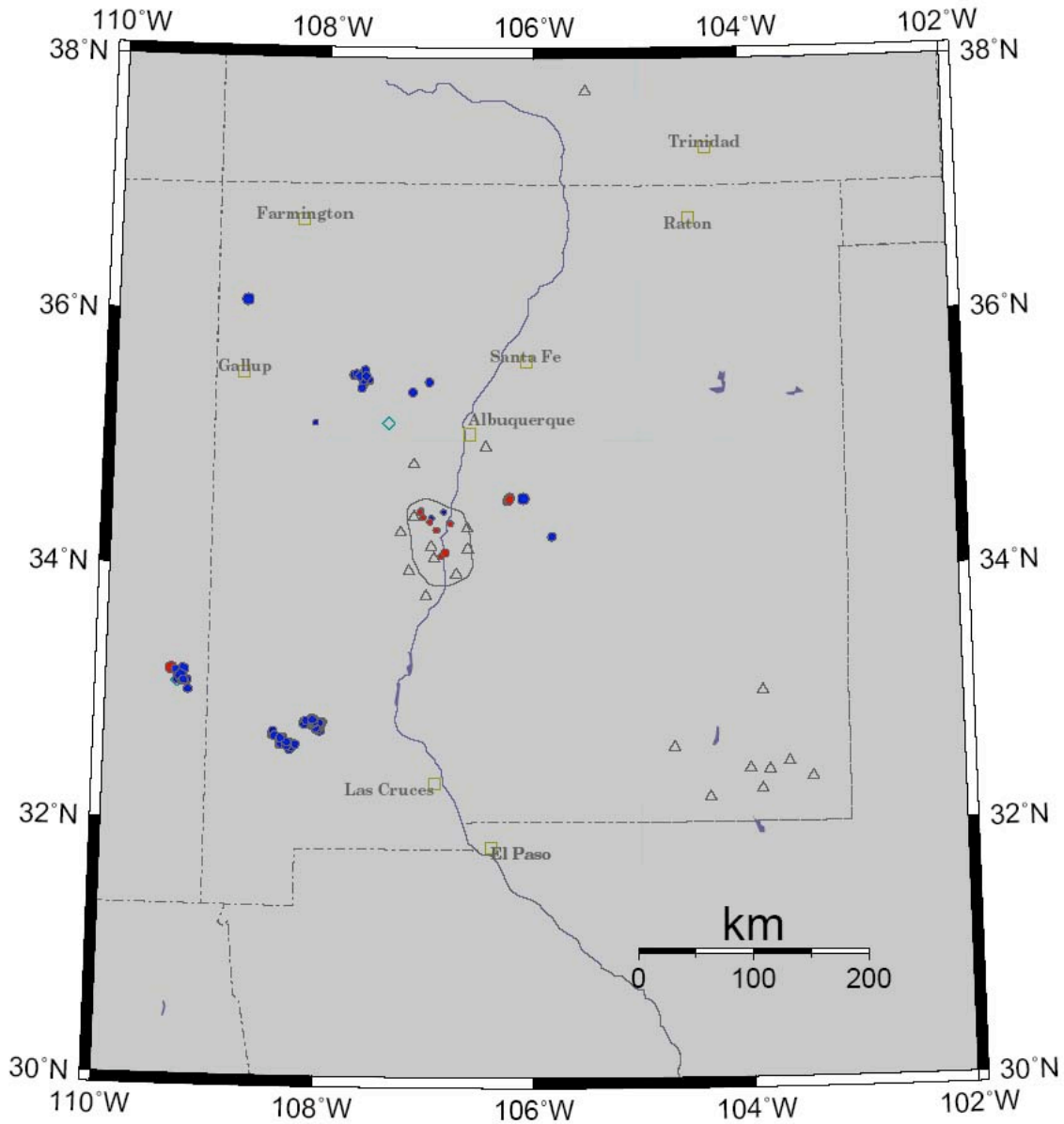


Figure 4-3. Epicenter distribution of seismic events around New Mexico (January, 1998 - February, 1998). Earthquakes are in red circles (dark circle in gray plot) and mining blasts are in blue circles (gray circles in gray plot). Triangles symbols are seismic stations. Diamond symbols are mining fields and square symbols are cities.

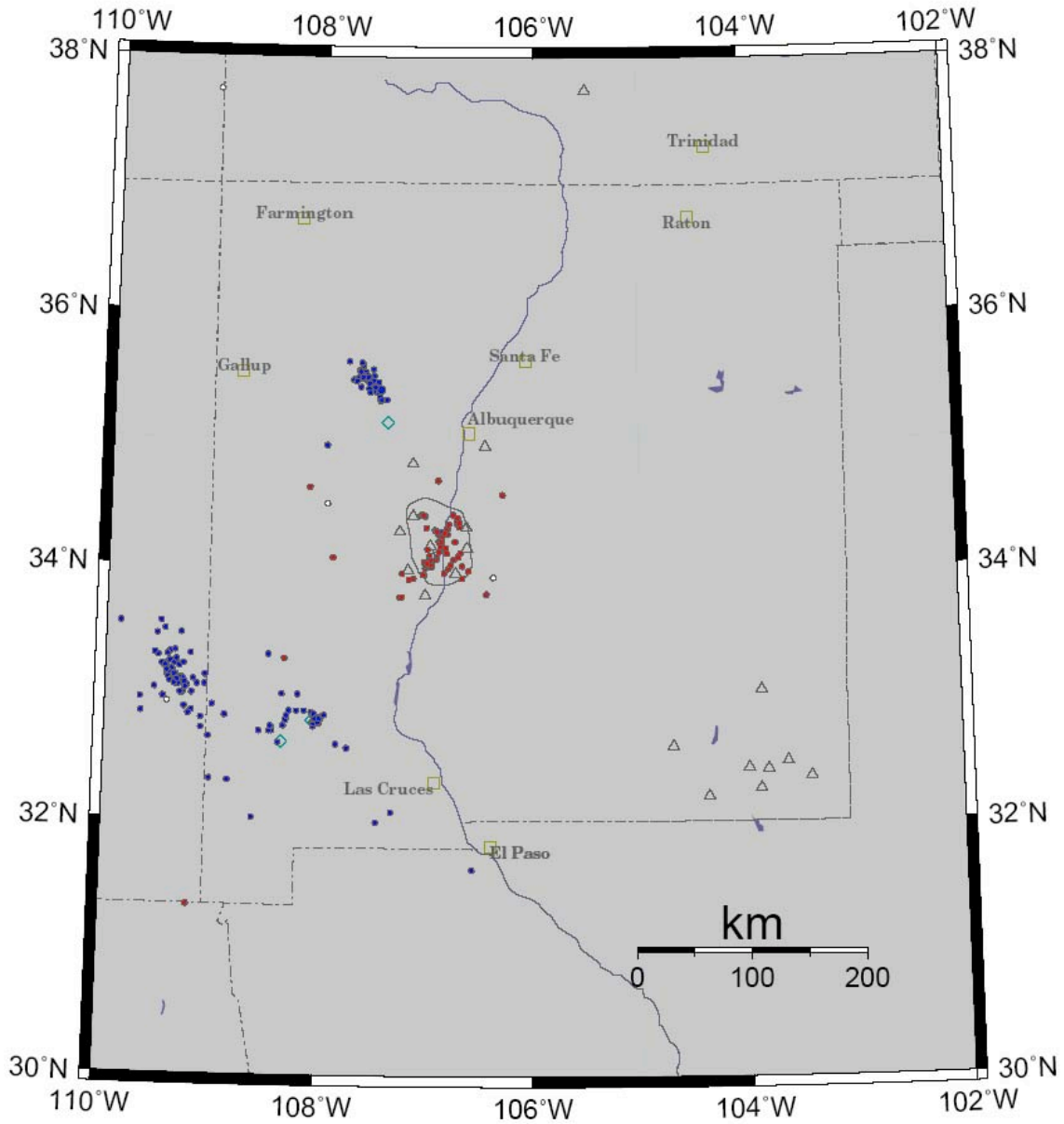


Figure 4-4. Epicenter distribution of seismic events around New Mexico (May, 2000 - August, 2000). Earthquakes are in red circles (dark circle in gray plot) and mining blasts are in blue circles (gray circles in gray plot). Triangles symbols are seismic stations. Diamond symbols are mining fields and square symbols are cities.

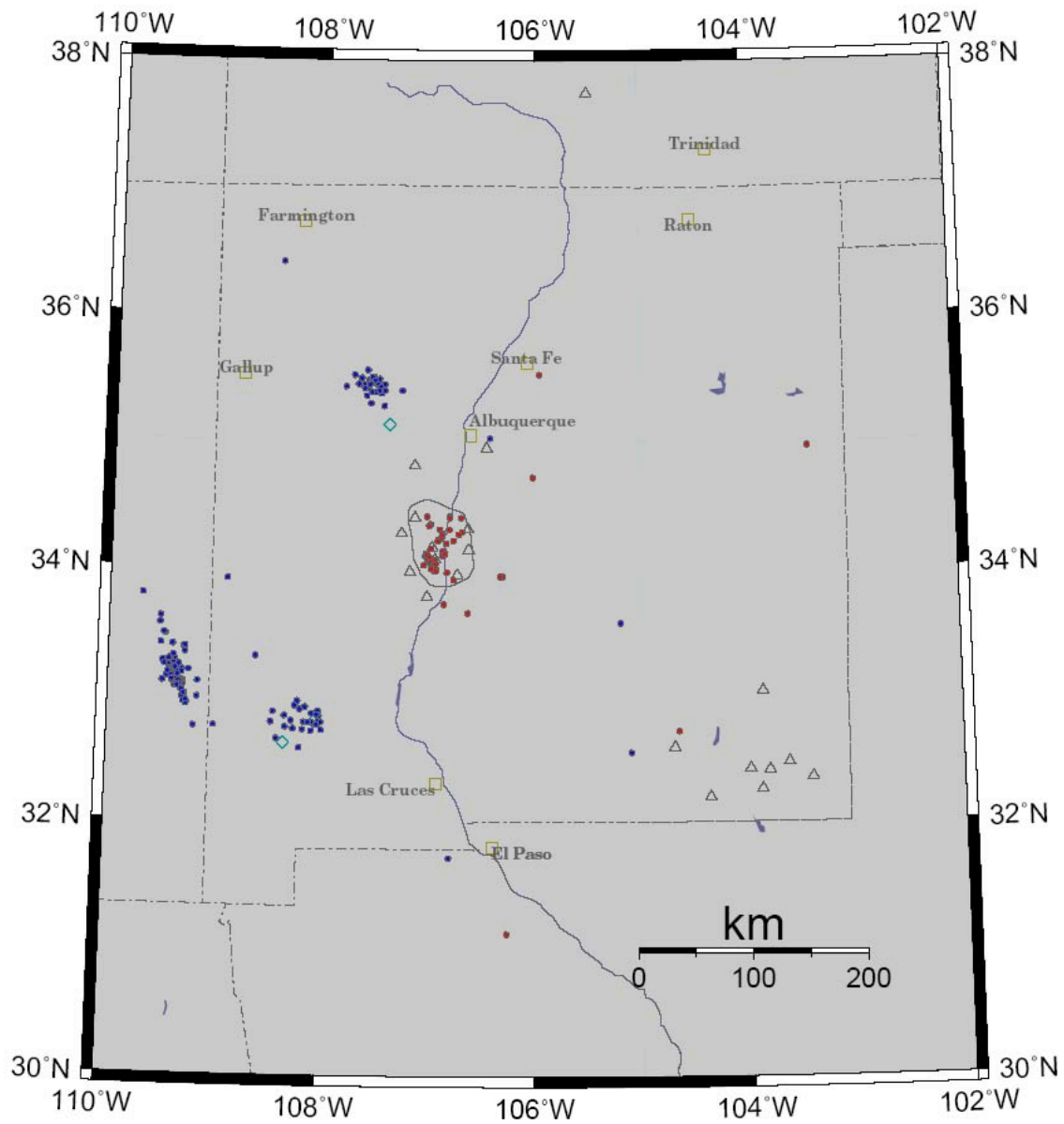


Figure 4-5. Epicenter distribution of seismic events around New Mexico (January, 2001 - April, 2001). Earthquakes are in red circles (dark circle in gray plot) and mining blasts are in blue circles (gray circles in gray plot). Triangles symbols are seismic stations. Diamond symbols are mining fields and square symbols are cities.

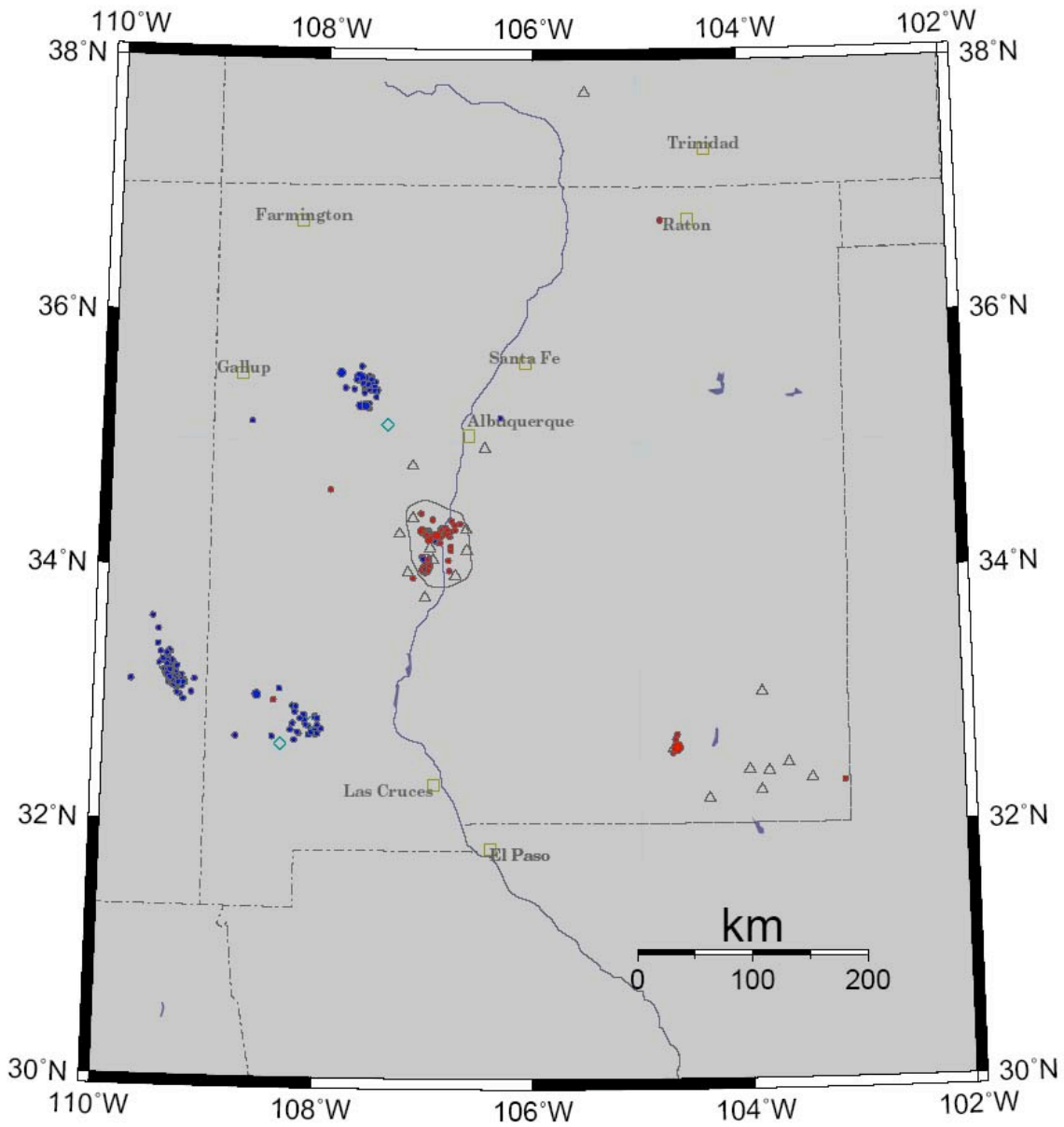


Figure 4-6. Epicenter distribution of seismic events around New Mexico (May, 2001 - August, 2001). Earthquakes are in red circles (dark circle in gray plot) and mining blasts are in blue circles (gray circles in gray plot). Triangles symbols are seismic stations. Diamond symbols are mining fields and square symbols are cities.

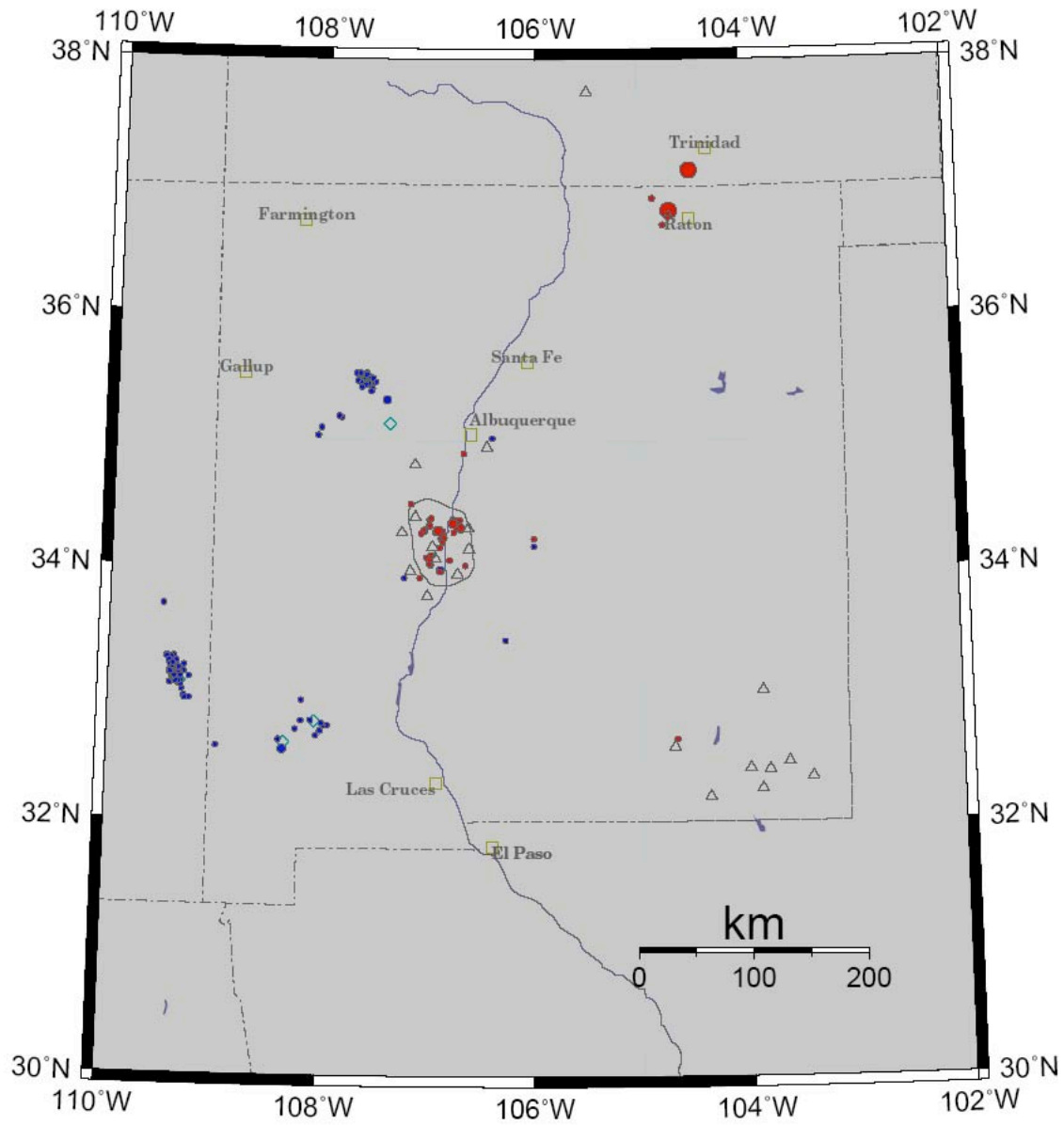


Figure 4-7. Epicenter distribution of seismic events around New Mexico (September, 2001 - December, 2001). Earthquakes are in red circles (dark circle in gray plot) and mining blasts are in blue circles (gray circles in gray plot). Triangles symbols are seismic stations. Diamond symbols are mining fields and square symbols are cities.

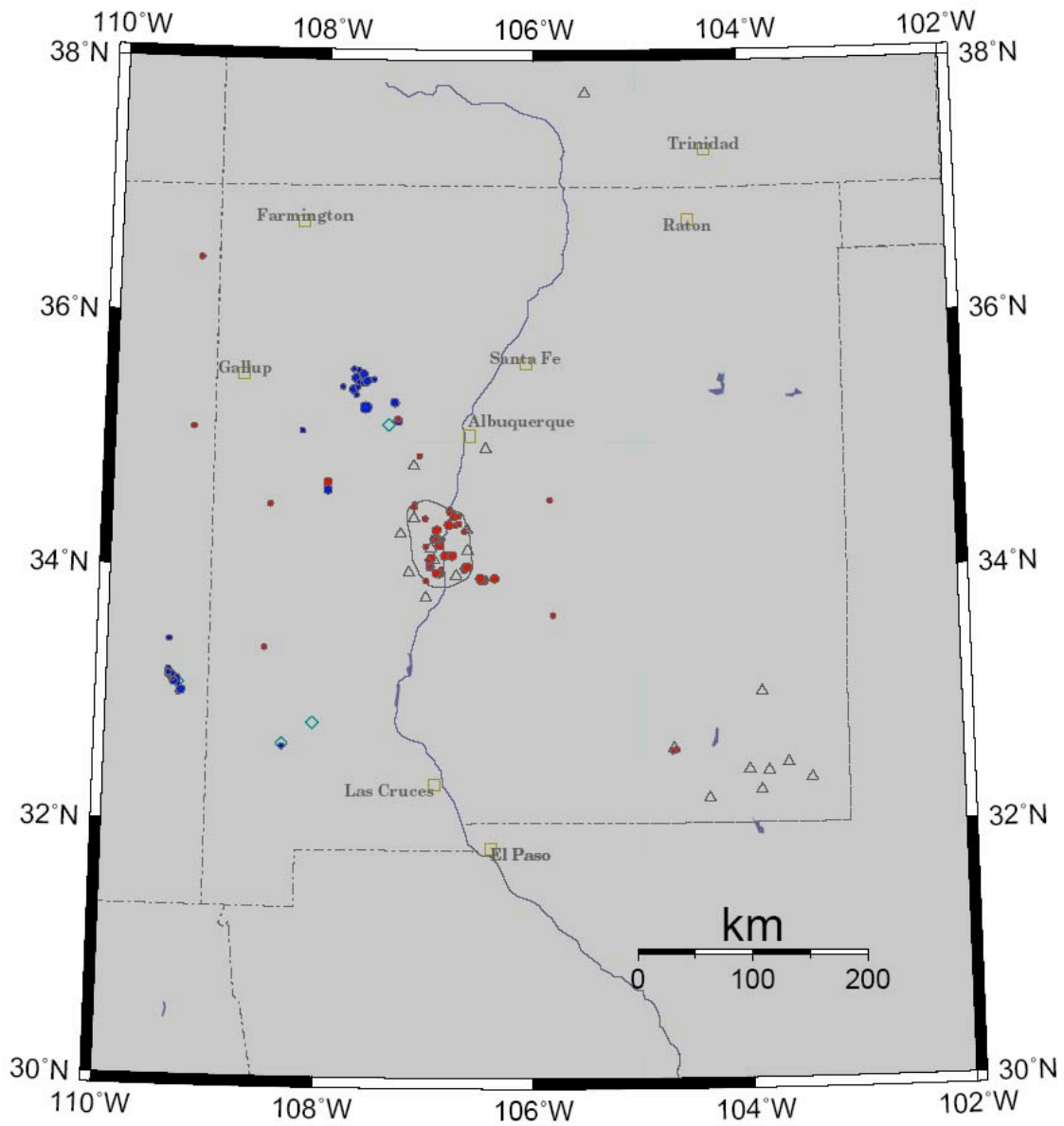


Figure 4-8. Epicenter distribution of seismic events around New Mexico (January, 2002 - April, 2002). Earthquakes are in red circles (dark circle in gray plot) and mining blasts are in blue circles (gray circles in gray plot). Triangles symbols are seismic stations. Diamond symbols are mining fields and square symbols are cities.

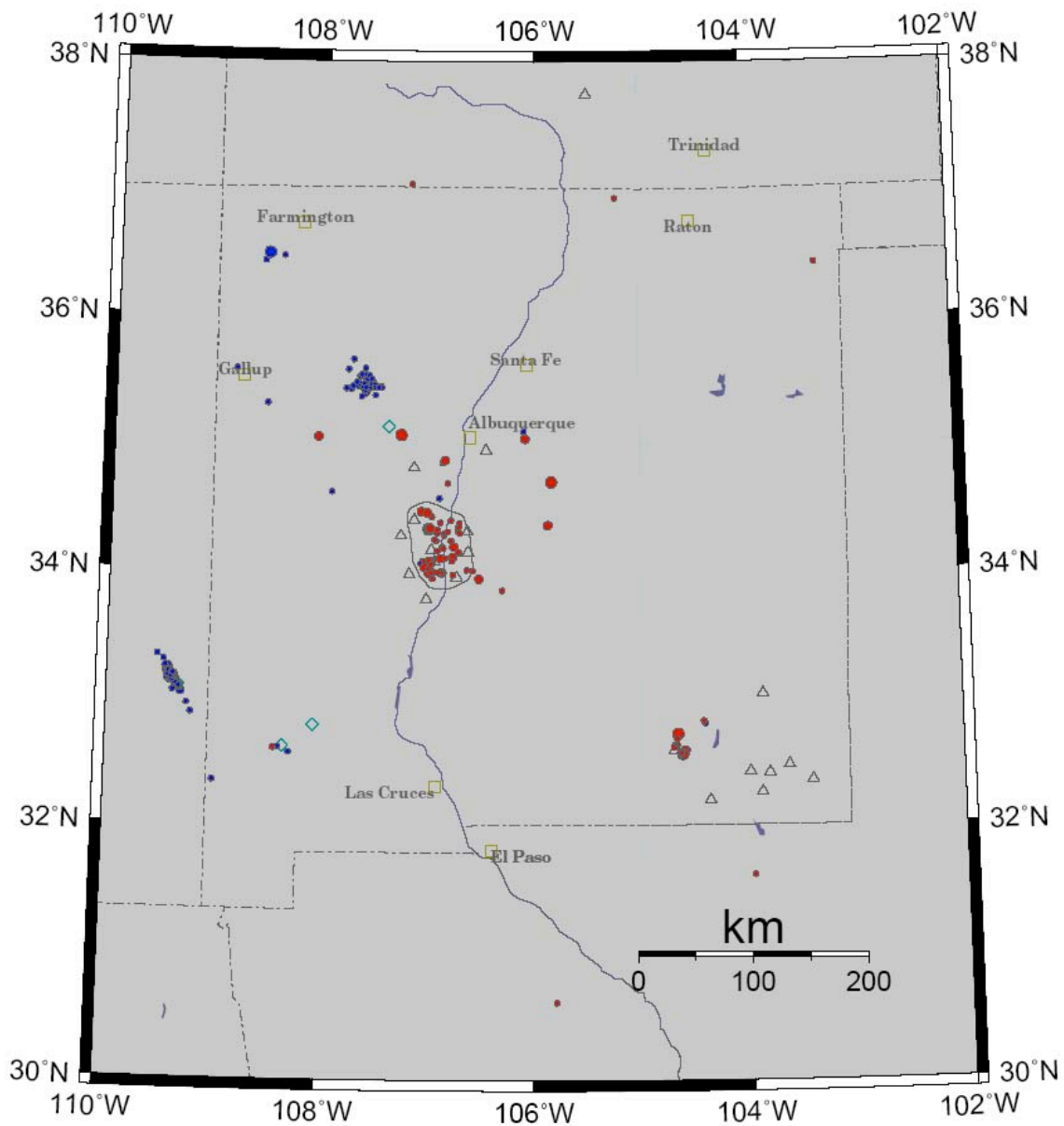


Figure 4-9. Epicenter distribution of seismic events around New Mexico (May, 2002 - August, 2002). Earthquakes are in red circles (dark circle in gray plot) and mining blasts are in blue circles (gray circles in gray plot). Triangles symbols are seismic stations. Diamond symbols are mining fields and square symbols are cities.

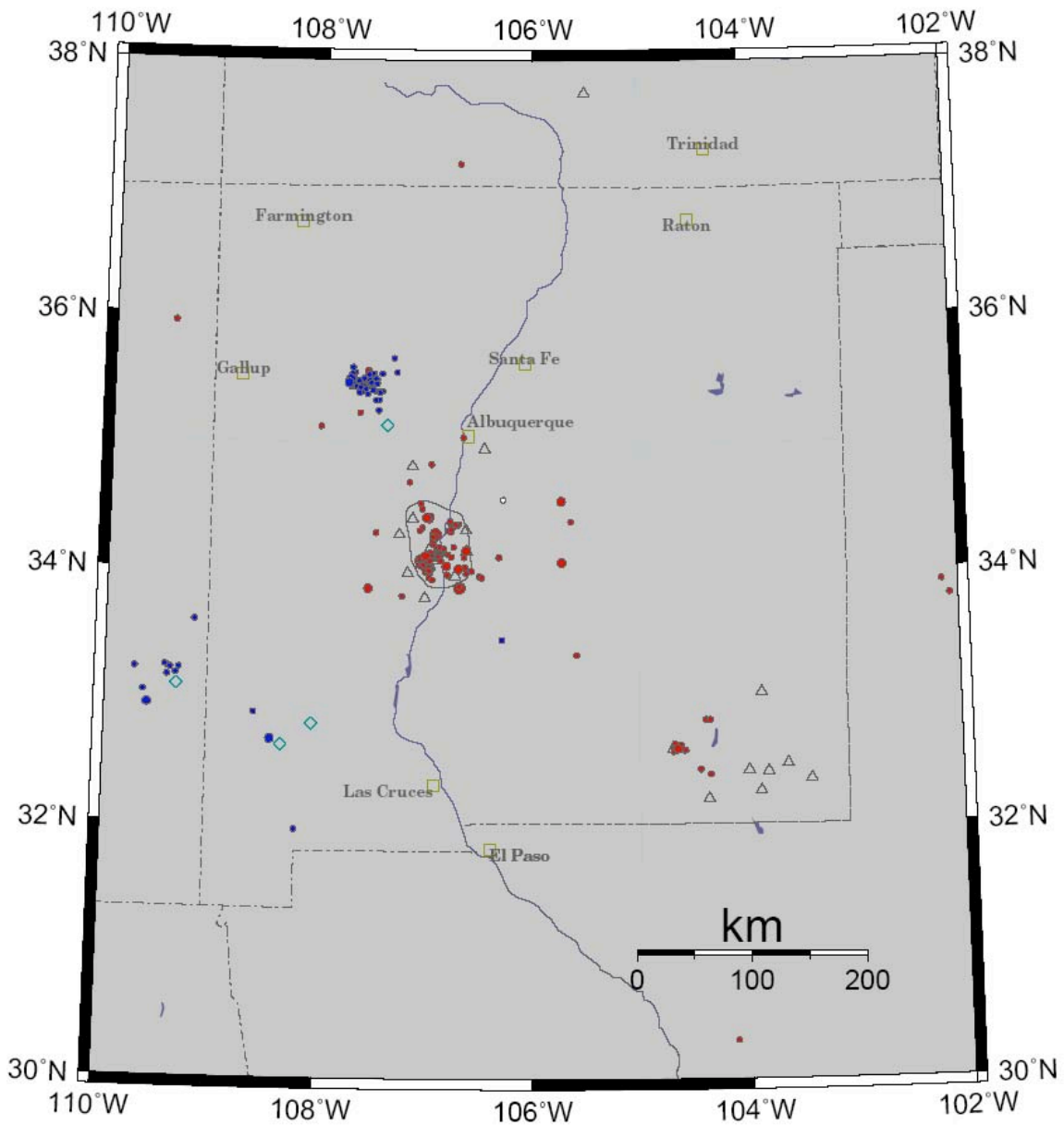


Figure 4-10. Epicenter distribution of seismic events around New Mexico (September, 2002 - December, 2002). Earthquakes are in red circles (dark circle in gray plot) and mining blasts are in blue circles (gray circles in gray plot). Triangles symbols are seismic stations. Diamond symbols are mining fields and square symbols are cities.



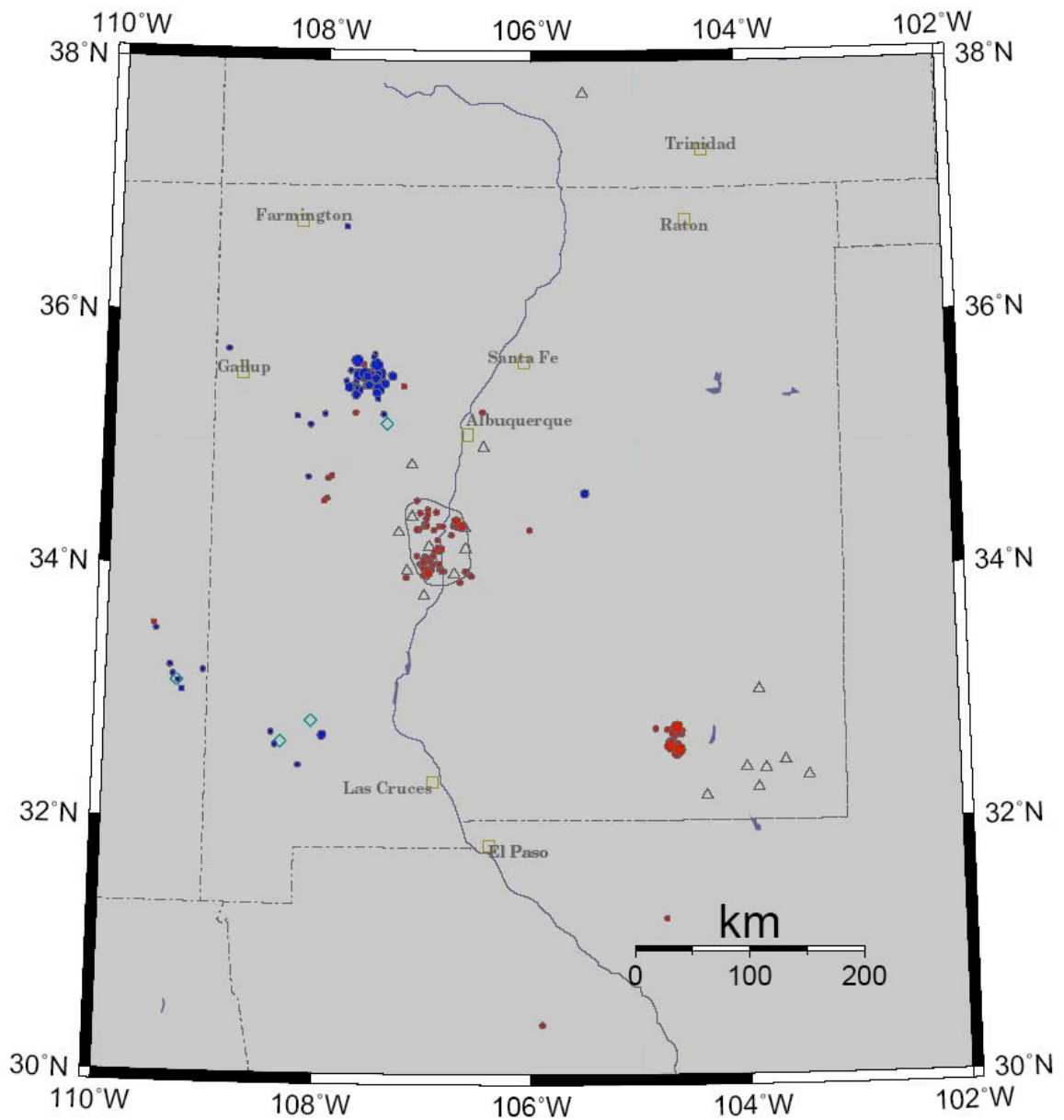


Figure 4-11. Epicenter distribution of seismic events around New Mexico (January, 2003 - April, 2003). Earthquakes are in red circles (dark circle in gray plot) and mining blasts are in blue circles (gray circles in gray plot). Triangles symbols are seismic stations. Diamond symbols are mining fields and square symbols are cities.

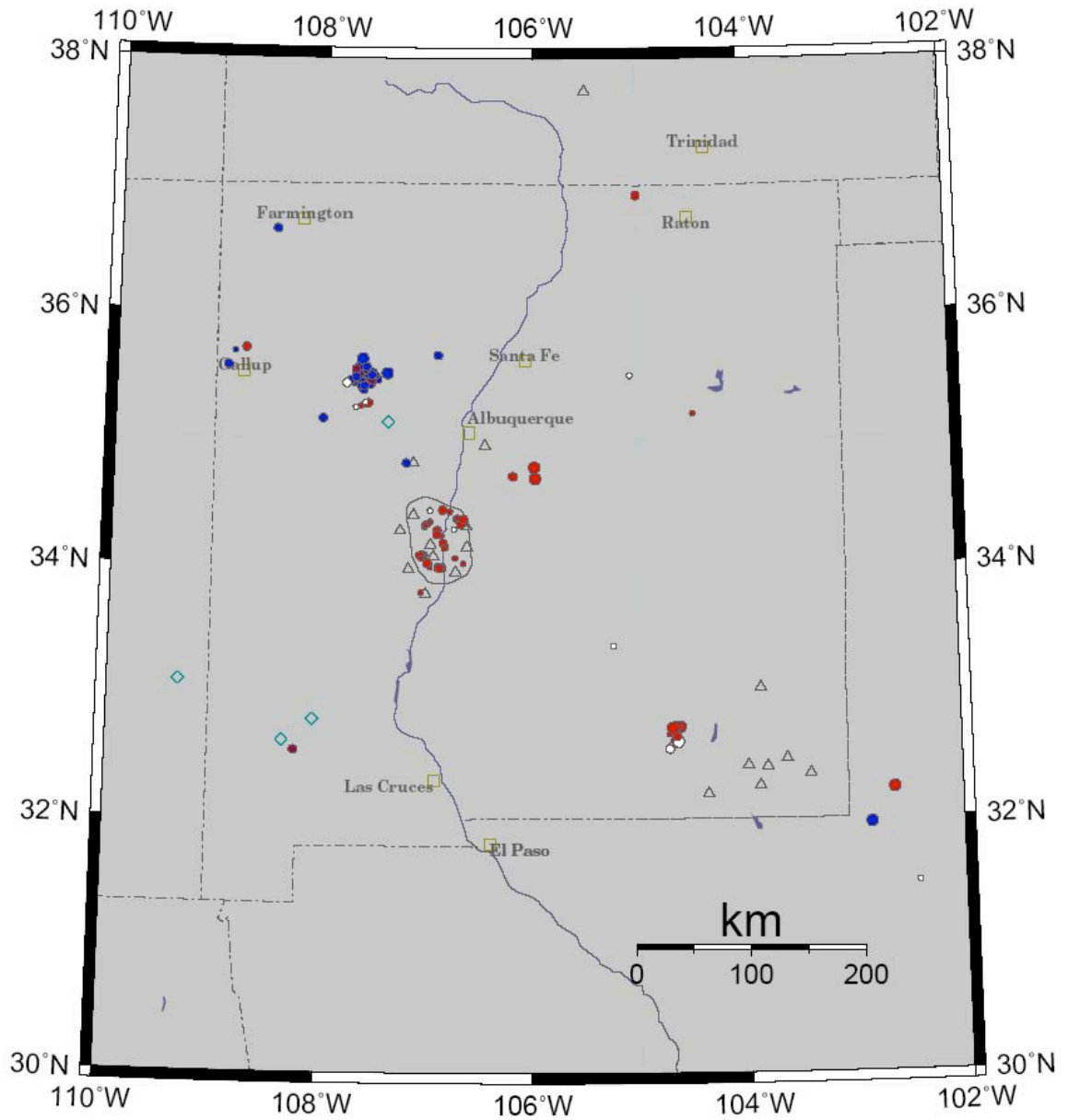


Figure 4-12. Epicenter distribution of seismic events around New Mexico (May, 2003 - August, 2003). Earthquakes are in red circles (dark circle in gray plot) and mining blasts are in blue circles (gray circles in gray plot). Triangles symbols are seismic stations. Diamond symbols are mining fields and square symbols are cities.

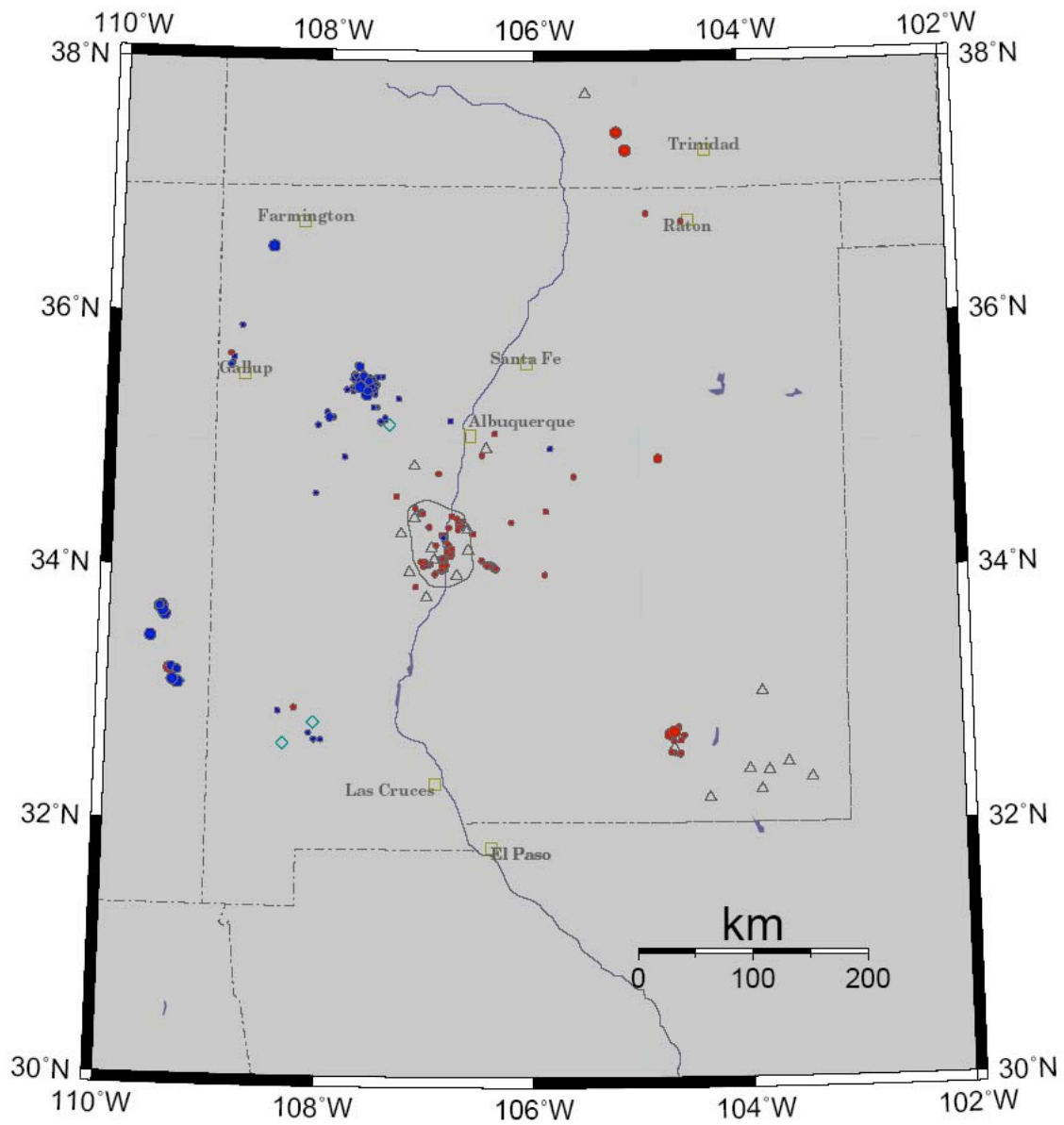


Figure 4-13. Epicenter distribution of seismic events around New Mexico (September, 2003 - December, 2003). Earthquakes are in red circles (dark circle in gray plot) and mining blasts are in blue circles (gray circles in gray plot). Triangles symbols are seismic stations. Diamond symbols are mining fields and square symbols are cities.

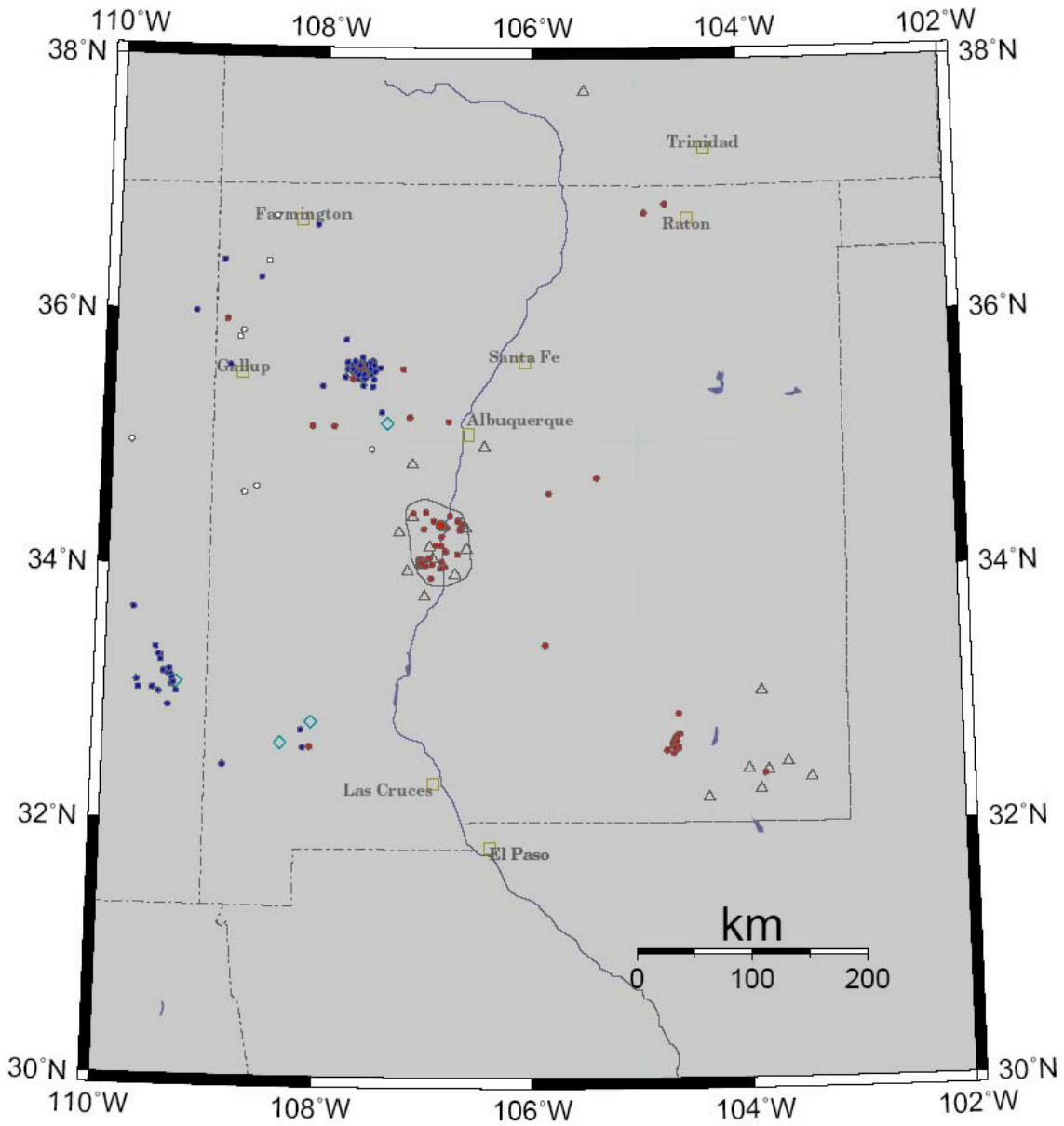


Figure 4-14. Epicenter distribution of seismic events around New Mexico (January, 2004 - April, 2004). Earthquakes are in red circles (dark circle in gray plot) and mining blasts are in blue circles (gray circles in gray plot). Triangles symbols are seismic stations. Diamond symbols are mining fields and square symbols are cities.

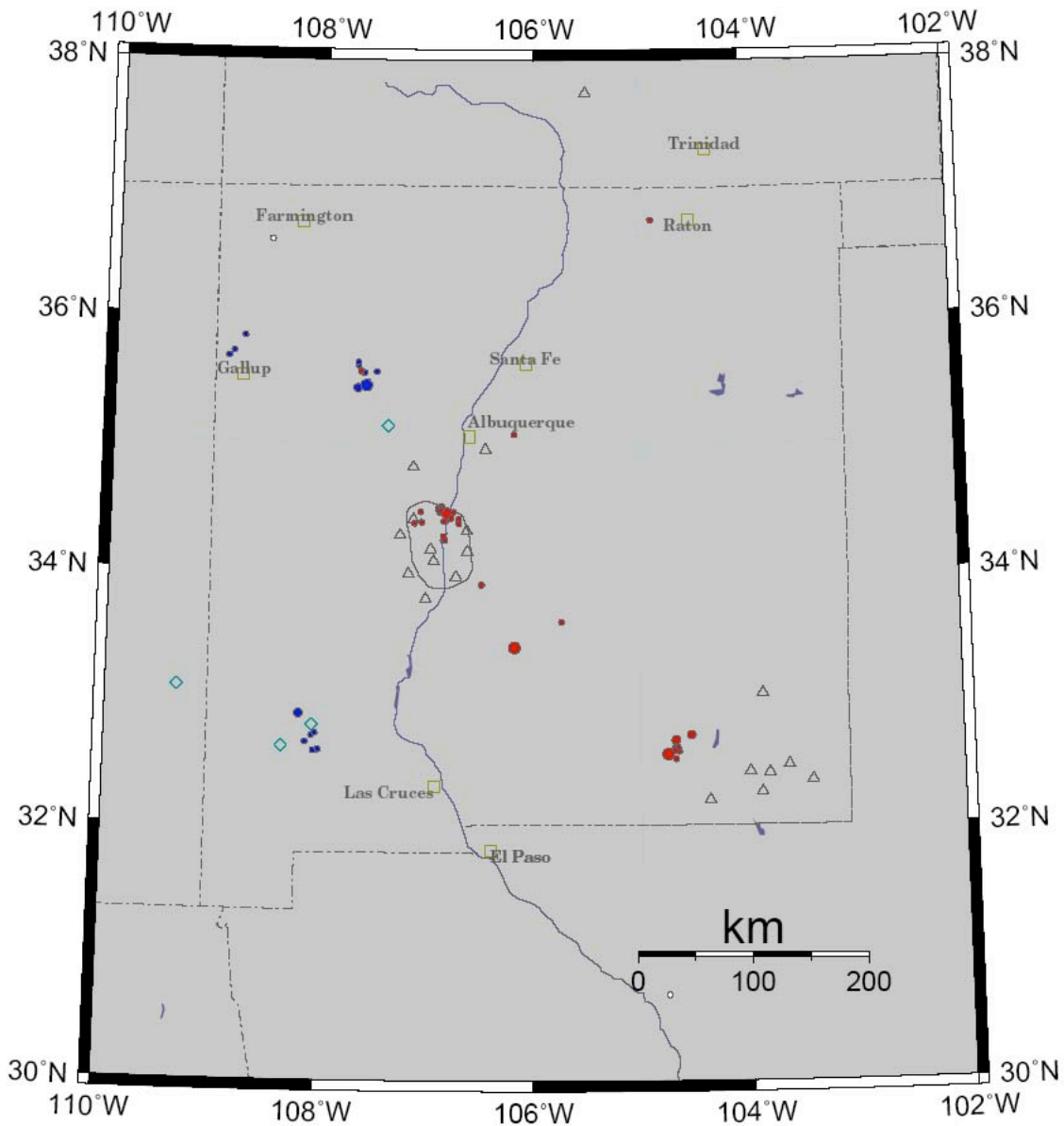


Figure 4-15. Epicenter distribution of seismic events around New Mexico (May, 2004 - June, 2004). Earthquakes are in red circles (dark circle in gray plot) and mining blasts are in blue circles (gray circles in gray plot). Triangles symbols are seismic stations. Diamond symbols are mining fields and square symbols are cities

## **SUMMARY, SUGGESTIONS AND CONCLUSIONS**

### **Summary**

The intention of this independent study was to find a way to automatically process seismic data for New Mexico (to save waveform data with triggered signals, to identify seismic events, to pick seismic phases and to find their locations) and use this method to locate seismic events in New Mexico from 2000 to 2004. The Earthworm provides a front-end for the processing of seismic data for the New Mexico network and produces automatic triggered events based on the STA/LTA algorithm. However, less than half of the triggers are typically seismic events because of various sources of transient noise. By considering the seismic characteristics of New Mexico seismograms (source regions are somewhat independent of each other, and waveforms of events occurring in the same source region are relatively similar), we developed an automatic program (**PLRR**) to detect seismic events and find seismic phases and locations. **PLRR** uses a reference event-waveform database that includes seismic events from all possible source regions of New Mexico to compare an unknown event with each event in the database with the waveform cross-correlation technique, and associate new events with similar past events prior to manual review.

The **PLRR** program is a package of **Matlab** m-files and depends on the support of **Matseis**. The only acceptable seismic data format for **PLRR** is the CSS3.0 format. A perl script was written to convert the daily triggered events of Earthworm from SAC format to CSS3.0 flatfile format. At the beginning of each new month, analysts can run **PLRR** to process the daily CSS3.0 data of the previous month and use it to detect seismic events and pick phases

automatically. There are two cross-correlation methods available with **PLRR**. One is the *waveform-pair cross-correlation matching* method, and another is the *event-pair cross-correlation matching* method. The latter one was found to be more robust than the former. Due to the layout of Socorro network, seismic events located about 200 km away from Socorro are often mistakenly identified as occurring in Socorro area by the *event-pair cross-correlation matching* method. This is because the amplitudes of S-phases in these events decrease to the background level when their waveforms arrive at the seismic stations from 200 km away. But in the frequency domain, the amplitudes of their S-phases are very strong in the band of 6-10 Hz. We developed the *spectrogram event-pair cross-correlation matching* method to increase signal-to-noise ratios and thereby avoid these types of gross location errors.

Using **PLRR**, we processed the seismic data of New Mexico from 2000 to 2004 automatically, then went back and repicked seismic phases and relocated seismic events manually. All in all, 3,023 events around New Mexico were phase-picked and located, most located in the SMB and major mining fields. Some events from the SMB had clearly defined magma body reflection phases, enabling us to find the depths of these events more precisely. During the data procesings, seismic events occurring in new source regions were added in the reference database so that the phase-picking and epicenter-locating ability of **PLRR** could adapt to future occurrences of the new waveforms.

### **Suggestions**

During the last two and a half years' development, I made mistakes along with my progress. Although I got everything done, this work is far from perfect. Here are some suggestions which might be good for future development:

1. Whenever a new type of seismic event is found, it is supposed to be added to the reference set. When the number of reference events becomes very large, the processing time to find the phases and location for one event becomes intolerable. A cluster algorithm (e.g. dendrogram tree) might be needed to organize the reference set, and a new event will be compared with a sub group of events or a master event list instead of whole events in the reference set. This will decrease the calculation time dramatically.
2. Most events in the reference set do not have waveforms for all seismic stations around New Mexico. It would increase the event-detection ability if some synthetic waveforms were added to stations missing waveforms in the reference set, or if new reference events with all stations were used to replace them.
3. **PLRR** could have new future capabilities. The phase repicking program (developed by C. Rowe and R. Aster Rowe's Ph.D. work) could to be incorporated into it. A double-difference relocation program (e.g. hypoDD or tomoDD) also could be integrated. In the long run, **PLRR** could be implanted into a real-time system to detect seismic signals and produce very high resolution seismic event catalogues automatically.
4. Some seismic stations need to be set up in the three corners of New Mexico besides Carlsbad to get better seismic event coverage, thereby increasing the accuracy and effectiveness of **PLRR** for the study region.

### Conclusions

In this independent study, I have presented an adaptive program that automatically detects seismic phases and locates epicenters by using the waveform cross-correlation technique. I



have shown how to combine it with the **Earthworm** system to process seismic data in New Mexico automatically. Two waveform cross-correlation matching methods were presented and the *event-pair cross-correlation matching* method was found to be better than the *waveform-pair cross-correlation matching* method. *Event-pair cross-correlation matching* method in the frequency domain can prevent the seismic events far away from the seismic network from being misidentified as events occurring locally. I also presented the seismic event location results in New Mexico from 2000 to 2004. This study will be helpful for people in the future because it may lead to a system that will vastly decrease the amount of time researchers spend on manual data processing.

## REFERENCE

- [1] <http://www.ees.nmt.edu/Geop/Museum Posters/NMseismology.html>
- [2] J. Schlue and Jaksha L., A Report On the SEISMICITY OF THE WIPP SITE for the period 1 July 2003 through 20 September 2003, 2003.
- [3] A. R. Sanford, K. Lin, I. Tsai and Jaksha L. H., Earthquake catalogs for New Mexico and bordering areas: 1869-1998, New Mexico Bureau of Geology Mineral Researches, Circular 210, 2002.
- [4] R. S. Balch, H. E. Hartse, A. R. Sanford and Lin K., A New Map of the Geographic Extent of the Socorro Mid-Crustal Magma Body, *Bull. Seism Soc. Am.*, Vol. 87, 174-182, 1997.
- [5] Sanford, A. R. and C. R. Holmes, Note on the July 1960 earthquakes in central New Mexico, *Bull. Seism. Soc. Am.* 51, 311-314, 1961.
- [6] Hans E. Hartse, Simultaneous hypocenter and velocity model estimation using direct and reflected phases from microearthquakes recorded within the center Rio Grande rift, New Mexico, Ph. D. Dissertation, New Mexico Institute of Mining and Technology, Socorro, New Mexico, 1991.
- [7] H. E. Hartse, Sanford A. R. and Knapp J. S., Incorporating Socorro Magma Body reflections into the earthquake location process, *Bull. Seism Soc. Am.*, 82, 2511-2532, 1992.
- [8] Xiaobing Zhang, Processing of Seismic Events collected by the New Mexico Tech Seismic Network Using a Matseis-based Review Tool and a Postgres Database, *Master degree thesis*, New Mexico Institute of Mining and Technology, 1997.
- [9] Withers Mitchell, An Automated Local/Regional Seismic Event Detection And Location System Using Waveform Correlation, *Ph. D. Dissertation*, New Mexico Institute of Mining and Technology, Socorro, New Mexico, 1997.
- [10] C. J. Young, B. J. Merchant and Aster R. C., Comparison of cluster analysis methods for identifying seismic events, 23<sup>rd</sup> Annual DTRA/NNSA Seismic Research Review, 229-238, 2001.
- [11] C. A. Rowe, R. C. Aster, B. Borchers and C. J. Young, An automatic, adaptive algorithm for refining phase picks in large seismic data sets, *Bull. Seism Soc. Am.*, 92, 1660-1674, 2002.
- [12] Sanford, A. R. and P. Einarsson, Magma chambers in rifts, *Continental of Ocean Rifts Geodynamics Series*, Volume 8, 147-168, 1982.
- [13] Papanicolaou, Lecture on "Mathematical Geophysics and Uncertainty in Earth Models" summer school, Golden, Colorado, June 14-25 2004.
- [14] Song, X.D., P.G. Richards, Observational evidence for differential rotation of the Earth's inner core, *Nature*, 382, 221-224, 1996.

[15] Aster, R.C., P.M. Shearer, and J.C. Berger, Reply to Comment on "Quantitative measurements of shear wave polarizations at the Anza Seismic Network, Southern California: implications for shear wave splitting and earthquake prediction", *Journal of Geophysical Research*, 96, 6415-6419, 1991.

[16] Bergthora S. T., and J. C. Pechmann, Constraints of relative earthquake locations from cross-correlation of waveforms, *Bulletin of the Seismological Society of America*, 77, 1626-1634, 1987.

[17] Young, C. J., B. J. Merchant and Aster R. C., Comparison of cluster analysis methods for identifying regional seismic events, 23<sup>rd</sup> Annual D'TRA/NNSA seismic Research Review, 229-238, 2001

[18] Homepage of Matseis (<https://www.nemre.nnsa.doe.gov/cgi-bin/prod/nemre/matseis.cgi>)

[19] Webpage of Earthworm (<http://folkworm.ceri.memphis.edu/ew-doc/modules.html>)

[20] L. Carmel, Y. Koren and D. Harel, Visualizing and Classifying Odors Using a Similarity Matrix, *Proceedings of the 9th International Symposium on Olfaction and Electronic Nose (ISOEN'02)*, Aracne, pp. 141--146, 2003.

THE USER'S MANUAL OF PLRR-MATSEIS

MatSeis-PLRR 0.1

User's Manual

Waveform Cross-Correlation (WCC)

Phase Picking based on

Analyst Reviewed Reference Data Set

Wenzheng Yang, Rick Aster and Darren Hart

February 16, 2004

## Objective

The main goal of this project was to develop and test an automated method for phase picking based on the new events' similarity to events archived in a reference database. The measurement of similarity used to compare events is waveform cross correlation (WCC). The typical processing is envisioned as follows: 1) Typical seismic activity (manmade and natural) for a region is cataloged in a reference database, 2) A new event database is compiled based on a real-time triggering process (e.g. Earthworm is used at New Mexico Tech), 3) WCC is used to determine if new event triggers are found in the reference database. If the correlation between the reference waveform and the new event waveforms is above a set threshold, then phase picking is done. When four or more phases are picked for a new event, then the new event can be run through a location program to determine its hypocentral location.

## Introduction

MatSeis-PLRR, or PLRR, is a modified version of MatSeis v1.7 with some special functions. Currently the PLRR package could not be run without the underlining support of MatSeis v1.7. PLRR was developed and tested on the Sun Unix system. Before using PLRR, users should be sure that they have already installed Matlab v6.0 R12 or v6.5 R13 and MatSeis v1.7 on their system. Currently, PLRR only supports one database structure: CSS3.0 Flatfile format, which one of the database formats supported by MatSeis. To find out more about the CSS3.0 flatfile schema, download a document from the MatSeis website (<https://www.nemre.nnsa.doe.gov/cgi-bin/prod/nemre/matseis.cgi>).

**PLRR** represents a four-step process: **Pick-Locate-Repick-Relocate**. Pick represents the initial phase picking process, or phase identification problem. The second step, Locate, will assign the event as originating from a region. At this stage in development, we only implement the pick and locate parts within this package. Our future goal is to integrate the automatic initial phase picking and locating with phase repicking (Rowe et. al. 2000) and relocating (Haijiang and Thurber et. al. 2003) for groups of similar events.

The phase pick processing is called automatic waveform cross-correlation phase picking method because it is based on a reference set. Before running this process, a user will have to compile a reference database that contains typical seismic data for their special research region. The reference database should include events with high signal-noise-ratio waveforms and good phase picks (e.g. P and S or any particular characteristic phases). The reference database should be analyst reviewed and contain events that are most typical for that region. In the New Mexico area, the three types of activity of interest are earthquake swarms, recurrent earthquakes, and mining blasts. Typical mining activity, up 400 events annually, is recorded coming from the Morenci, Tyrone and Santa Rita open pit mines to the southwest of the New Mexico Tech Seismic Network (NMTSN). Natural seismic events are less likely to occur in the same relative location unless their part of a long duration earthquake swarm event or a recurrent seismic source. Because of this, our main focus is to recognize and locate the mining, swarm, and recurrent events.

However, an adaptive tool is envisioned which could be used to insert short duration seismic activity into the reference database or a subset-reference database.

There are two automatic picking methods available in PLRR besides the manual picking method provided in Matseis (i.e. Measure-Tool). The first automatic picking method we call the waveform matching method. For each waveform available for a new event, this method finds the most similar waveform from the same station in the reference database set and assigns all its phase pick information to the new event waveform. If more than 4 phases are picked by this method, then the Matseis locator tool, **mslocator\_tool**, is used to locate the new event.

The second method for automatic picking method we call the event matching method. For a new event, this method finds the most similar event to it from the reference set and assigns the corresponding phases and location of the reference event to the new event.

This user's manual provides the following information:

1. Installation
2. Start-up procedure of PLRR
3. PLRR main window and menus
4. How to construct a reference set.
5. How to use waveform matching method
6. How to use event matching method
7. The collection of an automatic picked event.
8. Example

We assume that users who want to use PLRR have some experience using MatSeis. New users should read through the MatSeis manual, as we will not explain any MatSeis features in this manual. A downloadable copy can be obtained from the MatSeis web page: <https://www>

## 1. Installation of Matseis-PLRR

Download the `matseis_plrr.tar.gz` file from Wenzheng's web site and save it to the desired location. To extract the files use:

```
gunzip matseis_plrr.tar.gz | tar -xvf
```

This will create three directories: `plrr`, `example` and `input`?? The important directory here is `plrr`. The `plrr` directory contains the core code necessary for running this package.

Three environmental variables will be set on startup or can be set prior to startup if needed.

## 2. Start-up of Matseis-PLRR

Before you run the package of **PLRR**, please make sure that Matlab and Matseis are already installed on the same Unix computer. `Plrr` is an executable c-shell script used to set environmental variables `MATSEIS_HOME`, `EXAMPLE`, `MS_CONFIG_FILE`, starts matlab and finally starts the **PLRR** package. The `MATSEIS_HOME`, `EXAMPLE`, and `MS_CONFIG_FILE` directory paths will have to be modified to match your system before attempting to use it. The following are the contents of `plrr` (lines starting with `#` are considered commented):

```
#!/bin/csh  
## MatSeis-PLRR startup script for unix users  
  
# set home directory of MatSeis startup executable  
setenv MATSEIS_HOME /bin/matseis-1.7  
  
# set example home directory location
```

```
setenv EXAMPLE /users/wyang/plrr_package/example
```

```
# config file from command line
```

```
setenv MS_CONFIG_FILE /users/wyang/plrr_package/example/example.config
```

```
# start matlab with no desktop, add the path to MATSEIS_HOME and finally start PLRR
```

```
matlab -nodesktop -r "addpath('$MATSEIS_HOME');plrr;"
```

Once the above changes have been made to the plrr file, check that it is found in the current search path by typing:

Which plrr

should produce a similar results as below, (on our system)

```
/user/wyang/plrr_package/
```

There are three database used by PLRR: input, reference and output. The directory locations of these databases are set from the main GUI window under the **File** pulldown menu. The contents inside \*.input file are nearly the same as the contents inside \*.input file used in **Matseis** except that we've added several lines to set the directory for reference set. **Matseis** only has the input and output directories setting.

After finishing setting the directories, you can launch the GUI window of **PLRR** by executing plrr script at the UNIX command prompt. Figure 1 shows the main GUI window of **PLRR**.



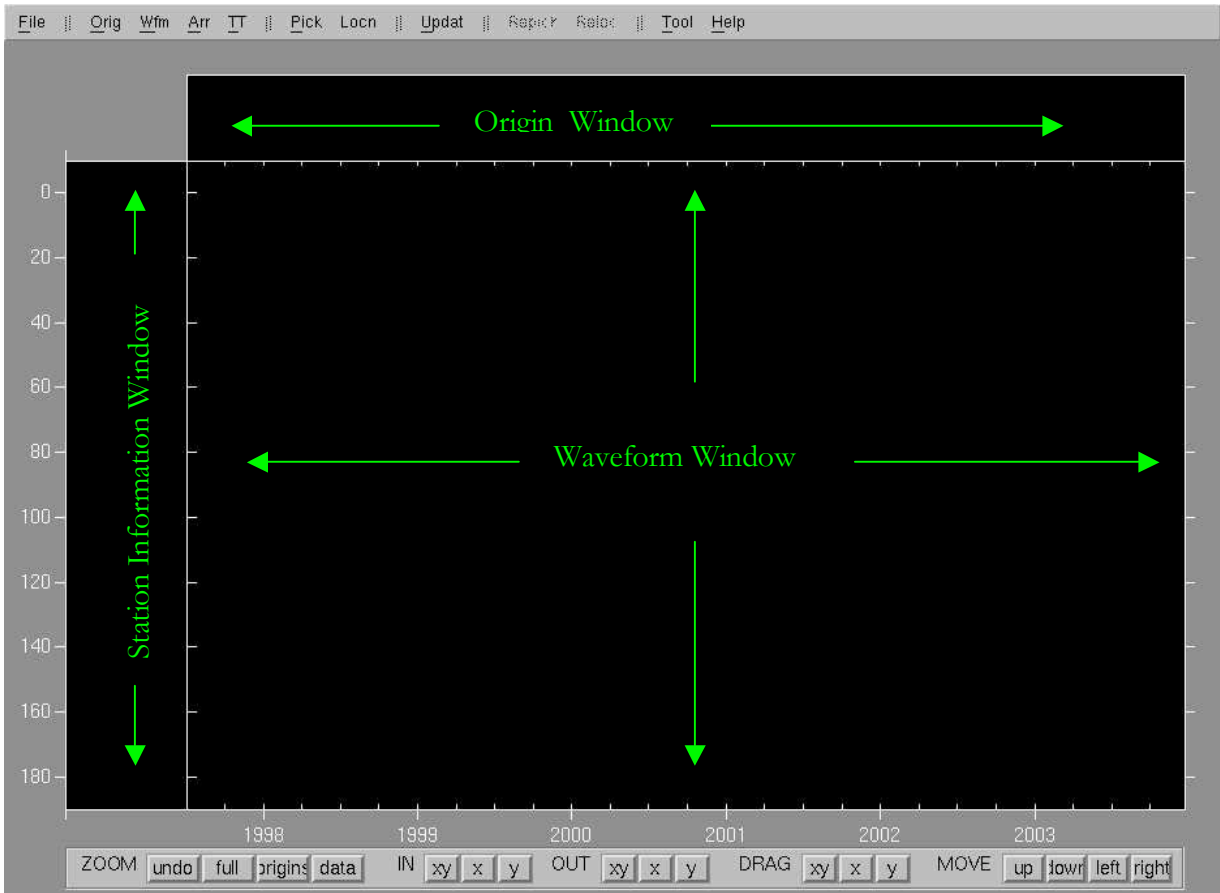


Figure 1. The main GUI window of PLRR is shown here. The origin, station, and waveform windows are used to display data just like in MatSeis.

The menu selection structure of PLRR is shown in figure 2.

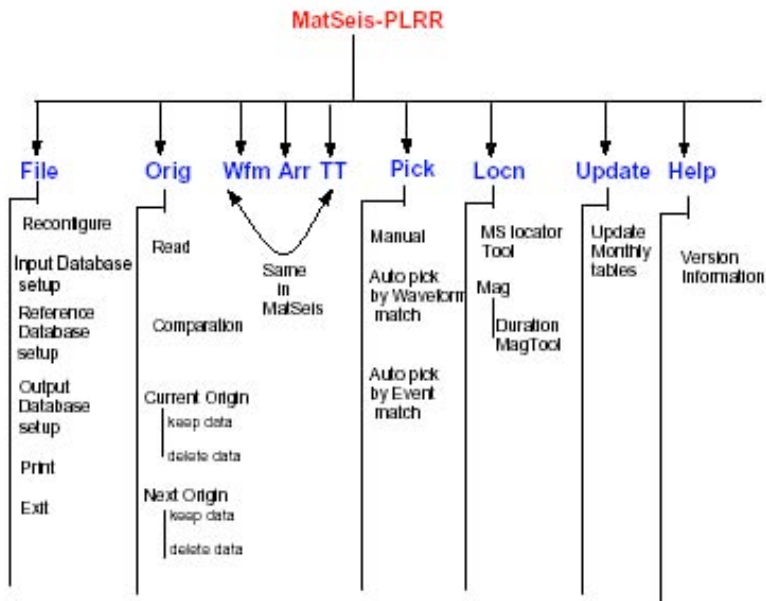


Figure 2. Gives the menu structure within Matseis-PLRR.

**PLRR** uses three CSS database references in its analysis: the Input, Reference and Output databases. The input database represents the data for the new events to be processed using PLRR methods. The reference database, as its name implies, is the database holding the set of cataloged events that the new events are being compared to using WCC. The last database is the output database, in which the processed events are placed with locations obtained through the PLRR processing method. The setup windows for all of three databases are similar and are shown in Figure 3. It is most convenient to save all information into \*.input files and use Reconfigure to read it.

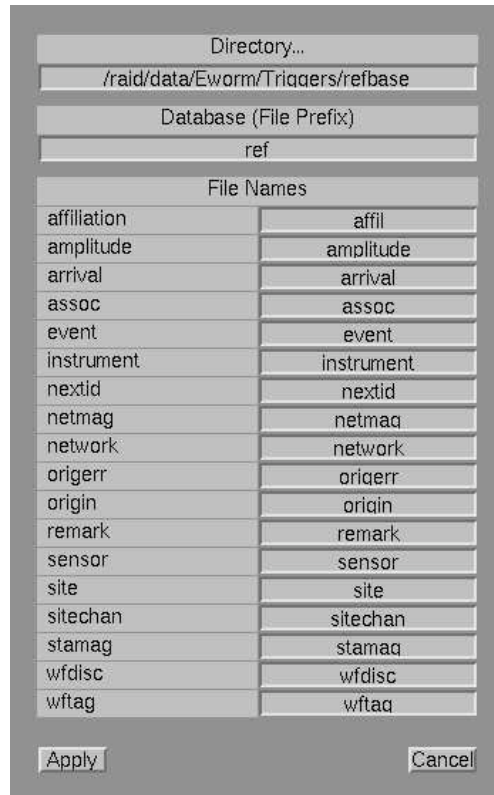


Figure 3. Database Setup GUI window for the Reference database. Set the directory path and database prefix for the database.

Once the database setup is ready, you can use **Read** under **Orig** sub-menu to read the Origin table from the input database into the main GUI window.

**Comparison** is a tool to plot simple dendrograms, cross-correlation matrices, and waveforms between one waveform in the main GUI window and selected reference waveforms for the same station. You cannot run **Comparison** if you do not select *orid* in the main GUI window. The functions of **Current Origin** and **Next Origin** are the same as they are in Matseis.

**Wfm**, **Arr** and **TT** are kept as same as they are in **Matseis 1.7**; you can find a more detailed explanation in the **Matseis** user manual. The sub-menu **Pick** provides three types

of picking methods. Manual is the **Measure-Tool** in **Matseis**, a tool to assist in detecting arrival phases manually. I use it to help find accurate phases for the reference set.

**Auto pick by Waveform match** and **Auto pick by Event match** are two automatic phase detecting tools using the WCC technique. The former one is more elementary than the latter. It is based on the comparison of waveform-pair. The latter one is based on the comparison between event-pair, and it is more robust than **Auto pick by Waveform match**. These methods were developed to process the seismic data of SC network, and you might have to modify them if you have a problem processing seismic data from your research region.

All tools under **Locn** are the same as they are in **Matseis**. **Auto pick by Waveform match** does not provide the location information, and you have to use **MS locator Tool** under **Locn** sub-menu to locate the selected event after you have applied it. Duration magnitude has been used to calculate seismic magnitude in New Mexico for several decades. You have to find a duration magnitude equation for your region if you want to apply this tool to calculate the magnitude. There is only one function, **Update Monthly Tables**, under **Update** sub-menu. Once you have finished processing a new event, you can run **Update Monthly Tables** to add this event into your output database.

As shown in figure 4, the **Update Monthly Tables** GUI window shows the contents of three CSS tables: *Origin*, *Arrival* and *Wfdisc*. You can modify the information in *Origin* table and select which waveforms and arrivals to save into the output database. You have to click **Change** before you finally click **Update** to save these CSS tables. The button **Restore** is empty now.

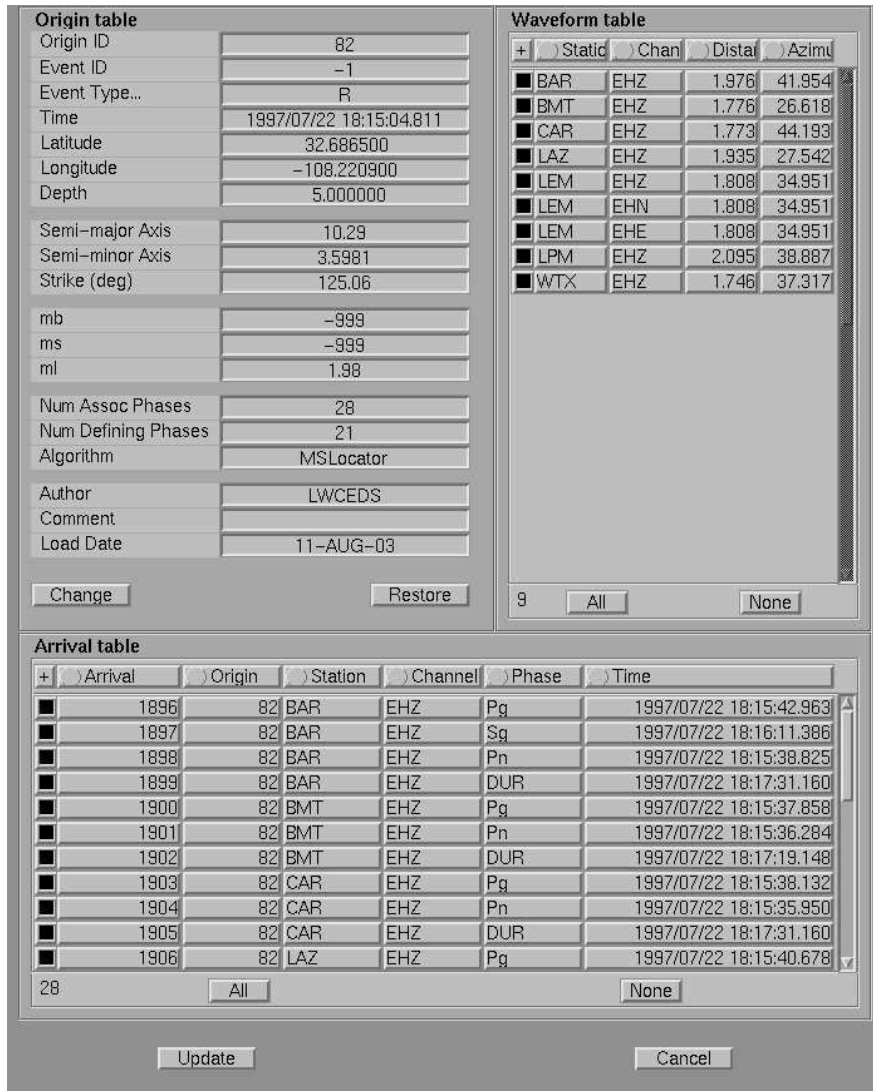


Figure 4. Update Monthly tables GUI window

## 2. Reference set

The quality of the reference set dominates the accuracy of the automatic phase picking. Based on our limited experience, we put forward the following criteria to select events for the reference set:

1. Only include waveforms with high signal to noise ratio. The maximum absolute amplitude ratio between signal and noise should large than 3.0.
2. Only include waveforms with distinguishable phase structure.
3. Remove glitches from the waveform.

There is an express way to build reference set:

1. Set the output database to be the reference set.
2. Read in an event into the main GUI window of **Matseis-PLRR** and manually pick phases and locate it.
3. Use **Update Monthly tables** to write this event into the reference set.
4. Run **sort\_wfdisc** in **Matlab** command window to sort and generate *wftag* table.

### 3. Automatic phase detection

Once there is an event selected in the main GUI window, you can click **Auto pick by Waveform match** to find the most similar reference waveform to each waveform of the event station by station. You also can find the phases for several waveforms by selecting them before running **Auto pick by Waveform match**. Since **Auto pick by Waveform match** is specially designed to process seismic data collected in SC network, you have to modify it if your data is significantly different from our data. The seismic data in New Mexico consists of short-period data (EHZ) and broadband data (BHZ). The sampling rate of EHZ station is 100 samples per second, and the sampling rate of BHZ station is 40 samples per second. In the source code of **Auto pick by Waveform match** (**plrr\_pic\_auto.m**), we band pass the EHZ seismic data using `butter(3,[0.1, 0.6])` and band pass the BHZ seismic data using `butter(3,[0.2, 0.6])`. We also default 0.55 to be the `cutoff` for cross-correlation in **plrr\_pic\_auto.m**. If the process cannot find the cross-correlation value for one waveform larger than the `cutoff` value, there is no phase assignment to that waveform. **Auto pick by Event match** has a GUI window (figure 5), through which you can select any reference event to match to any station available to that event. Since the GUI window is still being developed, only the **Match Method** option selection is available for use. The other option selections are fixed, and the button **Matrix** is empty in version 0.1.

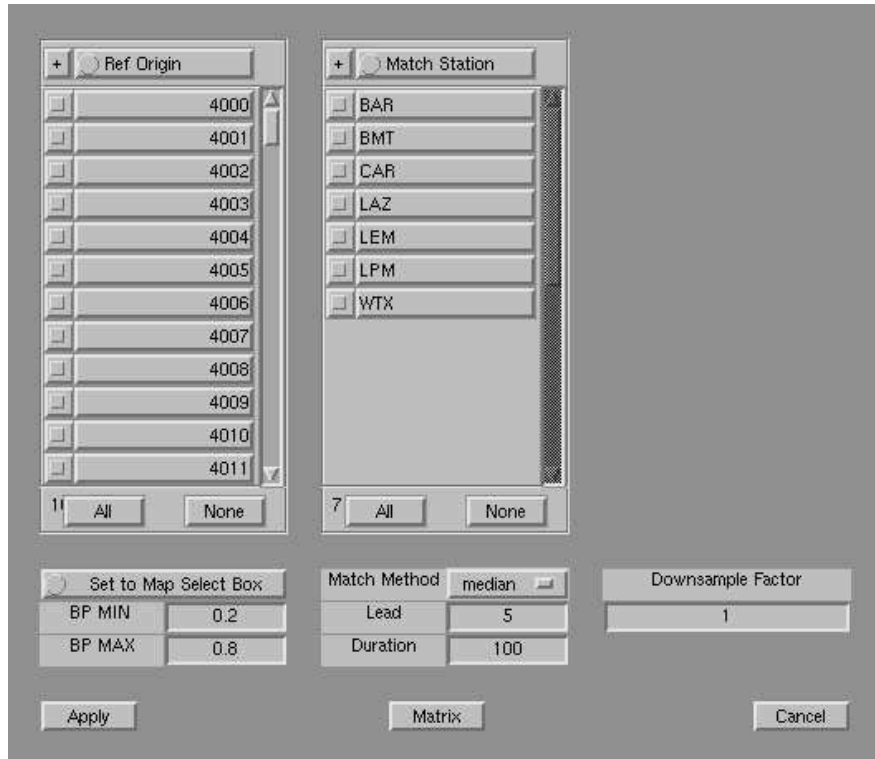


Figure 5. Event Match GUI window

**eventmatch.m** is the Matlab code to setup the event match GUI window, and **findmatch.m** is the Matlab code adjusted by **eventmatch.m** to calculate the cross-correlation value and stack the CCc curves based on different types of match methods. In **findmatch.m**, we set the default sampling rate to be 100 samples per second, and resample BHZ data from 40 or 20 samples per second to be the default sampling rate. We also set the default waveform window to be the first 100 seconds of the waveform before carrying out cross-correlation calculation. If the length of waveform data is less than 10000 (Default sampling rate  $\times$  Default window), we zero pad the data. If the length of waveform data is large than 10000, we only take the first 10000 samples. After cutting the data to fit the window, we band pass the data ( band pass parameters can be set in **Event Match** GUI window). BP MIN is the low cutoff and BP MAX is the high cutoff. We also use *butter* and *filtfilt* functions in Matlab to finish band pass filtering. Waveform data, after processing, are converted to Hilbert envelope in order to increase signal noise ratio. In **Match Method** option selection, **median** takes the median of all waveform-pair cross-correlation coefficient curves. **maximum** takes the maximum of all waveform-pair cross-correlation coefficient curves. **mean** takes their mean and **weighted** weights the curves one by one over epicenter-station distance. **Auto pick by Event match** finds the most similar event in the reference set to the selected event in the main GUI window.

database or lots of small CSS databases, you can write a Matlab code to finish the job. **plrr\_auto\_nov.m** and **plrr\_auto2\_nov.m** are two examples which we used to process

daily CSS databases in November 2003 recorded by Earthworm system of SC network using **Auto pick by Waveform match** and **Auto pick by Event match** methods, respectively.

## 5. Example

In the example subdirectory of this package, there is one CSS database named *ref\_ex* (*ref\_ex.\** tables + **wf**), which is an example of a reference set, and there is another CSS database named *newdata* under **newdata** sub-subdirectory, which is an example of a new event. You can use these data to check the two automatic phase picking methods, but you have to modify the dir set in *wfdisc* tables to both of these two CSS databases. You can use **emacs** or **dbset** in **Datascope** to do the job.

Due to limited time, we could not include as many details as we wanted. If you encounter any problems in **Matseis-PLRR**, contact **Wenzheng Yang** ([wyang@ees.nmt.edu](mailto:wyang@ees.nmt.edu)) for more detailed help. Any and all comments or suggestions are welcome.

*Appendix II*

**PLRR SOURCE CODE LIST AND RELATED CODES IN AUTOMATICALLY  
PROCESSING OF THE SC EARTHWORM SYSTEM SEISMIC DATA**

**Plrr M-file list**

plrr\_a\_update.m

plrr\_ms\_zoom.m

plrr\_update\_menu.m

cat2gmt.m

plrr\_nextid.m

plrr\_w3\_rem.m

cataloglist.m

plrr\_nextorigin.m

plrr\_w\_menu.m

corr\_speed\_test.m

plrr\_o\_comp.m

plrr\_w\_read.m

css\_edit.m

plrr\_o\_menu.m

rcorr.m

csstableitem.m

plrr\_o\_popup.m

ref\_waveedit.m

cutsac.m



plrr\_o\_read.m  
rick\_example.m  
db\_setup\_ref.m  
plrr\_o\_update.m  
shiftdata.m  
demean.m  
plrr\_output\_setup.m  
sort\_wfdisc.m  
epicenter.m  
plrr\_pic\_auto.m  
specgram1.m  
event\_swcc\_match.m  
plrr\_auto\_dec.m  
plrr\_pic\_menu.m  
event\_wcc\_match.m  
plrr\_auto\_nov.m  
plrr\_picbyevent\_auto.m  
example\_spec.m  
plrr\_auto\_oct.m  
plrr\_r\_chan.m  
filterwave.m  
plrr\_cc\_matrix.m  
plrr\_r\_site.m  
findmatch\_swcc.m

plrr\_config.m  
plrr\_r\_sitechan.m  
findmatch\_wcc.m  
plrr\_css\_out.m  
plrr\_r\_sta.m  
load\_css.m  
plrr\_dendro.m  
plrr\_r\_wave.m  
ms\_draw.m  
plrr\_exit.m  
plrr\_reconfig.m  
update\_ref.m  
mslocator\_tool.m  
plrr\_f\_menu.m  
plrr\_ref\_setup.m  
waveform2sacm.m  
mslocator\_tool\_aedit.m  
plrr\_input\_setup.m  
plrr\_ref\_sort.m  
waveform\_edit.m  
o\_popup.m  
plrr\_loc\_menu.m  
plrr\_setup.m  
whole\_shift\_phase.m

plrr.m

plrr\_monthly\_update.m

plrr\_tool\_menu.m

write\_css.m

plrr\_a\_menu.m

plrr\_ms\_fig.m

plrr\_tt\_menu.m

### Source Perl script of “dailycss”

```
#!/usr/local/bin/perl -w
#####
# daily routine to convert new coming event of this day to CSS format
# Wenzheng Y. 2003
#####
#

$incomingpath="/raid/data/Eworm/Triggers/incoming";
$outgoingpath="/raid/data/Eworm/Triggers/outgoing/";

open(WOUT,"date '+%Y%m%d' |");
$TODAY =<WOUT>;
close(WOUT);
$TODAY =~ s/\s+$/; #get rid of the end stuff
##$TODAY=$TODAY-2; # 2 days before
$TODAY--; # yesterday
open(WOUT,"date '+%Y%m'|");
$subpath=<WOUT>;
$subpath =~ s/\s+$/; #get rid of the end stuff
close(WOUT);

## If you want to deal the data of last month
## you have to specific the date as the following
## $TODAY='20030630';

print("\n Daily routine to convert Earthworm SAC incoming into CSS format \n Current date:
$TODAY\n");
$PAT="$incomingpath/$subpath/$TODAY*";
print(" The pattern of TODAY ($TODAY) is:\n $PAT\n\n");

open(WOUT,"/raid/users/wyang/outgoing/foreach.pl $PAT echo |");
@newevents=<WOUT>;
close(WOUT);
```

```

# judge new events coming in or not
#

$test=@newevents[0];
$mark=rindex($test,"_");
if ($mark== -1) {
die ("No events came in today, please check earthworm\n");
}
print "New Events coming, in processing ... \n";

# if month directory does not exists, make it.
opendir(DIR, "$outgoingpath$subpath") || mkdir("$outgoingpath$subpath",0775);

#
# readin new events
#
$COU=1;

## loop begin

foreach $event (@newevents) {
print(" New event[$COU]: $event");
$event=~ s/\s+$//; #get rid of the end stuff
@oo=split(/_/, $event);
$orid[COU]=$oo[2];
$eventdir="$TODAY$oo[1]";
print ("ORIGIN = $eventdir  ");
print ("ORID = $orid[COU]\n");

# copy new event of today to yearmonth directory
#
print "cp -rf $event $outgoingpath$subpath/$TODAY$oo[1]\n";
system("cp -rf $event $outgoingpath$subpath/$TODAY$oo[1]");
system("rm $outgoingpath$subpath/$TODAY$oo[1]/init");
system("rm $outgoingpath$subpath/$TODAY$oo[1]/quicklook");
system("rm $outgoingpath$subpath/$TODAY$oo[1]/repick");
system("rm $outgoingpath$subpath/$TODAY$oo[1]/sacelist");

# convert SAC into CSS format using SAC2000
print "/raid/users/wyang/outgoing/sac2css_sc $outgoingpath$subpath/$TODAY$oo[1] -- in processing\n";
system("/raid/users/wyang/outgoing/sac2css_sc $outgoingpath$subpath/$TODAY$oo[1] $TODAY");
print "\n\n";

# stack the CSS tables and fix the problem of ORIGIN table
# generate a CSS table for events of today
#
open(WOUT,"$outgoingpath$subpath/$TODAY$oo[1]/$TODAY.origin");
$originline=<WOUT>;
#close(WOUT);
$temp1=substr($originline,0,47);
$temp2=substr($originline,48,7);
$temp3=substr($originline,57,180);

#
# format output the origin number between 1~1000
if ($orid[COU] <10) {

```

```

$temp2="    $orid[COU] ";
}
if ($orid[COU] <100 && $orid[COU]>=10){
$temp2="    $orid[COU] ";
}
if ($orid[COU] <1000 && $orid[COU]>=100) {
$temp2="    $orid[COU] ";
}
if ($orid[COU]>=1000) {
$temp2="    $orid[COU] ";
}
$out="$temp1$temp2$temp3";

# append the public origin file

open(FILE,">>$outgoingpath$TODAY.origin" || die ("Could not open file. $!"));
#flock(FILE,$exclusive_lock);
print FILE "$out\n";
close(FILE);
# end of append ORIGIN table
$COU++;
}
#
# loop end
print "update wfdisc table\n";
system("cat $outgoingpath$subpath/$TODAY*/$TODAY.wfdisc > $outgoingpath$TODAY.wfdisc");

```