LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

UCRL-TR-229420

A Model-Based Signal Processing Approach to Nuclear Explosion Monitoring

A. Rodgers, D. Harris, M. Pasyanos

March 26, 2007

This work was performed under the auspices of the U.S. Department of Energy by University of California, Lawrence Livermore National Laboratory under Contract W-7405-Eng-48.

A Model-Based Signal Processing Approach to Nuclear Explosion Monitoring

Arthur Rodgers, David Harris and Michael Pasyanos

Ground Based Nuclear Explosion Monitoring Program
And Seismology Group
Atmospheric, Earth and Energy Sciences Department
Energy and Environment Directorate
Lawrence Livermore National Laboratory
Livermore, CA 94551

Report for LDRD Project 05-ERD-019
March 2007

Summary

This report describes research performed under Laboratory Research and Development Project 05-ERD-019, entitled "A New Capability for Regional High-Frequency Seismic Wave Simulation in Realistic Three-Dimensional Earth Models to Improve Nuclear Explosion Monitoring". A more appropriate title for this project is "A Model-Based Signal Processing Approach to Nuclear Explosion Monitoring". This project supported research for a radically new approach to nuclear explosion monitoring as well as allowed the development new capabilities in computational seismology that can contribute to NNSA/NA-22 Programs.

Current methods for seismic monitoring reduce seismograms to a small number of derived measurements from observed signals, such as the arrival times and amplitudes of waves with specific propagation paths through the Earth (i.e. phases). These measurements are interpreted in terms of deterministic one-dimensional (1D) Earth models, possibly with empirical or model-based corrections for three-dimensional (3D) structure. These models are often inaccurate due to our lack of knowledge of detailed sub-surface structure on a wide range of length-scales. Our lack of knowledge of structure generally becomes more important as the frequency of interest increases, inhibiting our ability to model high-frequency (> 1 Hz) signals and requiring the use inherently limited empirical calibration strategies to monitor small events at regional distances (< 1000 km).

We present a new paradigm for seismic monitoring of seismic events, including earthquakes and underground nuclear tests. The method has the potential to lower detection thresholds over conventional methods, by taking advantage of information in the entire waveform using correlation methods. The use of correlation methods has been shown to lower the detection thresholds with empirical waveform templates (e.g. match filtering and subspace detection). The method presented here uses theoretical signals, rather than observed signals, to build templates for correlation detection.

We demonstrate a model-based signal processing approach composed of several elements: the simulation of seismic waves in stochastic 3D Earth models and coherent signal-processing to represent observed seismograms in terms of theoretical seismograms. Rather than use a single "optimal" Earth model to predict observations, we use sets of stochastic models of Earth structure. These models incorporate constraints from multiple data sets using a Markov Chain Monte Carlo method and provide a range of plausible three-dimensional models for a region whose variability reflects limits of our knowledge of the true structure. We then compute the complete seismic response of each model using the Spectral Element Method (SEM) and high-performance computing or more computationally efficient path-specific 1D averages. The resulting theoretical signals are combined using coherent signal processing to represent an observed signal in terms of a linear combination of theoretical signal templates.

Results indicate that the 3D synthetics can be combined to represent the observed signals very well, but for low frequencies. This is impractical for lowering detection thresholds, but promises to be effective as resolution of seismic structure improves and computational methods and power facilitate the calculation of synthetic seismograms. In order to increase the bandwidth of theoretical signals, we took an alternative approach useing more computationally efficient theoretical seismograms based on path-specific 1D averages through the 3D velocity models. This allowed us to increase the frequency content and the number of template waveforms for correlation detection. This method performs well and very clearly detects a moderate (m_b = 5.1) mainshock earthquake for which the templates were designed. It also detects a smaller (m_b = 4.4) event that is difficult to detect with conventional energy detection. Combining up to four stations in a network at regional distances (460-1060 km) improves detection confidence and reduces spurious detection of teleseismic events. Finally, we show that when templates were designed with an explosion source at shallow depth we could not detect earthquakes that occurred at normal crustal depths (~10 km).

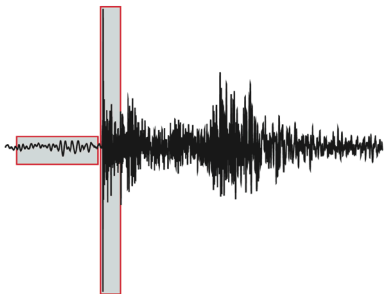
We suggest a two-pronged strategy for advancing this technology: 1) exploitation of increasingly realistic 3D stochastic earth models will improve correlation of theoretical

templates with event signals; and 2) exploitation of larger networks of seismic stations to reject background events that from target source(s). Both of these strategies take advantage of current trends in the broader seismology community to improve models, deploy ever-greater numbers of seismic stations and take advantage of a progressive drop in computational costs. The results of this study established the framework for Model-Based Signal Processing by showing that theoretical signals can be used in subspace detection and that network processing improves network detection performance.

Table of Contents

1.0 Introduction

Conventional methods for seismic monitoring of underground nuclear explosions are heavily dependent on the use of measured quantities from seismograms and deterministic models of Earth structure. For example, seismic events are typically located by minimizing the difference between observed and predicted arrival times of major P- and S-wave phases from one-dimensional (1D) Earth models, possibly with corrections for or three-dimensional (3D) structure. This procedure works well when the event is large enough to be observed with high signal-to-noise ratio and the travel time predictions are accurate. Event identification works by comparing the amplitudes of high-frequency P-waves (through body-wave magnitude, m_b) with low frequency surface wave amplitudes (through the surface wave magnitude, M_S). For smaller events when M_S cannot be estimated ($m_b < 4.0$) high-frequency (0.5-10 Hz) regional discriminants (e.g. Pn/Sn, Pn/Lg, Pg/Lg amplitude ratios) are used. These methods rely on detections made using energy detectors that compare short-term average to long-term average (STA/LTA) amplitude ratios in a pre-defined frequency band (Figure 1.1). For example, in the 1970's and 1980's large underground nuclear tests at known test sites were detected, located and identified using observations at teleseismic distances (> 2500 km) and the wealth of knowledge gained from previous tests. These signals were typically large enough to be observed with high signal-to-noise ratios at large distances, even worldwide.



Regional distance seismogram showing windows for traditional short-term average/long-term average (STA/LTA) detection. This seismogram has high signal-to-noise ratio and can be easily detected with the STA/LTA method.

As new nuclear states emerge it has become increasingly important to monitor broad regions without previous tests and to monitor at lower the detection thresholds. Smaller events do not generate as much low-frequency energy, inhibiting use of the surface waves. This requires the use of regional distance (< 1000 km) signals at higher frequencies, typically above the microseismic noise peak at 0.17 Hz. Regional distance signals at these frequencies are very complex because the wavefield interacts with the heterogeneous crust, including surface geology and topography. The seismic monitoring community has responded by performing detailed region-specific calibration of travel times (for event location) and amplitudes (for event identification). Empirical calibration by design depends on using the available observations and these can be limited. A further challenge to modeling the behavior of regional distance seismic observations for monitoring is the fact that our knowledge of detailed seismic velocity structure is poor and the error in travel time and amplitude predictions is often unknown. An obvious limitation to conventional seismic monitoring calibration is reached when signal amplitudes are comparable to noise levels. In this case, the main pulses of energy cannot be detected with STA/LTA methods because the signals are comparable or weaker than the noise and the subsequent functions of location based on travel times and identification based on amplitudes cannot be completed. Events that are not detected will be missed by the monitoring system.

2.0 Coherent Signal Processing

We address these limitations to conventional seismic monitoring by relying on correlation detection, where more information present in the whole waveform is used, rather than the derived observables such as travel time and amplitude. Coherent signal processing relies on correlation properties between signals and requires a signal model for comparison with a data stream. This model can be an observed signal or set of signals or, in the case of Model-Based Signal-Processing, a set of theoretical signals. In this section we describe the application of coherent signal processing to seismic monitoring.

2.1 *Correlation Detection*

The simplest correlation detector is a match filter. In this case a single template is compared to data stream. At each sample the current data signal window is compared to the template signal and a detection statistic, such as the correlation coefficient, is computed. The template can be constructed from a single or multiple channels, such as single-station three-component or seismic array data. The detection statistic between the data stream and the template is recorded for each instance of the stream. When the detection statistic exceeds a certain value, a detection is declared (Figure 2.1). In the presence of Gaussian noise there is a well-established relationship between the probability of a false alarm and the probability of a missed detection. When the detection threshold (i.e. the value of the detection statistic) is set low the operator must deal with the consequences of a high false alarm rate.

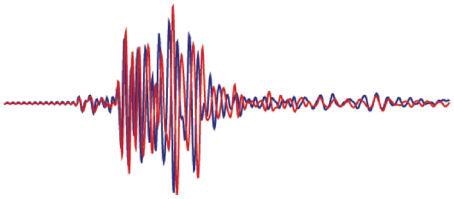


Figure 2.1 Example of a match-filter where a new signal (red) is detected by a high correlation with the template signal (blue).

Match filters are excellent for observing an exact (or near exact) repeat of an event one has already recorded. However, when the source occurs at a slightly different location or depth, or has different source properties (e.g. magnitude, moment tensor, source time function and/or directivity) the correlation between the new event and the template is degraded. In such cases it is possible for the detection statistic to fall below the user-defined detection threshold and the event will be missed.

2.2 *Sub-Space Detection*

In the case where a match-filter misses an event because of degradation in the correlation between the signal for a new event and the template signal, a remedy is possible but requires additional template signals. When multiple template signals are available they can be combined to achieve optimal correlation with a given instance of a data stream.

The templates should have some variability that arises from differences in the sources or propagation paths. This is the principle of a sub-space detector. Multiple templates of a target event are gathered, aligned and arranged in a matrix. The matrix is decomposed into its eigenvectors, ranked by eigenvalues, using singular value decomposition. A subset of the eigenvectors is then used to form linear combinations that optimally fit each instance of the data stream (thus the name sub-space). The sub-space method requires the choice of sub-space dimension, that is the number of eigenvectors to use to form linear combinations that mimic the data stream. These methods are described in Harris (1989, 1997) and Harris and Paik (2006).

3.0 Stochastic Earth Models

In this section we describe the three-dimensional (3D) earth models used to compute synthetic seismograms for model-based signal processing.

3.1 *The Markov Chain Monte Carlo Method to Estimate Seismic Velocity Structure*

Conventional methods to estimate seismic velocity structure rely on formal inversion of observables, such as body-wave travel times and surface wave dispersion measurements. Such methods typically rely on a linear approximation relating the observables to seismic velocity structure and invert a large linear system of equations to solve for the model parameters. These inversions often predict the observations used to estimate the model, but do not predict other observables due to different sensitivities of data to structure. For example structures estimated from surface wave dispersion likely do not predict P-wave travel times because of the differing sensitivities to P- and S-wave velocity structure and different depth sensitivities. In order to reconcile models estimated from different data researchers have begun to combine different data sets. An example of this is the joint inversion of surface wave group velocity dispersion and receiver functions for shear velocity structure near a broadband seismic station (Julia et al., 2000).

In order to reconcile different sensitivities of seismic data to velocity structure, a new methodology has been applied, called Markov Chain Monte Carlo (MCMC). Rather than invert the data for velocity structure, MCMC uses stochastic sampling and extensive forward calculations to estimate velocity structure. Shapiro and Ritzwoller (2002) employed an MCMC method along with *a priori* constraints, linearized inversion and simulated annealing to estimate S-wave velocity structure from surface wave dispersion measurements. Recently, more extensive application of the MCMC method has been used to estimate three-dimensional velocity structure of the crust and uppermost mantle for a large region of eastern Asia (Pasyanos et al., 2004). The MCMC method and its application to the models used in this study are described in Pasyanos et al. (2004).

3.2 *MCMC Models for the Yellow Sea-Korean Platform Region*

For the purposes of this study we used models of seismic velocity structure of the crust and uppermost mantle, because broadband regional seismograms are most sensitive to the P- and S-wave structure of this part of the earth. The MCMC models parameterize seismic velocity in the Yellow Sea-Korean Peninsula region on a 2° grid of points. At each point the depth-dependent structure is parameterized with 7 homogeneous layers. The MCMC method estimates the thickness and elastic properties (V_P, V_S and ρ, the P- and S-wave velocity and density, respectively) of each layer. The 7 layers considered at each point are: water, upper sediments, lower sediments, upper crust, middle crust, lower crust and uppermost mantle. The water layer properties are fixed and the thickness is set using known topography/bathymetry. Thus there are 4 parameters for 5 crustal layers and 3 parameters for the mantle half-space for a total of 23 parameters for each lateral grid point. The MCMC models span a region of eastern Asia with latitude 23°-57° and longitude 109°-147°.

Figure 3.2.1 shows an example of the variability in crustal thickness for four models. Note that the large-scale features are similar, in that the oceanic crust is thin and the continental crust is thicker and thickens toward the interior. However, the models

demonstrate variability on scale of a few grid points, 2°-6°. This smaller-scale structure leads to differences in the predicted waveforms, as will be shown below.

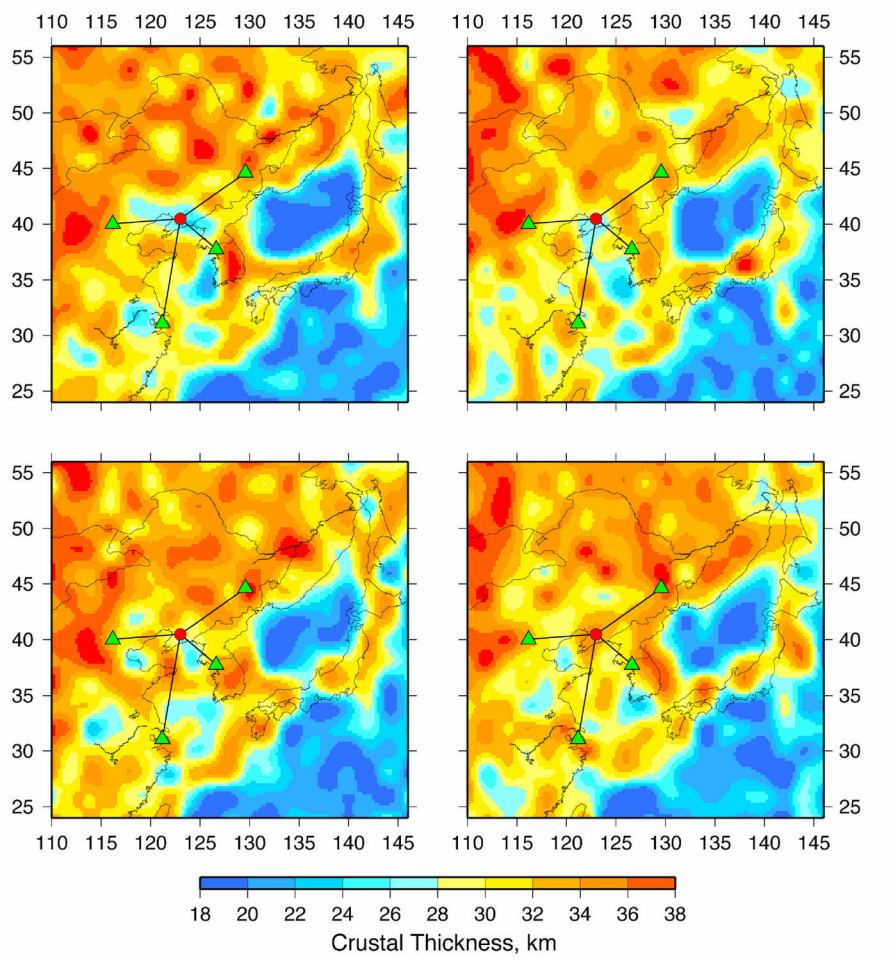


Figure 3.2.1 *Crustal thickness for four different MCMC models. Note that the large-scale features, such as ocean-continent differences, are similar but smaller-scale details are different. Also shown are the locations of the earthquake (red circle) and stations (green triangles) used in our study.*

Under this LDRD project higher resolution (1° grid spacing) models were developed. This effort revealed some aspects of the MCMC application that required improvement, due to the higher computational demand of higher resolution models for the same broad area. Future studies will benefit from this development.

4.0 Small-Scale Stochastic Heterogeneity

The seismic velocity models described in the previous section characterize relatively long-wavelength structure (a 2° grid spacing corresponds to 222 km). A Rayleigh wave with a period of 20 seconds has a wavelength of 60 km. Higher frequencies (longer periods) have even shorter wavelengths. Thus the MCMC models are too coarse to cause scattering of the higher frequencies of interest. Studies by Frankel and Clayton (1986), Levander and Holliger (1992), Wagner (1996) and Pullammananipillil et al. (1997) employed finite difference modeling to explore the nature of stochastic crustal heterogeneity, finding a range possible scales lengths from 200m to 10 km. To address this issue we originally planned to add small-scale (10-220 km) stochastic heterogeneity to the large-scale (220 km) MCMC models. Unfortunately, we were not able to implement this feature within the time constraints; however we were able to develop new codes and methodologies for stochastic heterogeneity that will support future efforts.

4.1 *Fourier Method*

The specification of stochastic seismic wavespeed heterogeneity is commonly done using Fourier transform methods described in Frankel and Clayton (1986) and Frankel (1989). This method can be implemented in two slightly different ways: one can either start with a random field, transform to the Fourier wavenumber domain, filter it according to a prescribed auto-correlation function then transform back to the spatial domain; or one can start with the auto-correlation function in the Fourier wavenumber domain, perturb the phase and transform to the spatial domain.

We developed software to generate stochastic seismic velocity models using the Fourier method. The code works in two- or three-dimensions and accepts a variety of correlation function types: Gaussian, exponential, von Karman (self-affine or fractal). The code allows for different correlation lengths in different coordinate directions, so anisotropic heterogeneity can be represented. Figure 4.1.1 shows two random seismic velocity models in 2D with Gaussian and exponential correlation functions and equal correlation

lengths of 5000 m. Note that the field based on the Gaussian correlation function is smoother than that based on the exponential correlation function. This has been observed previously and many studies find the exponential or von Karman correlation functions are more realistic for representing seismic velocity heterogeneity.

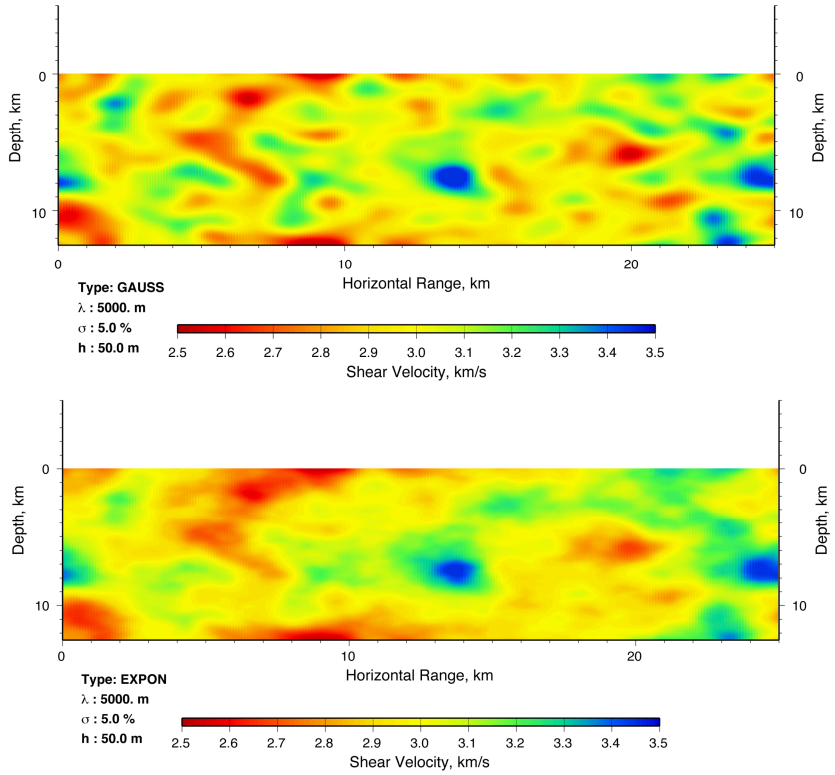


Figure 4.1.1 *Examples of stochastic seismic velocity models generated by our Fourier method code. The fields have (upper) Gaussian and (lower) exponential correlation functions and equal correlation lengths of 5000 m.*

Stochastic heterogeneity affects the seismic wavefield. Figure 4.1.2a shows a 2D random seismic velocity imposed on a background plane-layered model. We computed nine realizations of the stochastic model using an exponential correlation function with root-mean square (rms) amplitude of 4% and horizontal and vertical correlation lengths of 4000 m and 1000 m, respectively. The waveforms computed for a shallow explosion source at a distance of 100 km are shown in Figure 4.1.2b. Notice that the seismograms have the same main P- and S-wave arrivals, but the details and coda of the main arrivals are different due to scattering of the wavefield by heterogeneity.

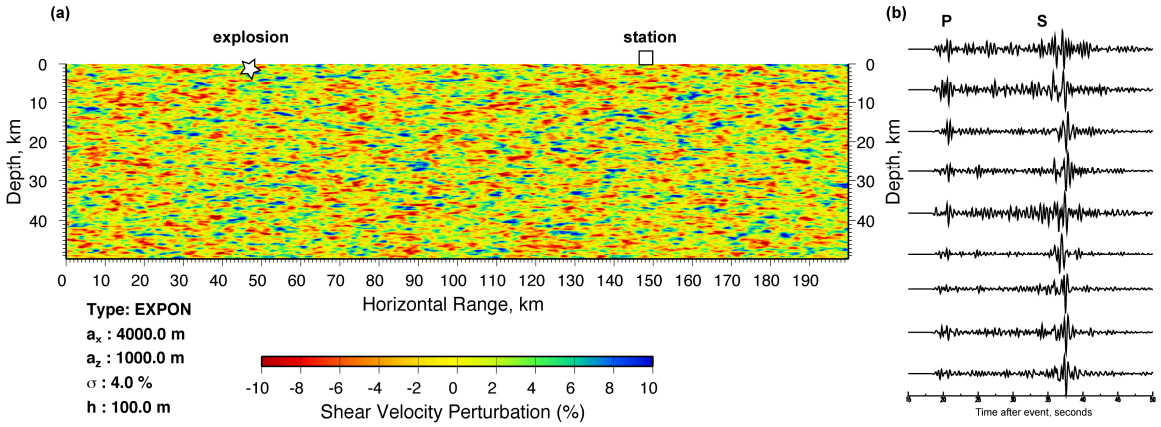


Figure 4.1.2 *(a) Stochastic seismic velocity model with anisotropic exponential correlation structure. (b) Vertical component velocity seismograms for shallow explosion source at 100 km using nine different realizations of the stochastic velocity model with the same correlation structure shown in (a).*

4.2 *Karhünen-Loève Method*

We explored the Karhünen-Loève expansion as an alternative method to generate stochastic heterogeneity. This method uses an eigenvector decomposition of the theoretical covariance matrix linking all points in a grid. As such the Karhünen-Loève method can use very general forms of correlation structure, including non-stationary variation in the heterogeneity. In contrast, the Fourier method applies a constant correlation structure in each coordinate direction for an entire Cartesian domain.

In practice the Karhünen-Loève method works in the following way. First an ordered grid of points in specified and the covariance structure is defined. Then the covariance matrix linking the correlations between all points is formed and decomposed into its ranked eigenvectors (in descending order). Realizations of the stochastic heterogeneity are then created by summing the eigenvectors with random coefficients, weighted by the eigenvalues:

An example of the generality of the Karhünen-Loève method, we show an anisotropic covariance structure in Figure 4.2.1 (a) with the orientation of the layers not parallel to the coordinate axis. This structure can have a spatially varying (non-stationary) correlation structure as is shown in Figure 4.2.1 (b). In this case the correlation length increases along the vertical axis and the structure becomes spatially more smooth as the vertical coordinate increases.

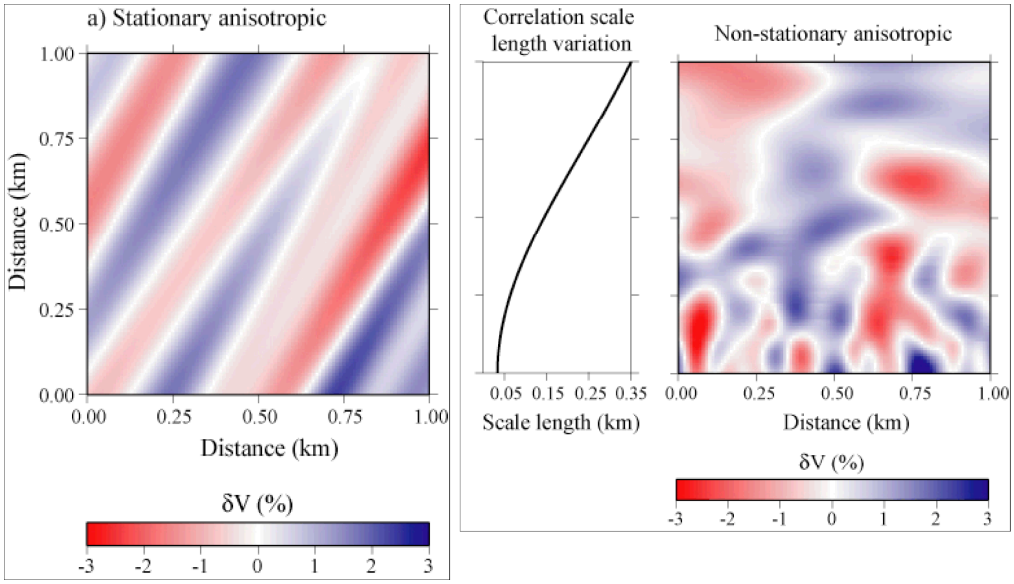
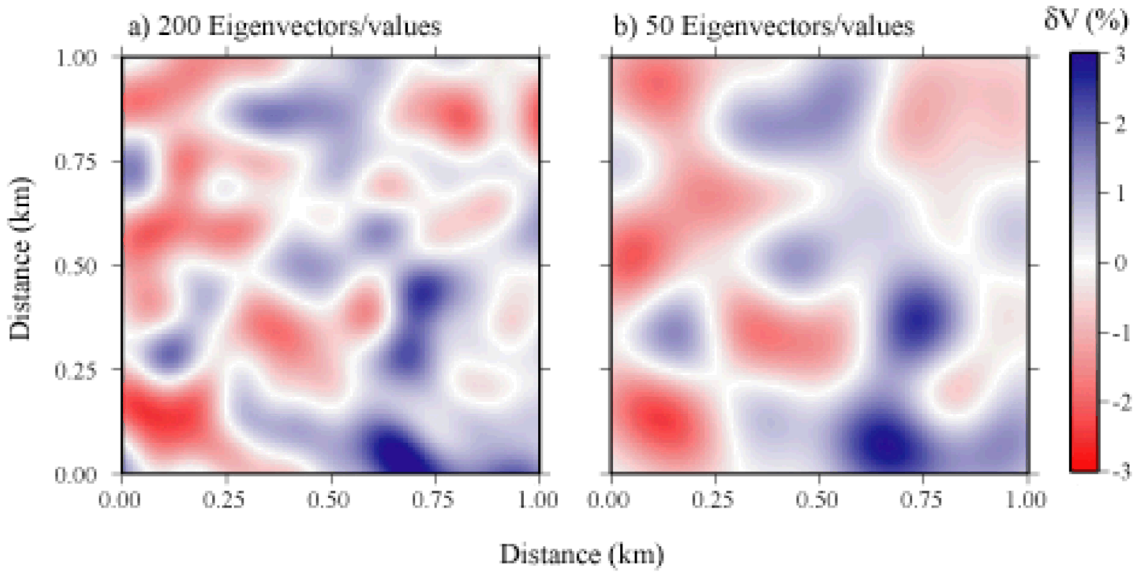


Figure 4.2.1 *(a) Stationary anisotropic covariance structure with the symmetry axis not paralle to the coordinate axes. (b) Non-stationary covariance structure with the correlation length increasing along the vertical axis and the resulting stochastic heterogeneity showing smoother structure for large values of the vertical coordinate.*

Finally, a kriging algorithm is used in order to render the stochastic field onto a dense grid for simulating the elastic response. Figure 4.2.2 shows the effect of eigenvector truncation on the resulting stochastic field. Ideally all eigenvectors should be used, however this is computationally intensive and it is appealing to truncate the eigenvectors to the most significant. For example, a 2D grid with 10,200 points corresponds to 10,200 eigenvector/eigenvalue pairs. It was found that the first 50 eigenvectors included 76% of the total power and 200 eigenvectors included 95.5 %. Figure 4.2.1 shows the stochastic

field generated using either 200 (a) or 50 (b) eigenvectors. Clearly the case with 200 eigenvectors has more variability in the structure.



Realization of the stochastic field with the Karhünen-Loève method using (a) 200 and (b) 50 eigenvectors.

In (a) the remaining 10,000 eigenvectors included only 0.5% of the total power, illustrating that truncation of the eigenvector expansion is appropriate and judicious.

While the Karhünen-Loève method allows for very general covariance structure of stochastic heterogeneity, in practice it is very computationally intensive because the covariance matrix for all points in the grid, as big as the total number of points in the grid must be stored in memory and decomposed into its eigenvectors. These memory requirements limit the applicability of the method in 3D. Michael Thorne started work on this as a summer student (July-August 2004). He continued this effort and presented complete work on the Karhünen-Loève method in 2D at the Fall 2006 American Geophysical Union meeting (Thorne et al., 2006). He is currently working on improvements of the method for efficient application in parallelized codes for seismic wave propagation.

5.0 Synthetic Seismograms

Model-based signal processing requires the calculation of synthetic seismograms in three-dimensional models, such as those described in the previous sections.

5.1 *The Spectral Element Method*

We began by using the Spectral Element Method (SEM, Komatitsch and Villotte, 1998; Komatitsch and Tromp, 1999). An open-source code for computing synthetic seismograms with the SEM in spherical geometry (called SPECFEM3D) was obtained from California Institute of Technology. This code is able to represent all the important features required to model seismic waves. It allows for fully 3D seismic velocities and density structure, surface topography and attenuation. Spherical geometry, rather than Cartesian, is important for modeling seismic waves along paths longer than about 1000 km. The SEM works by combining the flexibility of finite element methods and the accuracy of spectral methods (e.g. pseudospectral methods). Finite element methods are very flexible for representing complex geometries, including free surface topography. They are also very accurate because the boundary conditions at domain surface, the free surface in the case of the Earth, are applied before the equations are discretized. This allows for very accurate handing of the free surface topography and structure along major internal discontinuities (i.e. where material properties change abruptly). The equations are solved in the spectral domain using Legendre polynomials. The SEM uses a finite element mesh, but rather than solve the equations at points evenly spaced along the finite element edges, the SEM uses Gauss-Labotto points judiciously chosen to be the locations of zero values of Legendre polynomials. This provides improved accuracy over, say higher order finite element or finite difference methods, and also leads to diagonalization of the traditional mass matrix used in finite element methods, improving computational efficiency. The code runs in parallel using the Message Passing Interface (Gropp et al., 1994).

Figure 5.1.1 shows the SPECFEM3D mesh for the entire globe. The globe is composed of six chunks and the chunks are subdivided for domain decomposition parallelism. The

different colors in Figure 5.1.1 indicate the domain treated by a single processor. Within each processor's domain are five-by-five mesh elements. The SPECFEM code can be run with a single chunk for domains up to 90° by 90°. Smaller chunks are possible using improvements developed by Brian Savage (University of Rhode Island) and included in more recent distributions of the code.

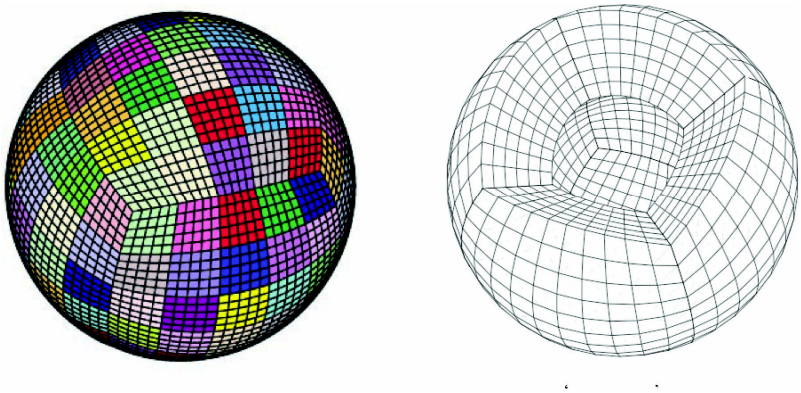


Figure 5.1.1 *Spectral element method mesh used by the SPECFEM3D code for (a) all six chunks that compose the globe and (b) the global mesh with one chunk removed.*

We obtained the code (http://www.gps.caltech.edu/~jtromp/research/downloads.html) and complied on Livermore Computing parallel computers (MCR, Thunder). We modified the code to read our user-specified seismic velocity and density models. The MCMC models for the Yellow Sea-Korean Peninsula region are specified on a 2° grid of points. For seismic wave propagation simulations we must render the material properties on a much finer grid. Typically these methods require 5-10 grid points per wavelength of the highest frequency wave. For 0.1 Hz (10 second period) waves this corresponds to 2.5 km spacing (assuming 2.5 km/s minimum wave speed). For 1 Hz (1 second period) waves this corresponds to 250 m grid spacing.

We modified the SEM global code to read in models specified on a 2° grid, like the MCMC models. Figure 5.1.2 shows the region covered by the MCMC models. These models are specified to a depth given as a mantle half-space, usually less than about 45 km. In order to account for structure in the mantle, we merged the models with the global CUB2.0 model (Shapiro and Ritzwoller, 2002). Also shown in Figure 5.1.2 is a cross-section through the CUB2.0 model showing variations in mantle shear velocities.

Note that the CUB2.0 model includes crustal structure, but that is not shown in the cross-section.

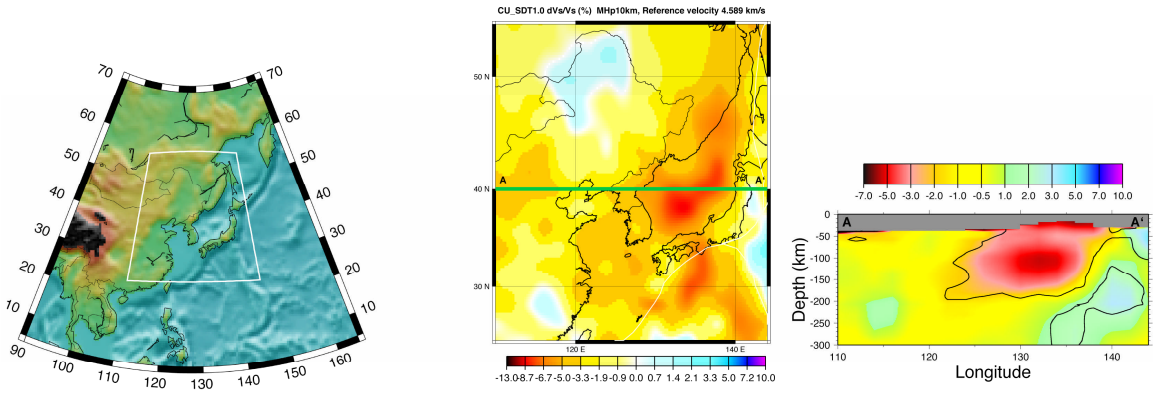


Figure 5.1.2 *(left) Map of eastern Eurasia, with the white lines indicating the coverage of the MCMC models. (right) Cross-section through the CUB2.0 model showing shear velocity variations.*

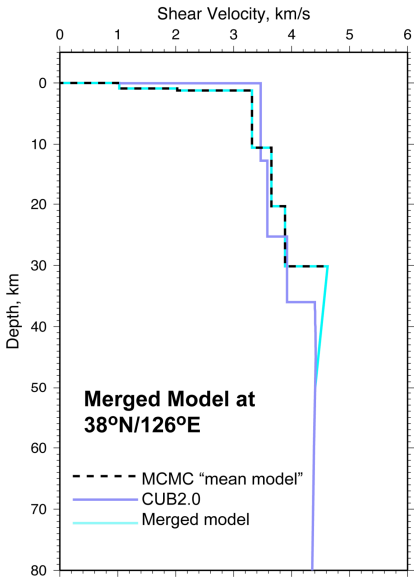


Figure 5.1.3 *Illustration of the merging of MCMC and CUB2.0 models at a location in the Korean Peninsula. The MCMC mean model and CUB2.0 shear velocity profiles are shown (blacked dashed and blue lines, respectively). The merged velocity profile for this location is shown as the cyan line.*

The MCMC and CUB2.0 models were merged to smoothly transition the models at a given location over a depth range from the local Moho depth to 60 km. Figure 5.1.3 illustrates this process. The SEM code requires the velocity and density model at unevenly spaced points, because the Gauss-Labotto points and Legenedre polynomials and the mesh are non-uniform. In order to provide the velocity and density values we developed an algorithm to smooth the models using a 3D Gaussian smoothing operator. The eight models points surrounding a target location were weighted by the Gaussian width of 200 km. This provided a smooth and continuous specification of the velocity and density model.

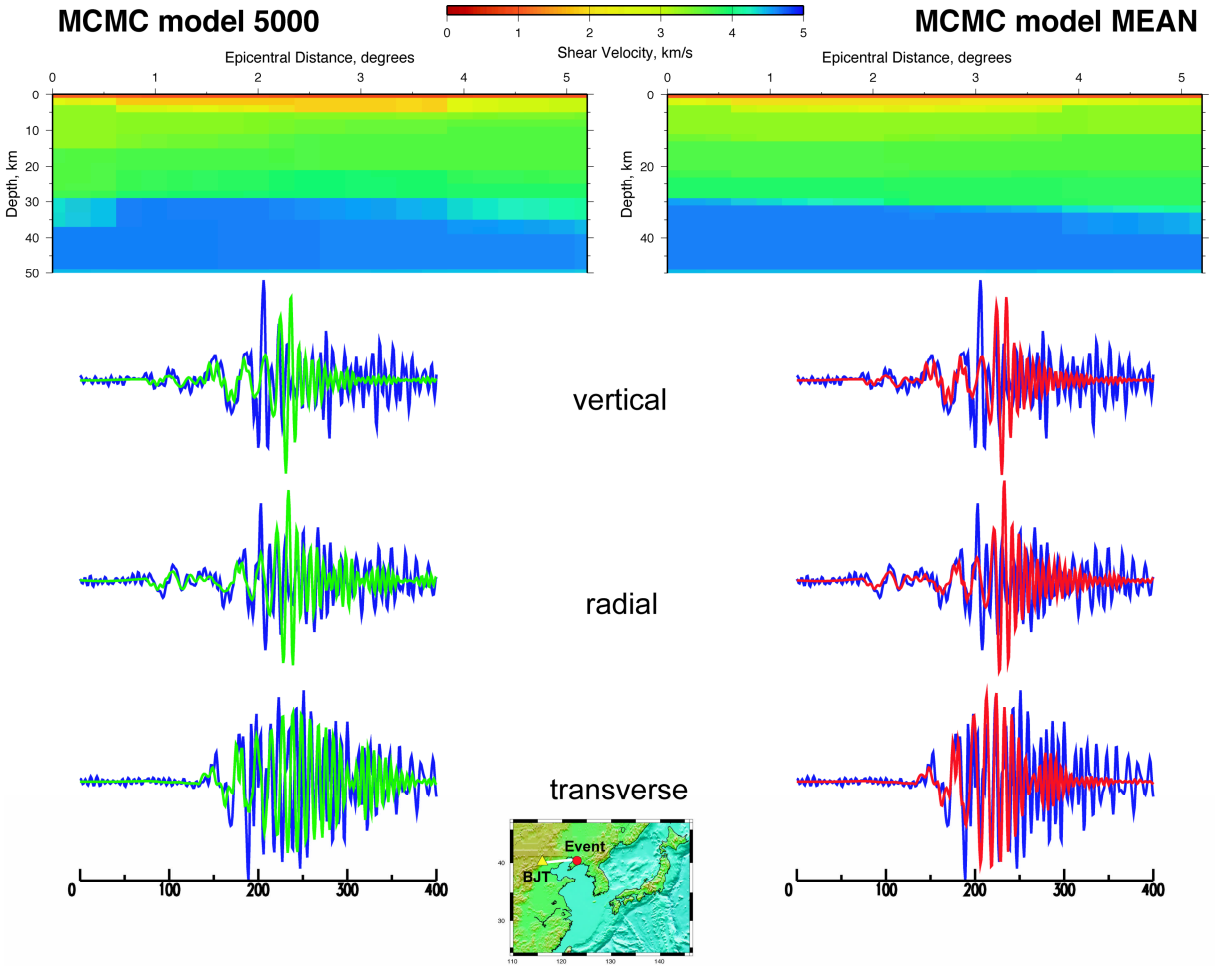


Figure 5.1.4 *Cross-sections and waveforms for two MCMC models (model_5000 and the mean model). The observed (blue) and synthetic waveforms (green and red) are shown. Note that the models differ in the near-surface and this impacts the surface waves.*

We computed the SEM synthetics using "one-chunk" of the global mesh. This covered a solid angle of 90°. We ran the SPECFEM3D code on 144 CPU's using 768 spectral elements along each side of the chunk. This allows accurate simulation of the wavefield for frequencies 0.0-0.1 Hz (periods down to 10 seconds). Simulations were mostly performed on the MCR cluster. Figure 5.1.4 shows the observed and synthetic seismograms and the MCMC model cross-section along the path from our test event to station BJT. Note that the models have slightly different velocities in the near-surface and this results in different surface wave response. Specifically, the lower near surface velocities for model_5000 cause longer duration surface waves than the mean model.

5.2 Wave Propagation Program

A new elastic finite difference code has been developed by LLNL with funding from the LSTO, under LDRD project 05-ERD-079. This code is based on a second-order node-centered finite difference method in Cartesian geometry (Nilsson et al., 2007). It solves the elastodynamic response to moment tensor and point forces with fully three-dimensional varying velocity and density. We ran this code as an alternative to the SEM in order to try to increase the frequency content of the synthetic seismograms. We found that we could only slightly increase the frequency content over the SEM synthetics, however this code will be useful for shorter-range synthetics (e.g. local to near-regional distances, less than 300 km).

6.0 Results

In this section we present the results of subspace detection analysis with model-based signals. We started by running relatively low frequency synthetic seismograms computed with the SEM code.

6.1 Test Event

For this proof of concept study we considered a moderate sized earthquake occurring near the China-North Korea border. The event occurred on January 11, 2002 and had a moment magnitude (M_W) of 4.89. Figure 6.1.1 (left) shows the study area, the event and station locations.

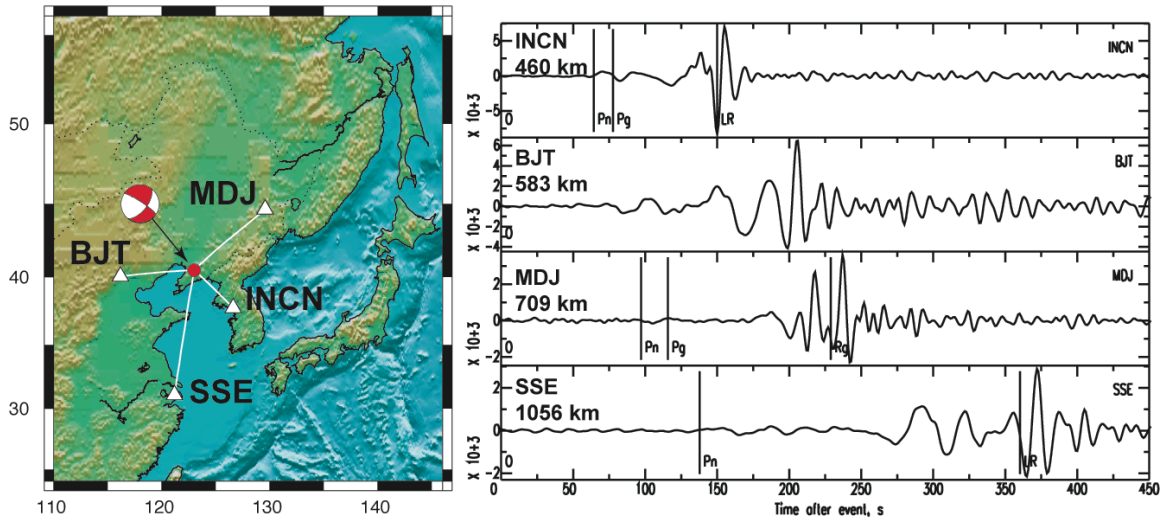


Figure 6.1.1 *(left) Map of eastern Asia showing the earthquake (red circle, focal mechanism) we studied and regional distance stations (white triangles) that recorded the event. (right) Vertical component waveforms (filtered 0.0125-0.1 Hz) from the event at the four regional stations. The station names and epicentral distances are indicated next to each waveform.*

Focal parameters were determined by William Walter (personal communication). Also shown in Figure 6.1.1 are the vertical component waveforms. Broadband waveforms were obtained from four regional stations operated by the Chinese Digital Seismic Network (CDSN, stations BJT, MDJ, SSE) and the Global Seismic Network (GSN, station INCN). These have good SNR for the surface waves at the frequencies of interest for this study. The instrument response was removed, the waveforms were integrated to displacement and the horizontal components were rotated to radial and transverse components. These stations are at regional distance 460 – 1056 km and are representative of typically monitoring conditions.

6.2 Subspace Analysis With Low-Frequency Signals

We computed the model-based waveform signals (synthetic seismograms) using the SEM code for 9 MCMC models. Figure 6.2.1 shows the resulting model-based and observed

waveforms. Generally the model-based signals show very consistent body-waves, with only slight variations in the timing of arrivals. However, the surface waves, especially the later arriving short-period energy, display differences likely related to dispersion and scattering. Note especially the data and synthetics for station BJT. This path (Figure 6.1.1) crosses the sedimentary structure of the Bohai Basin and the data reveal a complex response.

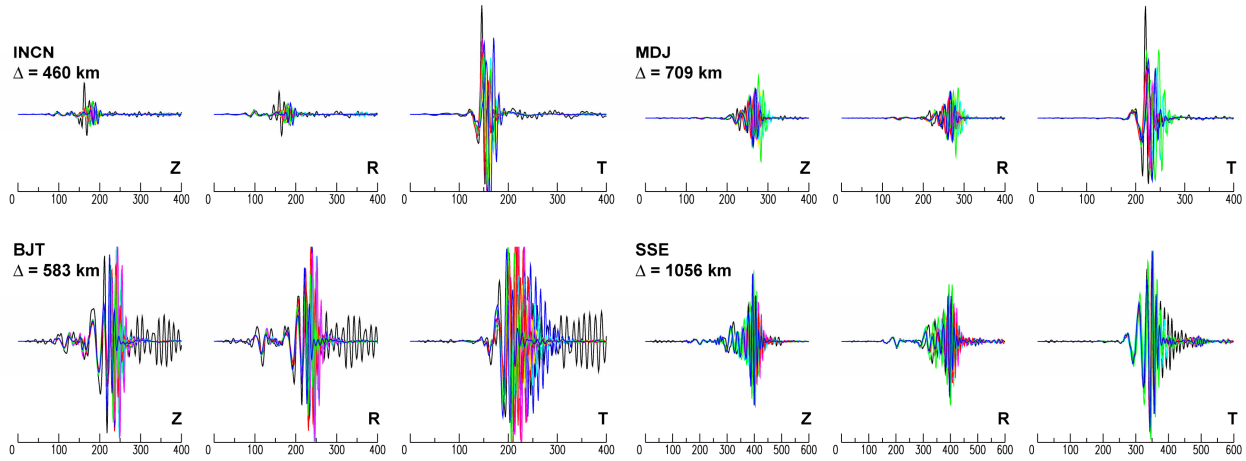


Figure 6.2.1 *Three-component observed (black) and synthetic (colored) waveforms for the four regional stations. There are synthetic signals for nine different models. Z, R, T corresponds to vertical, radial and transverse components, respectively. Data and synthetics are filtered 0.01-0.1 Hz. Note that the time scale is different for station SSE.*

Subspace detection analysis was performed for each station separately. The three-component waveforms (observed and synthetic) were multiplexed into channel sequential order, forming a single vector with *M* total points. The *N* model-based multiplexed vectors were formed a matrix of length *M* and width *N*. Following the subspace methodology the matrix of template waveforms was decomposed into its singular vectors and sorted by most-significant singular value. For this study we computed model-based signals for nine models (*N*=9). The length of signal time windows for each station varied such that the entire surface wave and coda were captured (400-600 s, similar to Figure 6.2.1).

The dimension of the subspace is a significant design parameter a strongly impacts the tradeoff between the probability of detection and the probability of false alarm. The larger the number of significant singular vectors used to represent the observed waveform the better will be the fit to potential signals and the higher the probability of detection. However, using a larger subspace dimension increases the probability of falsm alarm, due to misleading correlations with noise. Determination of the subspace dimension is made, in part, by considering the "energy capture". This is the fractional energy of each template waveform represented by the singular vector basis of dimension *1* to *N*. When all singular vectors are used, each of the N design templates will be perfectly represented, due to the linear dependence of the eigenvectors on the individual templates. The energy capture is computed for each of the *N* original template waveforms and plotted as a function of the subspace dimension.

Figure 6.2.2 shows the energy capture for the waveform templates computed for station BJT and using three different frequency bands. The energy capture indicates that for the low frequency (80-20 seconds) case a subspace dimension of only one or two is needed to represent 95% of the power in each original template waveform. However, the subspace dimension needed to represent the basis signals increases as the bandwidth increases. For the bands 80-15 seconds and 80-10 seconds we use subspace dimensions of 3 and 5, respectively. The subspace dimension must increase as additional complexity is added in the broader bandwidth waveforms. The key for the subspace methodology to work effectively is for the subspace dimension to increase slowly as the bandwidth increases.

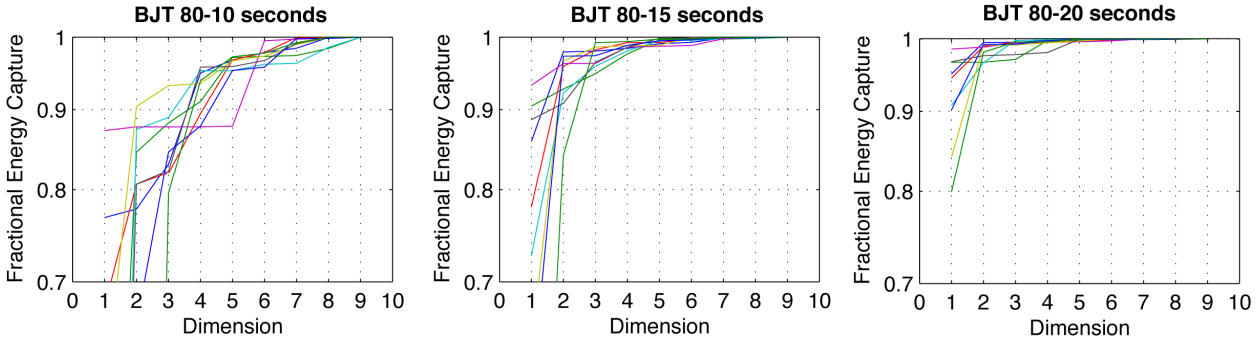


Figure 6.2.2 *The energy capture for templates computed for station BJT in three different frequency bands 0.0125-0.1 Hz (left), 0.0125-0.067 Hz (center) and 0.0125-0.05 Hz (right).*

The subspace representation is then applied to the observed waveforms. To evaluate the performance of the subspace detector we compute the linear correlation between the three-component synthetics for the nine individual MCMC models and for the subspace.

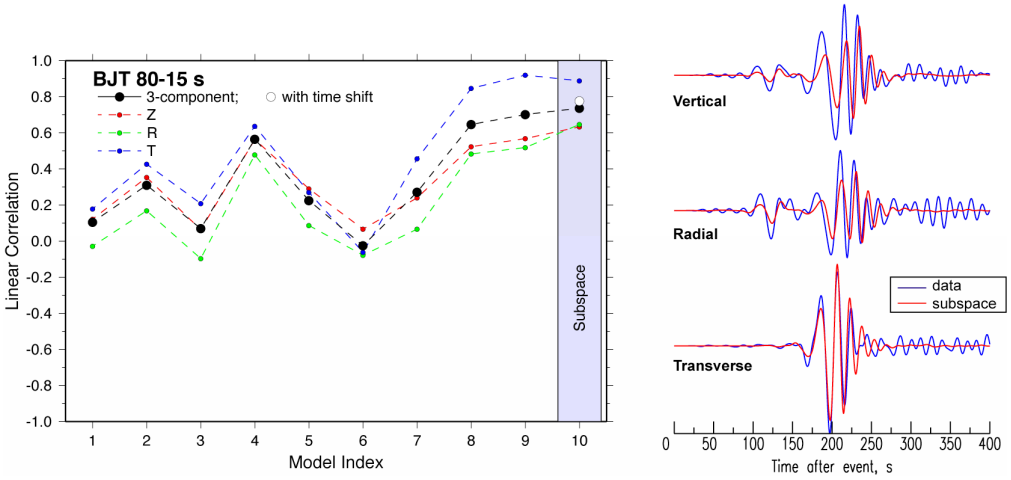


Figure 6.2.3 *(left) Linear correlation between the observed waveforms and the nine individual MCMC model-based signals, the tenth model is the subspace result, using a subspace dimension of three. The linear correlation is plotted for the three-component (black circles) and individual components (colored circles). (right) The resulting fit between the observed (blue) and subspace detector (red) waveforms.*

Figure 6.2.3 shows the linear correlations for the waveforms observed at BJT using the frequency band 80-15 seconds and a subspace dimension of 3. The linear correlations between the observed and individual model-based signals (model indices 1-9) vary between about 0.0 and 0.7. The subspace results in an improved waveform fit over any individual model. While values greater than about 0.5 indicate fairly good waveform similarity the subspace result (about 0.7) should be compared with the average correlation for the individual models (about 0.4) because there is no reason to choose any single model from the MCMC model set. Notice that the resulting waveform for the subspace has the proper surface wave dispersion. For these frequencies the individual

model-based signals do not reproduce the late-arriving scattered surface wave energy that was not present in the basis waveforms (Figure 6.2.1).

The performance of the subspace representation, as measured by the increase in the linear correlation between the observed and the individual model-based and subspace signals, improves as the bandwidth is increased. Figure 6.2.4 shows the three-component linear correlations between the model-based and observed signals for three frequency bands. Increasing the bandwidth introduces additional complexity in the observed and model-based signals and the improvement in linear correlation for the subspace is most dramatic for the broadest band comparisons (80-10 seconds).

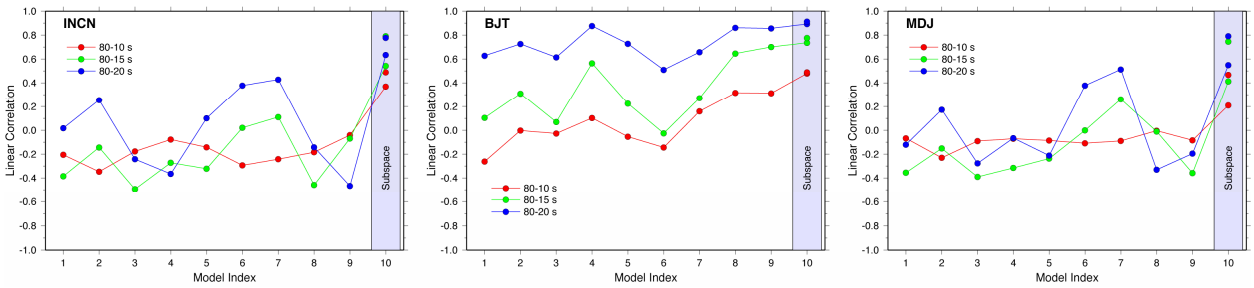


Figure 6.2.4 *Linear correlations between model-based signals (individual, 1-9, and subspace) and the observed three-component waveforms at three stations INCN (left), BJT (center) and MDJ (right). For each station the analysis was performed in three different period bands 80-10 seconds (red), 80-15 seconds (green) and 80-20 seconds (blue).*

The most dramatic increases in linear correlation between the observed and model-based signals are seen for stations INCN and MDJ. The broadband (80-10 s) comparisons are quite poor for the individual model-based signals, but these all increase dramatically when combined with the subspace methodology.

Finally, we tested the effect of adding noise to the observed signals in order to lower the SNR and found that we can still obtain improvements in the linear correlation with the subspace representation over the individual model-based signals (Figure 6.2.5).

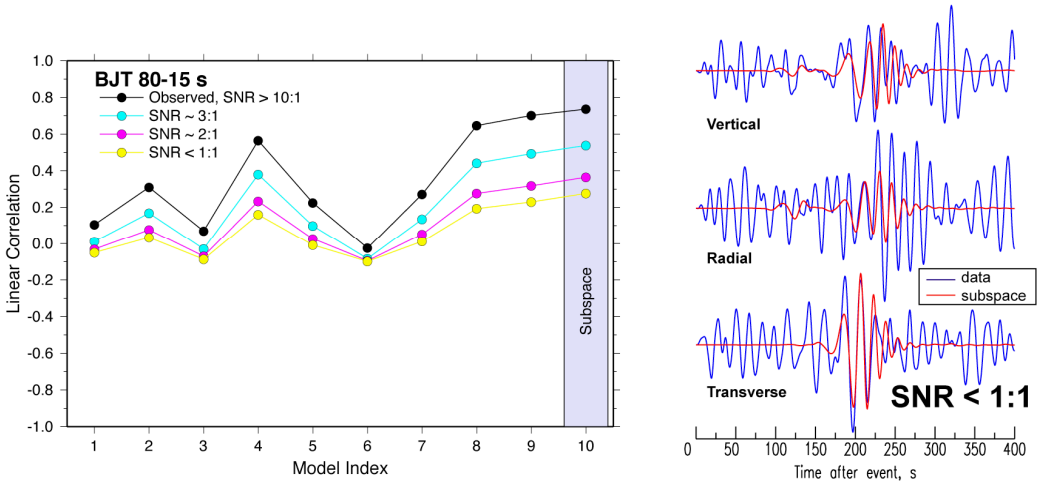


Figure 6.2.5 *(left) Linear correlation between the observed and individual model-based and subspace signals for station BJT (80-15 s) with the addition of noise. The observed waveform has a SNR to greater than 10:1. The addition of noise degrades the correlations, however, the subspace still improves the representation of the observed signal over the individual model-base signals. (right) The observed (blue) waveforms with noise added to overwhelm the signal and the subspace signal (red) show that the correlation is still possible.*

6.3 *Subspace Analysis With Higher Frequency 1D Signals*

The results of the previous section show that the subspace method can improve the correlation performance of model-based signals with data. However, the bandwidth of the 3D synthetics is limited and this inhibits the detection performance on continuous signals. In order to improve the performance in realistic monitoring conditions we increased the frequency content of the model-based signals by taking 1D (depth varying) path-averages of the structure through each 3D model. Synthetic seismograms for the 1D path-average models can be easily computed with the reflectivity method (Kennett, 1983; Randall, 1994).

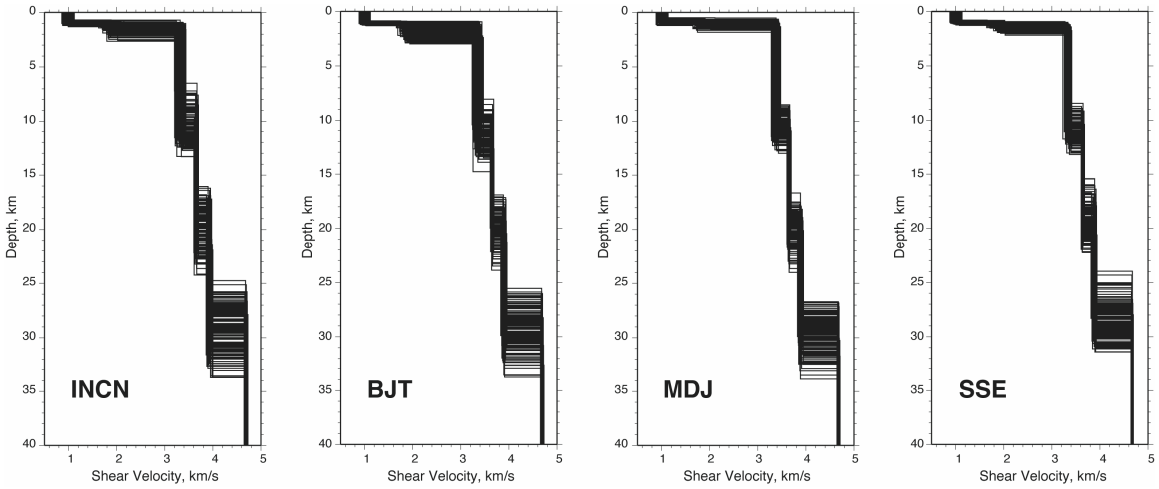


Figure 6.3.1 *Path-averaged 1D velocity models for the four paths considered. There are 50 models for each path (panel).*

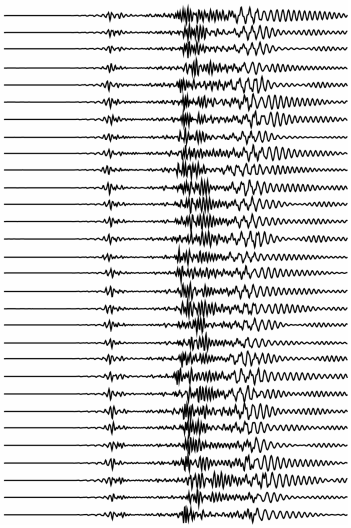


Figure 6.3.2 *Synthetic seismograms for the radial component at station BJT for 30 1D path-average models.*

This allowed us compute more synthetics with higher frequency content, which will perform better for low magnitude events. We computed the 1D path-average models from 150 3D MCMC models for each of the four paths considered (Figure 6.1.1). The model-based signals were computed to a frequency of 2.0 Hz. This is much higher than was possible with the 3D SEM synthetics. Figure 6.3.2 shows the radial component synthetic seismograms for 30 1D path-averaged models to station BJT. Notice that the

arrival times of the body- and surface waves varies, reflecting variability in our estimate of the earth structure along the path.

We performed the subspace detection on continuous data streams using software developed by David Harris and Timothy Paik. This software is described in the Appendix. To begin, we performed the subspace detection on six days of three-component single station data around the time of the January 11, 2002 event. This event had several aftershocks, reported by the Annual Bulletin of Chinese Earthquakes. The locations and origin times of these reported events should be reasonably good due to the relatively dense coverage of seismic stations in China.

The subspace detector reads the continuous data stream (three-component single station in this case) and at each sample computes the optimal combination of template signals. It outputs the detection statistic, which we show for each station in Figure 6.3.3. The detection statistic is the squared correlation between data and subspace signal and varies between 0.0 and 1.0. This shows that station INCN and MDJ detect the main event (m_b 5.1) very clearly (above the background detection statistic). These stations also detect a smaller aftershock (m_b 4.4). Stations BJT and SSE indicate a small peak for the main event. Interestingly the subspace detector gives high detection statistic values for teleseismic events indicated in Figure 6.3.3. This means that the teleseismic waveforms could be represented by the template waveforms for regional event.

When the four-station (three-component) data stream is treated coherently as a network, the detection performance improves. Figure 6.3.4 shows the detection statistics for two days of continuous data and compares the single-station versus network performance.

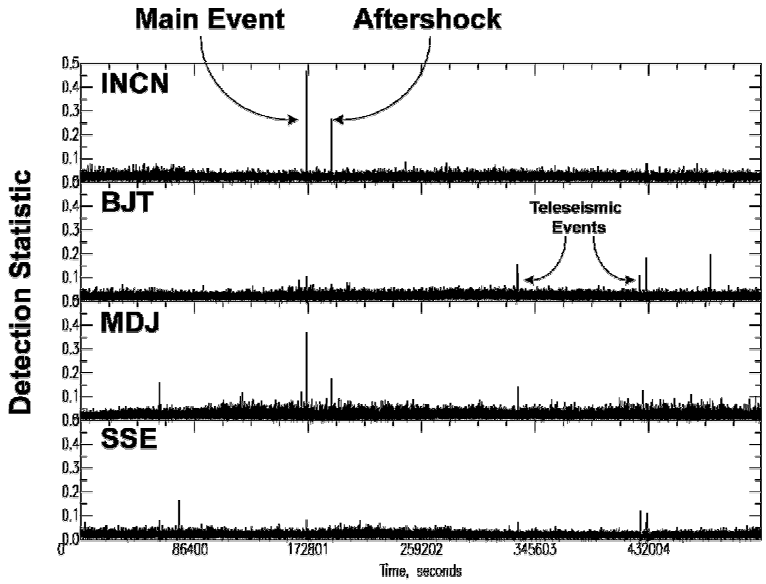


Figure 6.3.3 *Detection statistic for single-station three-component subspace detection using the 1D synthetic seismograms based on MCMC models. The plots shows six days of continuous data.*

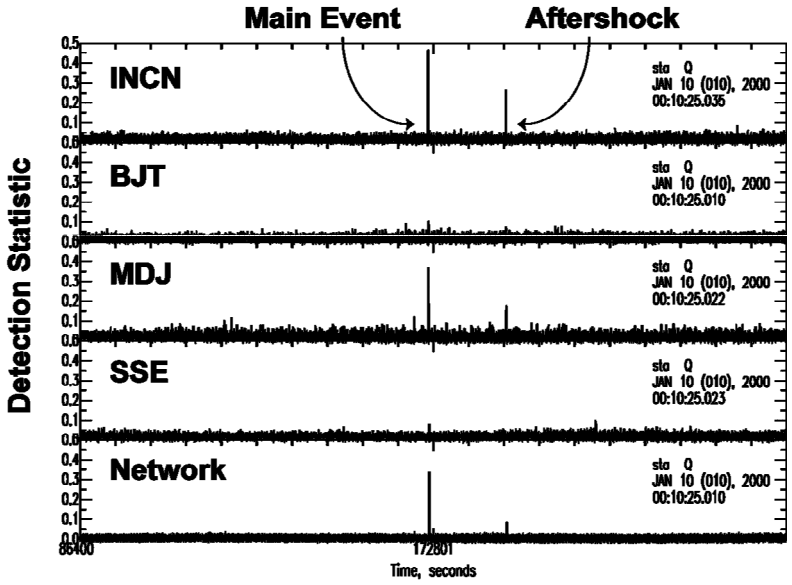


Figure 6.3.4 *(top) Detection statistic for four single station three-component detection (INCN, BJT, MDJ, SSE) same as Figure 6.3.3. (bottom) The detection statistic for the four station network. The plot shows two days of continuous data.*

The detection statistic for the network is reduced for the teleseismic events that are detected in some of the single-station cases. This illustrates the power of coherent

processing because the teleseismic events causes a disturbance across the four-station network that has different timing compared to the regional event. This phase information is crucial for designing a subspace detector to have spotlight detection performance for a location and event of interest.

In addition to location power, Model-Based Signal Processing has identification power as well. The template signals are computed for an event at a specific location and for a source with specific properties. In the example shown above the templates were computed for an earthquake with a double-couple focal mechanism at 10 km depth. To investigate the event identification power of Model-Based Signal Processing, we computed the template signals for an explosion source at 1 km depth at the same location as the January 11, 2002 earthquake.

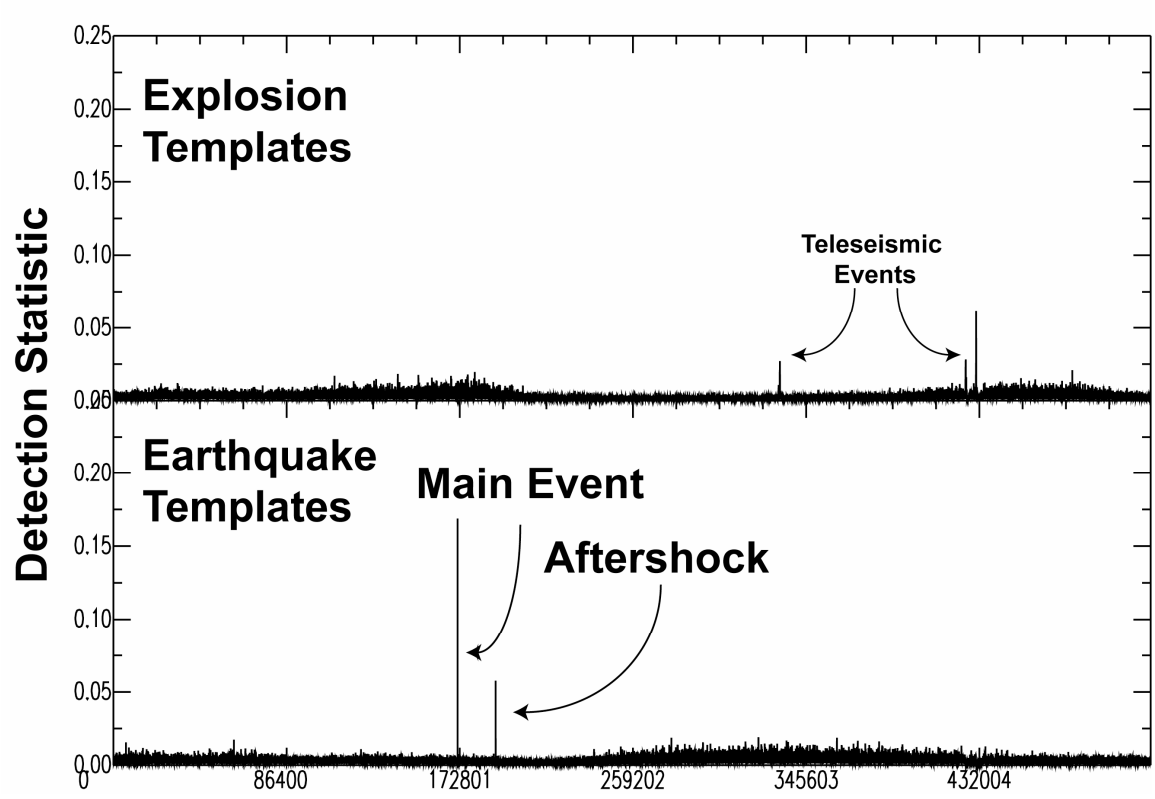


Figure 6.3.5 *Detection statistic for subspace detection with a four-station three-component network using template signals for (top) an earthquake and (bottom) ashallow explosion. The detection statistic is shown for six days.*

When the explosion templates are used they fail to give a high value of the detection statistic when the earthquake and aftershocks occur. They do however indicate high detection statistic values for the teleseismic events. This is probably due to the spurrious high correlation of an impulsive P-wave from the explosion source at regional distance with the teleseismic event. In realistic monitoring situations large teleseismic events will be easily detected by other means, such as a global seismic network.

7. Conclusions and Recommendations

In this study we have shown how Model-Based Signal Processing can improve seismic monitoring. Template signals computed from 3D models with structural variability can be used to correlate with observed signals using the subspace detection method. We found it difficult to compute template signals in fully 3D models with adequate bandwidth for the current monitoring challenges. However, the reduction of 3D models to 1D path-average models allowed us to greatly increase the bandwidth of the model-based template signals. This shows promise to impact spotlight detection of events in regions of interest.

We suggest a two-pronged strategy for advancing this technology: 1) exploitation of increasingly realistic 3D stochastic earth models will improve correlation of theoretical templates with event signals; and 2) exploitation of larger networks of seismic stations to reject background events that from target source(s). Both of these strategies take advantage of current trends in the broader seismology community to improve models, deploy ever-greater numbers of seismic stations and take advantage of a progressive drop in computational costs. The results of this study established the framework for Model-Based Signal Processing by showing that theoretical signals can be used in subspace detection and that network processing improves network detection performance.

While we did not fulfill all the accomplishments we had hoped to in our original project plan, we did establish that Model-Based Signal-Processing can detect small seismic

events and this technology will benefit from improved seismic models, densification of seismic networks and more powerful computers.

This project has enabled the development of important capabilities for seismic monitoring research, namely:

- computation of synthetic seismograms in 3D models
- familiarization of the PI with high-performance computing;
- specification of 3D models into synthetic seismogram codes;
- development of software for generating realistic stochastic heterogeneity;
- further improvement of MCMC models of the Korean Peninsula.

These capabilities have provided improved visibility of the LLNL GNEM Program as taking a forward looking approach to nuclear explosion monitoring.

8. Project Publications and Highlights

During the course of this LDRD project we actively presented results at scientific and programmatic workshops and conferences. Among the highlights are a briefing to NNSA/NA-22 Headquarters in Washington D.C. and presentation the annual nuclear explosion monitoring meeting (the Seismic Research Review). We gave presentations at national meetings of the Seismological Society of America (SSA) and the American Geophysical Union (AGU) Meeting. At two AGU meetings we gave oral presentations during special sessions on seismological applications of advanced signal-processing. We also presented elements of this effort at the Energy and Environment Directorate External Advisory Committee (March 2005) and the Directorate Review Committee Chairs (June 2006).

Harris, D., A. Rodgers, M. Pasyanos, S. Myers, J. Levatin, A. Franz and J. Tromp (2006). Trends in detection in earthquake seismology, presentation at the 151st Meeting of the Acoustic Society of America, June 6, 2006, UCRL-PRES-222128.

Rodgers, A. and J. Tromp (2005). Simulations of Seismic Wave Propagation with the Spectral Element Method and High-Performance Computing, poster presentation at the

Energy and Environment Directorate External Advisory Committee, March 15, 2005, LLNL, UCRL-POST-210540.

Rodgers, A. and J. Tromp (2005). Modeling Nuclear Explosions and Earthquakes with the Spectral Element Method and High-Performance Computing, abstract and poster presented at the Annual Seismological Society of America Meeting, April 27-29, Incline Village, NV, UCRL-ABS-208922.

Rodgers, A. and J. Tromp (2005). What Will USArray Data Look Like?: Spectral Element Simulations on LLNL High Performance Computers, poster at Incorporated Reseacrh Institutions for Seismology (IRIS) Workshop, June 8, 2005, Dolce Skamania Lodge, Stevenson, WA, UCRL-ABS-211044.

Rodgers, A. (2005). Computational Seismology at LLNL: A National Lab Perspective, Joint Computational Infrastructure for Geodynamics (CIG) and Incorporated Reseacrh Institutions for Seismology (IRIS) Workshop, June 8, 2005, Dolce Skamania Lodge, Stevenson, WA, UCRL-PRES-212698.

Rodgers, A. (2005). Simulations of Seismic Waves on Livermore Computers with the Spectral Element Method, presentation to the Livermore Computing Monthly Users Meeting, June 14, 2005, LLNL, UCRL-PRES-213089.

Rodgers, A.,, D. Harris, J. Levatin, M. Pasyanos, S. Myers and J. Tromp (2005). A Model-Based Signal-Processing Approach to Seismic Monitoring: Stochastic Earth Models, Spectral Element Method Synthetics and Coherent Signal Processing, abstract and poster presented at the 2nd SPICE Workshop, September 4-10, Smolenice, Slovakia, UCRL-ABS-213711.

Rodgers, A, D. Harris, S. Ford and M. Pasyanos, (2005). A Model-Based Signal Processing Approach to Seismic Monitoring, abstract S32A-06, 2005 Fall American Geophysical Union Meeting, San Francisco, CA, UCRL-ABS-214846.

Rodgers, A., D. Harris and M. Pasyanos (2006). A Model-Based Signal-Processing Approach to Nuclear Explosion Monitoring, abstract and poster presented at the 28th Annual Seismic Research Review, UCRL-ABS-222007, UCRL-CONF-222691

Rodgers, A., D. Harris and M. Pasyanos (2006). Detection of Seismic Events with Model-Based Signal-Processing, abstract, extended abstract and poster presentation at the Center for Advanced Signal and Image Sciences (CASIS) Workshop, November 16-17, LLNL, UCRL-ABS-225122 and UCRL-PRES-226177.

Rodgers, A., D. Harris, M. Pasyanos, S. Blair and R. Matt (2006). Model-Based Signal Processing: Correlation Detection With Synthetic Seismograms, abstract S11B-01, 2006 Fall American Geophysical Union Meeting, San Francisco, CA, UCRL-CONF-224073.

Thorne, M., S. Myers, D. Harris and A. Rodgers (2006). Finite Difference Simulation of Seismic Scattering in Random Media Generated With the Karhunen-Loeve Transform, abstract S51E-05, 2006 Fall American Geophysical Union Meeting, San Francisco, CA, UCRL-CONF-224171.

9. Acknowledgements

This project was supported by the Laboratory Directed Research and Development (LDRD) Laboratory Science and Technology Office (LSTO), project 05-ERD-019. Several people contributed to this effort, described in detail below.

JoAnne Levatin (CAR, Computation Directorate) provided essential assistance on high-performance computing and development of computer codes for this project. She helped modify the SEM code to handle the MCMC models for our template waveforms. She also developed code to generate model files for the WPP code.

Jeroen Tromp (California Institute of Technology) provided the open source SEM code, SPECFEM3D and assistance with our applications.

Sean Ford (Arizona State University) was a summer student in July-August 2005. Sean wrote a prototype sub-space program in MATLAB.

Stephen Myers (GNEM Program, LLNL) provided his codes for the Karhunen-Loueve method and helped Mike Thorne understand the method.

Michael Thorne (Arizona State University) was also a summer student in July-August 2005. Mike wrote a code to generate stochastic fields with a variety of correlation functions using the Fourier method. He also worked with Steve Myers to develop a code for generating stochastic fields with the Karhunen-Loueve method. Mike continued to work on this after graduating in late 2005 as a postdoc at the university of Alaska Fairbanks. He presented this at the Fall 2006 AGU meeting.

Al Franz and Robert Matt (CAR, Computation Directorate) worked with Mike Pasyanos and ran MCMC models.

Katheleen McCandless (CAR, Computation Directorate) worked with JoAnne Levatin to build formatted MCMC models and run WPP synthetic seismograms.

Timothy Paik wrote the sub-space detector design program while a summer student at LLNL.

Jay Zucca provided moral support and guidance as well as helped to formulate a compelling proposal in the conception of the project.

High-performance computing resources were provided by Livermore Computing. We thank David Dannenberg and Greg Tomaschke for assistance with our LDRD and DAT allocations on MCR.

This work was performed under the auspices of the U.S. Department of Energy by University of California, Lawrence Livermore National Laboratory under Contract W-7405-Eng-48.

10. References

Frankel A. (1989). A review of numerical experiments on seismic wave scattering, *Pageoph* **131**, 639-685.

Frankel, A. and R. Clayton (1986). Finite difference simulations of seismic scattering: Implications for the propagation of short-period seismic waves in the crusta and models of crustal heterogeneity, *J. Geophys. Res*., 91, 6465-6489.

Gropp, W., E. Lusk and A. Skjellum (1994). Using MPI: Portable Parallel Programming with the Message-Passing Interface, MIT Press, Cambridge, MA, 307 pp.

Harris, D. (1989). Characterizing source regions with signal subspace methods: theory and computational methods, Lawrence Livermore National Laboratory, Internal Document UCID-02148, Livermore, CA, 41 pp.

Harris, D. (1991). A waveform correlation method for identifying quarry explosions, *Bull. Seism. Soc. Am.*, 81, 2395-24187.

Harris, D. (1997). Waveform correlation methods for identifying populations of calibration events, Proceedings of the 19th Annual Seismic Research Symposium on Monitoring a CTBT, September 23-25, 1997, 604-614.

Harris, D., A. Rodgers, M. Pasyanos, S. Myers, J. Levatin, A. Franz and J. Tromp (2006). Trends in detection in earthquake seismology, presentation at the 151st Meeting of the Acoustic Society of America, June 6, 2006, UCRL-PRES-222128.

Harris and Paik (2006).

Julia, J., C.J. Ammon, R.B. Herrmann, and A.M. Correig, Joint inversion of receiver function and surface wave dispersion observations, *Geophys. J. Int.,* 143, 1-19, 2000.

Kennett B.L.N. (1983), Seismic Wave Propagation in Stratified Media, pp. 342, Cambridge University Press.

Komatitsch, D. and J.-P. Vilotte (1998). The Spectral Element Method: An efficient tool to simulate the seismic response of 2D and 3D geologic structures, *Bull. Seism Soc. Am.*, 88, 368-392.

Komatitsch, D. and J. Tromp (1999). Introduction to the spectral element method for three-dimensional seismic wave propagation, *Geophys. J. Int.*, 139, 806-822.

Komatitsch, D., J. Ritsema and J. Tromp (2002). The Spectral-Element Method, Beowolf computing and global seismology, *Science*, 298, 1737-1742.

Levander, A., and K. Holliger (1992). Small-scale heterogeneity and large-sclae velocity structure of the continental crust, *J. Geophys. Res.*, 97, 8797-8804.

Nilsson, S., N. A. Petersson, B. Sjogreen, H.-O. Kreiss (2006). Stable difference approximations for the elastic wave equation in second order formulation, *SIAM J. Numer. Anal.* UCRL-JRNL-222276, reviewed and revised manuscript.

Pasyanos, M., G. Franz and A. Ramirez (2004). Stochastic inversion of seismic structure of the crust and upper mantle using the Markov Chain Monte Carlo approach, technical report, UCRL-TR-204963, Lawrence Livermore National Laboratory.

Pullammananappallil, S., A. Levander and S. Larkin (1997). Estimation of crustal stochastic parameters from seismic exploration data, *J. Geophys. Res.*, 15,269-15,286.

Randall, George E. (1994). Efficient Calculation of Complete Differential Seismograms for Laterally Homogeneous Earth Models, *Geophy. J. Int.*, 118, 245-254.

Rodgers, A, D. Harris, S. Ford and M. Pasyanos, (2005). A Model-Based Signal Processing Approach to Seismic Monitoring, abstract S32A-06, 2005 Fall American Geophysical Union Meeting, San Francisco, CA.

Rodgers, A., D. Harris, M. Pasyanos, S. Blair and R. Matt (2006). Model-Based Signal Processing: Correlation Detection With Synthetic Seismograms, abstract S11B-01, 2006 Fall American Geophysical Union Meeting, San Francisco, CA.

Shapiro, N.M. and M.H. Ritzwoller (2002). Monte-Carlo inversion for a global shear velocity model of the crust and upper mantle, Geophys. J. Int., 151, 88-105.

Thorne, M., S. Myers, D. Harris and A. Rodgers (2006). Finite Difference Simulation of Seismic Scattering in Random Media Generated With the Karhunen-Loeve Transform, abstract S51E-05, 2006 Fall American Geophysical Union Meeting, San Francisco, CA.

Wagner, G. (1996). Numerical simulations of wave propagation in heterogeneous wave guides with implications for regional wave propagation and the nature of lithospheric heterogeneity, *Bull. Seismo. Soc. Am.*, 86, 1200-1206.

Appendix. User Guide to LLNL's Subspace Detector Software

There are two codes for performing subspace detection: a code for designing a subspace detector and another for running the detector on a continuous data stream. The algorithms for these codes are described in Harris and Paik (2006). This User's Guide gives a brief summary of how the codes work. It is expected that the user have some knowledge of waveform correlation and subspace detection.

Subspace Detector Design

The first code, called `subspaceDesigner`, designs the subspace detector. It reads in a set of template seismograms that form the basis for creating a subspace detector, performs a number of calculations with the templates and outputs two files used by the detector. The templates may be multichannel waveforms from multiple stations. The code is written in JAVA. To run the code on the LLNL GNEM LAN, simply enter:

```
s34> subspaceDesigner
```

This will run the script `/usr/local/bin/subspaceDesigner` and run the JAVA code. Note that /usr/local/bin must be in your `PATH` environment variable. Commands are executed from a Graphical User Interface (GUI)

The code has four processing steps. It starts by prompting if you want to run on single or multiple events (Figure A.1). Execute your choice by clicking on the text for either: "Single event" or "Multiple events".

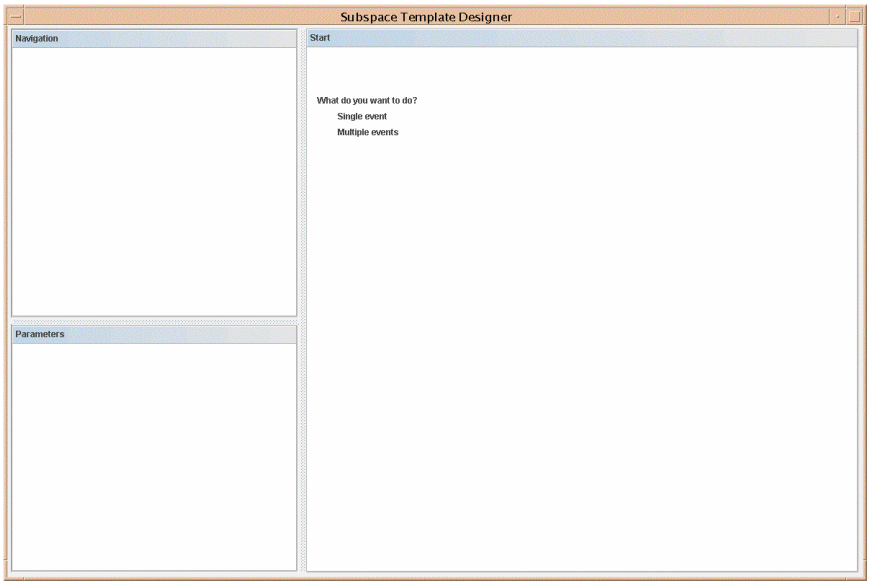


Figure A.1 *SubspaceDesigner GUI after launching the code. This is the main window used in all stages of the detector design. At this point the user must choose between single or multiple events.*

Regarding Data Format(s)

The template seismograms must be formatted in CSS or NDC format, and include at a minimum wfdisc and wftag tables. They may optionally include site and origin tables.

The wfdisc records must have unique wfid's and the `dir` and `dfile` fields must point to the waveform files. A partial listing of a valid wfdisc table has the following fields:

sta	chan	time	wfid……..	dir	dfile
INCN	BHE	947634236.70000	0	MODEL_0	INCN.BHE
INCN	BHN	947634236.70000	1	MODEL_0	INCN.BHN
INCN	BHZ	947634236.70000	2	MODEL_0	INCN.BHZ
BJT	BHE	947634236.70000	3	MODEL_0	INCN.BHE
BJT	BHN	947634236.70000	4	MODEL_0	INCN.BHN
BJT	BHZ	947634236.70000	5	MODEL_0	INCN.BHZ

The wftag table provides a link between the `wfid` and the `evid`, that is all the waveforms must be associated to a single event through the wftag table. The wftag file should have the following fields:

tagname	tagid	wfid	lddate
evid	0	0	2006/11/11
evid	0	1	2006/11/11
evid	0	2	2006/11/11
evid	1	3	2006/11/11
evid	1	4	2006/11/11
evid	1	5	2006/11/11

.
.

Event Editing Stage

The main window for the GUI will appear as in Figure A.2. There are three panels. The panel in the upper left marked "Navigation" indicates the stages and allows the user to jump to different stages. The panel on the lower left contains the dialogs for the current stage. The large panel on the right side will contain tables and plots generated by the processing at each stage.

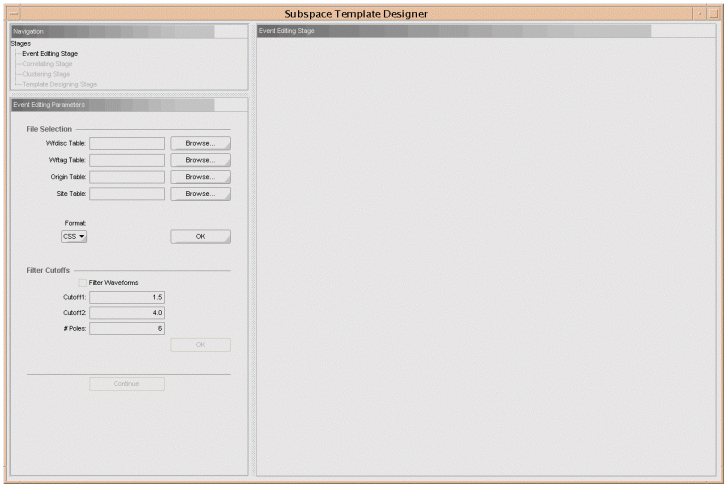


Figure A.2 *Event Editing Stage. The user must select the files for the template waveforms.*

The first stage defines and reads in the template waveforms, allows the user to window and filter the data. To select the data, type the wfdisc and wftag files names, or alternatively, click the "Browse" button to locate and select the files. Note that the data format must be defined correctly using the pull-down menu marked "Format:". Optionally the user can define the origin and site tables, similar to the wfdisc and wftag tables (Figure A.2).

After selecting the files and format for the template waveforms the data are read by clicking the "OK" button directly below the File Selection dialog. The code will write to the standard output "About to populate waveforms". The code will then display a table of channels (Channel View) with station and channel listed from the selected wfdisc table. Alternatively the user can view the events (Event View). The user can toggle between Channel and Event views by clicking the tabs on the lower center of the main window. Waveforms can be viewed by clicking the "View" button on the right side of the Channel or Event view.

It is recommended to view the data by event and include all station-channel data for one event. This way the use can view the full duration of the signals and define time windows to include the signals of interest. The menu in the upper right corner of the waveform viewing window allows the user to define how many waveforms are displayed.

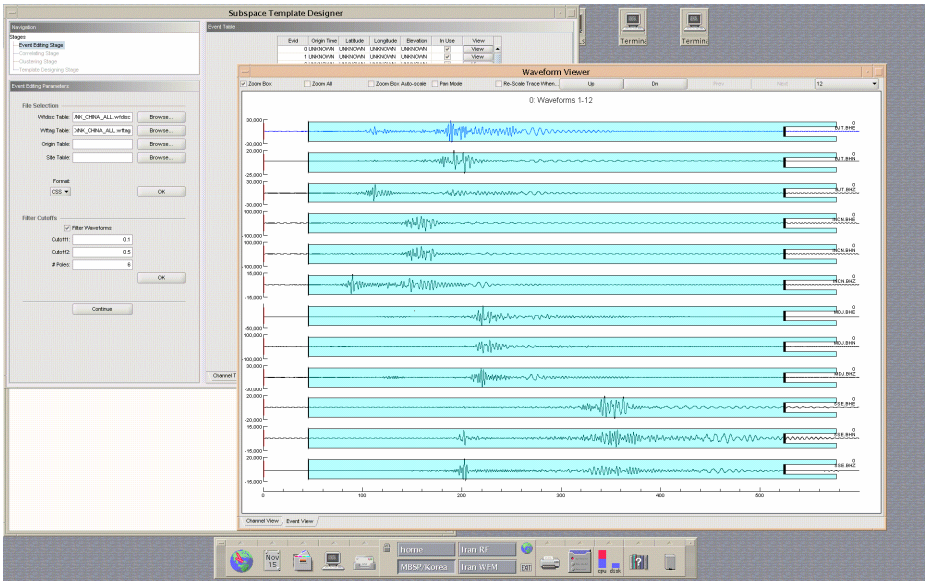


Figure A.3 *Event Editing Stage, showing filtered and windowed waveforms in the waveform viewer. The file and filter selections can be seen in the text boxes on the left side of the main window.*

At this point the user should enter the filter parameters, if filtering is desired. First, enter the low-, high-frequency (in Hz) cuts and number of poles for the filter in the text boxes for "Cutoff 1", "Cutoff2:" and "# Poles:" respectively. At this point if filtering is

desired, click the radio button marked "Filter Waveforms" above the filter parameter text boxes and then click the button marked "OK" to the lower right of the filter parameter text boxes. The filtered waveforms will appear in the waveform viewer at this point (Figure A.3).

The user must select the time window for all waveforms by sliding the cyan window to the desired start and ending times to capture the signals of interest.

At this point the user can proceed to the next stage by clicking the button marked "Continue" in the lower part of the Event Editing Dialog.

Correlating Stage

The Correlating Stage computes the pairwise waveform correlations. Before the correlations are computed the used can define the windows in the text boxes marked "Offset:" and "Window Length:". The offset is the time from the beginning of the waveforms and the window length is the window duration, both in seconds. To set these, type the number the text box and click the button marked "Set" one-at-a-time. Next, click the button marked "Correlate". This will initiate the calculation of (filtered) waveform correlations for all pairs of template seismograms and can take some time. When the calculations are completed the tool will display a histogram of all correlation values in the results panel (Figure A.4).

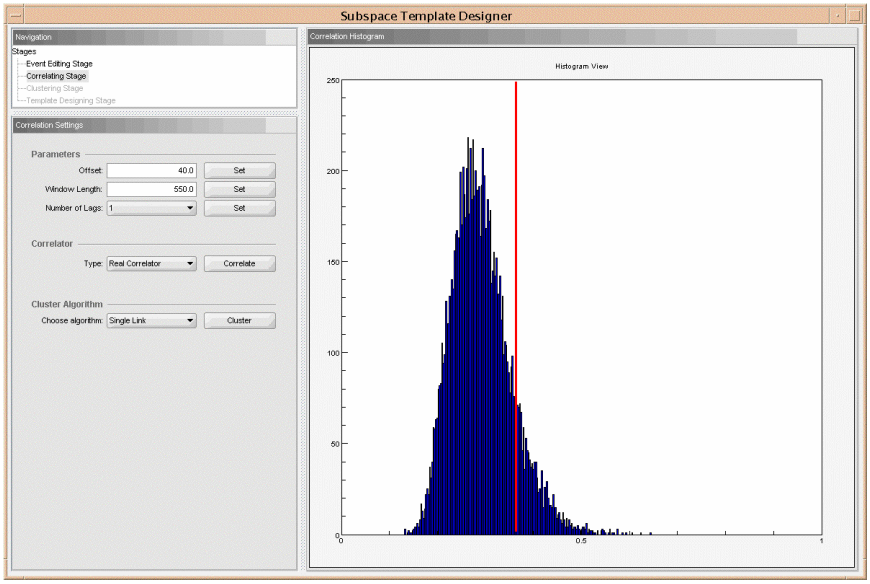


Figure A.4 *Correlating Stage. This figure shows the histogram of correlation values after correlating a large set of template waveforms.*

The tool first displays the histogram of correlation values with a vertical bar at the value 0.5. The user then can click on this bar and slide it to the left or right to define a threshold correlation value for clustering. In the case shown in Figure 4, the user has selected a value near 0.37. This means that only pairs of waveforms that correlate with

cross-correlation values of 0.37 or more will be considered in the Clustering Stage (next). There is a menu for choosing the clustering algorithm. Currently only the single-link clustering algorithm is available.

To proceed to the next stage, click the button marked "Cluster".

Clustering Stage

After clinking the "Cluster" button in the Correlating Stage the tool very quickly plots a dendrogram of the single-link cluster results (Figure A.5). A dendrogram-like output is written to the standard output.

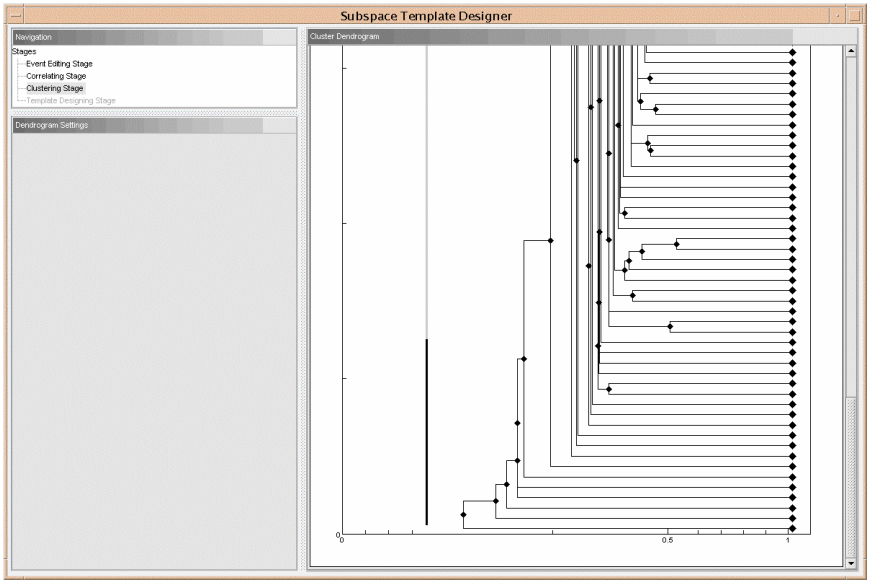


Figure A.5 *Clustering Stage. This figure sows the state of the tool after the dendrogram has been computed.*

At this point the user must select a cluster by clicking on one of the events. This is done by hovering the mouse over one of the black diamonds on the right side of the dendrogram. When the cluster is selected it will change color to blue and a waveform viewing window will be launched. This will take the tool to the next stage for designing the template.

Template Designing Stage

At this point the tool is nearly ready to create a subspace detector for the selected cluster. But before the detector files can be written a few steps must be completed. In the waveform viewer window, the user should select the "Event View" tab on the lower left side. This will display the waveforms for a single event. Be sure to select the number of waveforms to be the total number of channels in order to view all waveforms for the

event. Then select the window for all waveforms by sliding the cyan window to the desired start and ending times to capture the signals of interest.

The dialog for this stage allows the user to set the "False Alarm Rate" and "Degrees of Freedom". For cases with multiple stations, the tool reports the Degrees of Freedom (DOF) to be too small by 3/2 times the number of channels. The correct DOF should be the time-bandwidth product times the number of channels. The tool reports a smaller DOF because it assumes there are arrays which have redundancy and reduce the DOF.

The user can set the DOF in the text box marked "Degrees of Freedom" and click the button marked "Set".

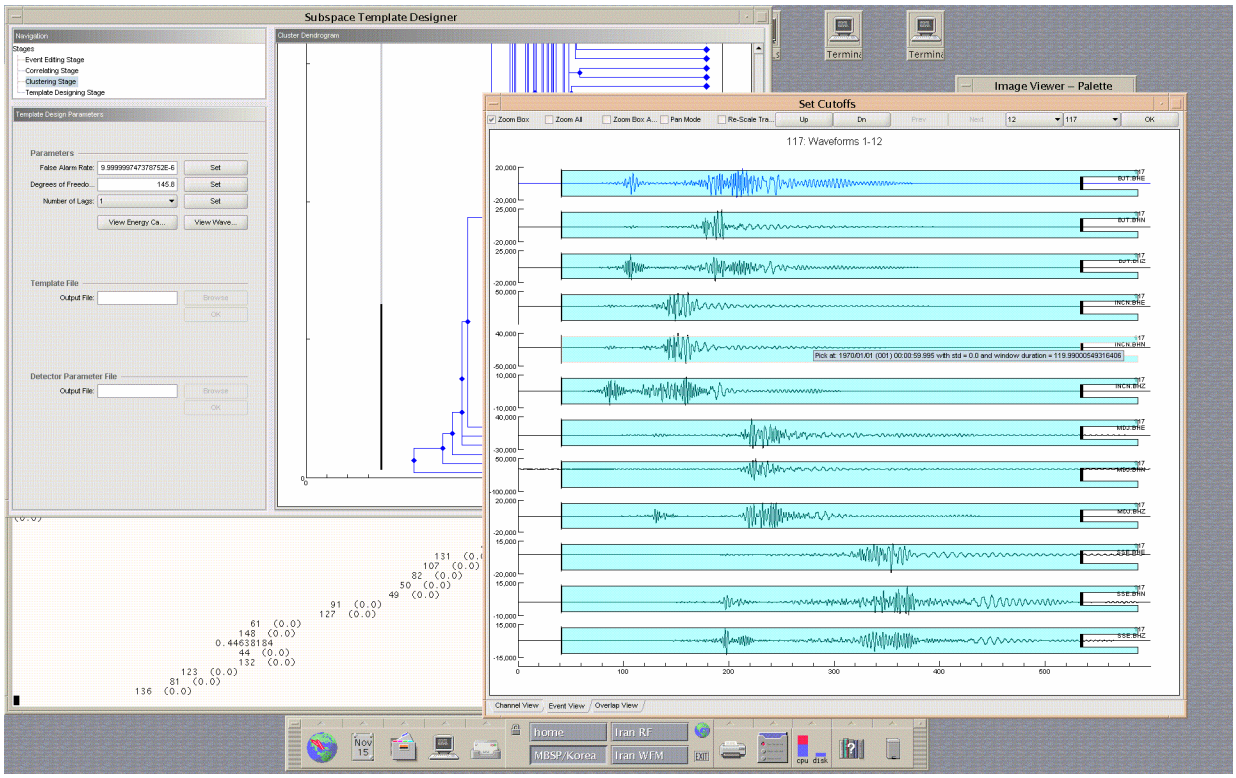


Figure A.6 *Template Designing Stage. Waveforms for the selected cluster are viewed and time windows are defined in the waveform viewing window launched after selecting the cluster.*

After the time window is selected, the user must click the button marked "OK" in the upper right corner of the waveform viewing window (Figure A.6). This defines the template waveform that will be used to form the matrix of basis templates.

The tool will generate a family of curves for the probability of detection at a fixed false alarm rate as a function of post-integration signal-to-noise ratio using a subspace dimension of 1-N, where N is the total number of template waveforms (Figure A.7). The user can also display the energy capture by clicking the button marked "View Energy Capture" in the Template Design Parameters dialog. The energy capture is displayed

in a separate window. This plot shows a family of curves illustrating how each of the N basis templates is modeled by a subspace eigenvectors of dimension n. The plot demonstrates that when n increases toward the total number of basis waveforms, N, more of the basis signal can be projected onto the subspace. The energy capture window allows the user to select the subspace dimension by sliding the black vertical bar to the desired dimension. The bar starts at dimension equal to zero and turns red when selected.

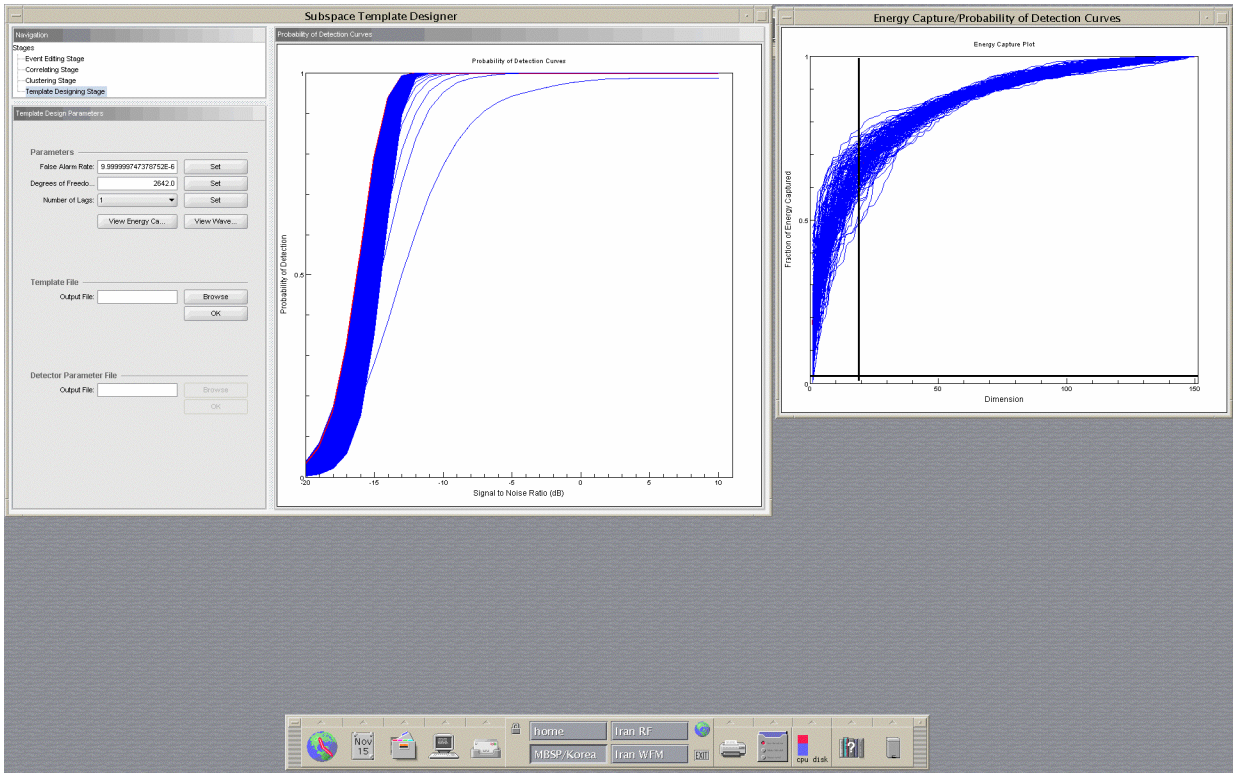


Figure A.7 *Template Designing Stage. The main window shows the Template Design Parameters and displays the Probability of Detection curves. The Energy Capture is displayed in another window. In this case, the user has selected a subspace dimension of 20.*

Finally, the user must write out two files that will define the subspace detector for use by the second (subspace detector) code. This is done by typing the file name or browsing (with the "Browse" button) in the lower part of the Template Design Parameters panel. The Template File is a large binary file containing the subspace waveforms in a compact format and is conventionally given a suffix of ".def". The Detector Parameter file is a small ASCII file with parameters and is conventionally given a suffix of ".par".

The parameter file looks like:

```
s34> cat all_dim20.par
#This is a parameter file
#Wed Nov 15 11:56:25 PST 2006
decrate=5
data-output-base-directory=./results
stations=BJT BJT BJT INCN INCN INCN MDJ MDJ MDJ SSE SSE SSE
```

```
fhi=0.5
channels=BHE BHN BHZ BHE BHN BHZ BHE BHN BHZ BHE BHN BHZ
ford=6
SubspaceDetector1=SSD1 /datax/MBSP/NKCHINA/Continuous/ALL/all_dim20.def 0.021565259
SubspaceDetector1
modfreq=0.30000000074505806
flo=0.10000000149011612
format=CSS
ftype=BP
delta=0.05000000075000001
```

Some of these parameters are fairly obvious, such as the low and high frequencies for the filter, the filter order, decimation rate, stations and channels, etc…

Subspace Detector

After a subspace detector has been designed by the `subspaceDesigner` code, the detector can be run on a continuous data stream with a simple command line. There are two JAVA jar files: `Detector.jar` and `Jampack.jar`. These can be placed in a directory and referenced in command line or with a C-shell script. The detector takes two arguments the parameter file and the CSS wfdisc file for the continuous data. The continuous data only requires a wfdisc table. Be sure that the continuous data have the same sample rate as the template waveforms. The detector writes the output (SAC file of the detection statistic and ASCII detections) to a sub-directory called `results`.

```
s34> cat run_detector
#!/bin/csh

# Check for results directory, make it if it does not exist
if ( ! -d results ) then
  mkdir results
endif

set JDIR = /datax/MBSP/JAVA
set PAR_FILE = all_dim20.par
set CONTDB = ALL_cont.wfdisc

/opt/java/jdk1.5.0/bin/java -cp $JAVADIR/Detector.jar:$JAVADIR/Jampack.jar llnl.gnem.
apps.subspace.detection.CmdLineProcessor $PAR_FILE $CONTDB
```

The code will read in the parameter (.par) and subspace (.def) files and then start streaming the continuous data and computing the detection statistic. When the detection statistic exceeds the threshold, determined by the false alarm rate, DOF and subspace dimension, it will declare a detection and write the time out to an ASCII file called null.detList. The detection statistic is written to a binary SAC file and can be viewed in SAC as the detector is running. When the detector has run on the complete continuous data stream it reports

Also, the SAC file of the detection statistic is re-written to have the correct absolute timing. To determine the absolute times of detections one must account for the Finite Impulse Response nature of the subspace detector.