A SOFTWARE TOOLBOX FOR SYSTEMATIC EVALUATION OF SEISMOMETER-DIGITIZER SYSTEM RESPONSES

Jessie L. Bonner[1], Robert B. Herrmann[2], Anastasia F. Stroujkova[1], and Mark R. Leidig[1]

Weston Geophysical Corporation[1] and St. Louis University[2]

Sponsored by the National Nuclear Security Administration

Award No. DE-FG02-09ER85548/Phase_I

ABSTRACT

Measurement of the absolute amplitudes of a seismic signal requires accurate knowledge of the seismometer-digitizer response. When the seismometer and digitizer characteristics are known, deconvolution of the digital system response is relatively straightforward. However, even with a known response, problems in timing and incorrect polarities can often contaminate any calibration parameters derived from these data. When either one or both components of the response for a seismometer system are unknown, the data are often not considered suitable for seismological tasks that require absolute amplitude measurements.

We have developed a set of software tools that can recover the sensor/digitizer response function from raw data when the response is either unknown or incorrectly known. These software tools include:

- **SACPSD**: a power spectral density (PSD) estimator for background noise spectra at a seismic station. SACPSD differs from the current PSD used by NEIC and IRIS in that the smoothed spectra are derived by smoothing over linear frequencies rather than logarithmic periods, resulting in smoothed PSD that tracks the unsmoothed PSD rather than exhibiting a bias where the raw PSD drops rapidly with increasing period. SACPSD is fast and can be used routinely to monitor network performance and metadata validity,

- **NOISETRAN**: a prototype method of estimating the broadband seismic system response for a seismic station based on the background noise at a nearby reference station,

- **GUIs/SCRIPTS**: we have developed Graphical User Interfaces (GUIs) and shell scripts that provide seismologists with easy access to the software in the system response recovery toolbox.

During Phase I of this Small Business Innovative Research (SBIR) project, we have also compiled an extensive database of seismic sensors and recorders that are in use or have been developed. Information on each instrument includes sensitivities, bit weights, and gains. The database also contains station information and response corrections for many stations actively or previously installed around the world. The information in this database provides another step in determining the proper response correction for an unknown station.

Our results indicate that our Phase I project demonstrated both technical feasibility and a high potential to achieve the GNEM R&D goals for system response recovery of important seismic datasets.

INTRODUCTION

For this Phase I SBIR FY2009 project, we began the development of a software toolbox to estimate, assess, and chronicle the seismometer and digitizer response from recorded seismic data. The specific technical objectives included developing a comprehensive database of known seismometer and digitizer responses. Next, we have assembled a set of seismic analysis tools (Figure 1) to iteratively recover the instrument response (either precisely or rough order-of-magnitude). We estimate a station response transfer function by fitting the seismic data at an unknown station to the background noise and recorded signals from nearby stations with known responses. We have also investigated tools for metadata validation and initiated development of an easy-to-use graphical-user-interface to assist the user in running the programs.

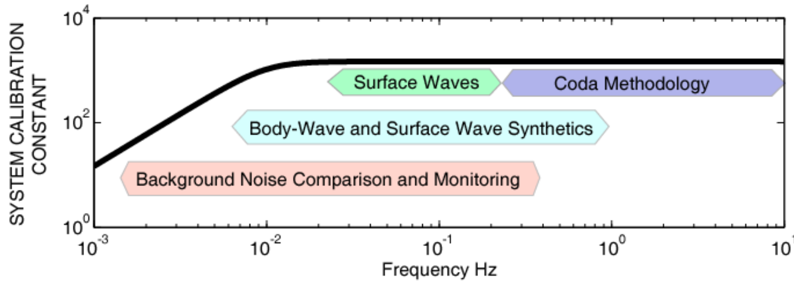

Figure 1. Schematic showing our techniques for recovering the seismometer-digitizer response for recorded seismic data. The methods include developing transfer functions using noise, surface wave and coda methods, developing synthetics for comparison to the observed data, and monitoring broadband background noise for timing, gain changes, and metadata discrepancies.

RESEARCH ACCOMPLISHED

Seismometer-Digitizer Database

During Phase I of the project, we compiled an extensive database of seismic sensors and recorders that are in use or have been developed. Information on each instrument includes sensitivities, bit weights, and gains. The database also contains station information and response corrections for many stations actively or previously installed around the world. The information in this database provides an initial step in determining the proper response correction for a station with unknown response.

The database consists of 400 seismometer, accelerometer, and infrasound sensors. Almost every sensor in the database includes the response sensitivity, with 107 instruments having more than one possible manufactured sensitivity. A more complete instrument response definition in the form of a pole/zero file is available for 21% of the sensors. There are 50 digitizers in the database, but only a small number of these are traditionally used for seismic recording. The bit weight needed to convert the digitized signal back to that output from the sensor was acquired for these most common recorders. Approximately half of the recorders can record at more than one gain setting.

The database contains varying amounts of information on over 16000 actively and previously deployed stations from around the world. Information includes the station name, location, dates of operation, instrumentation deployed, and instrument responses. A map of the 4016 stations with instrument response information is shown in Figure 2. These stations with known responses can be used to correct data recorded nearby with unknown responses. Examples of the amplitude responses for many of the United States National Seismic Network (USNSN) stations are shown in Figure 3.

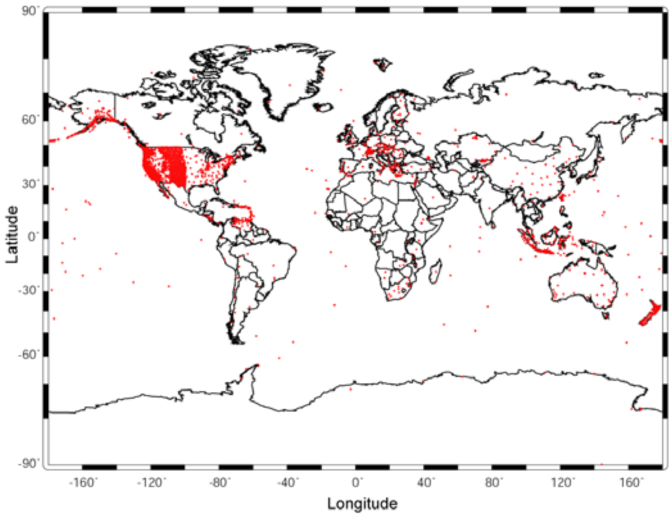

Figure 2. Map of pertinent stations (red dots) with instrument response correction information in the database.

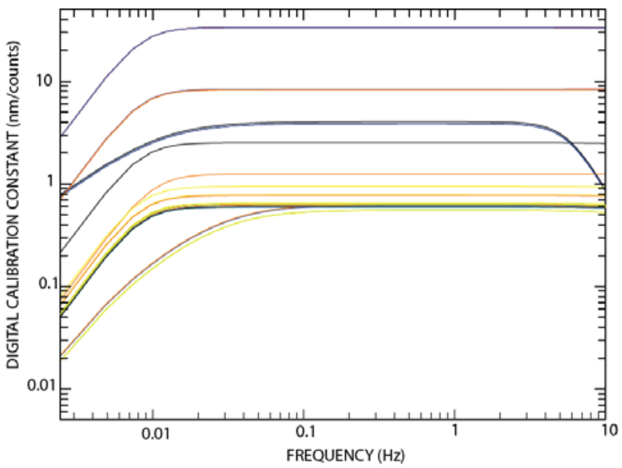

Figure 3. Seismometer-digitizer responses for United States National Seismic Network (USNSN) broadband stations (vertical component). There are 63 different stations plotted.

SACPSD

The determination of PSD of Earth noise is a useful tool for monitoring noise sources and for the performance estimation of seismic instruments (McNamara and Buland, 2004). As part of the Phase I project, we developed SACPSD, a GSAC program that calculates power spectral densities for seismic data. The program allows the use of routine PSD measurements as a tool for verifying the metadata for digital instrumentation and for checking on instrument performance. We also use SACPSD to help recover system responses based on noise estimates from a reference station—further discussed in the NOISETRAN section of this proposal.

The computation of power spectra follows Ifeachor and Jervis (1993). Given a time series x_k, we first correct that series to have a zero mean, apply a window, w_k, and correct amplitudes for the windowing function. Their equation (10.16) defines the corrected time series, s_k, as

$$s_k = c_2 w_k (x_k - c_1)\,, \quad (1)$$

where the constants are defined as

$$c_1 = \frac{\sum_{k=0}^{N-1} w_k x_k}{\sum_{k=0}^{N-1} w_k} \text{ and } c_2^2 = \frac{N}{\sum_{k=0}^{N-1} w_k^2}\,, \quad (2)$$

and N is the number of equally spaced observations, with sampling interval Δt. In SACPSD, the user may select either a Hanning window or a 10% sine taper for the windowing function.

Using an N-point discrete Fourier transform (DFT), the transform of s(k) is S(n), which is defined as

$$S_n = \sum_{k=1}^{N-1} s_k \exp(-j2\pi ft)\Delta t = \sum_{k=1}^{N-1} s_k \exp(-j2\pi nk/N)\Delta t \quad (3)$$

where n=0,...,N-1 , t = k Δt and f=n Δf. Because of the definitions of the forward and inverse DFT's, both s_k and S_n are periodic. The power spectrum density is defined as

$$PSD(f = n\Delta f) = 2S_n^2 / T \qquad (4)$$

where the time window $T = N\Delta t$. The factor 2 accounts for the negative frequency contribution.

For actual processing, the initial time series is assumed to consist of more than N=16384 points. Then using successive N point segments that overlap by N/2 points, the individual PSD's are summed and finally averaged by dividing by the number of segments, NSEG. The stacking and averaging reduces the variance in the PSD estimate, yielding a smoother PSD.

The estimate of the ground noise uses the frequency dependent modulus of acceleration sensitivity of the sensor, G (in counts/m/s^2) to give the noise estimate in decibels through the relation

$$N(f) = 10\log_{10} \frac{1}{NSEG} \sum_{i=1}^{NSEG} \frac{PSD_i}{G^2} \qquad (5)$$

For comparison with other noise-PSD algorithms, a smoothed noise spectrum is computed from

$$\frac{1}{NSEG} \sum_{i=1}^{NSEG} \frac{PSD_i}{G^2} \qquad (6)$$

by applying a simple smoother, currently a 5-point averaging operation in the equi-spaced frequency domain. Other smoothers may be implemented in the future. We note that McNamara and Buland (2004) apply a smoothing operation over period. Finally the smoothed noise spectrum is linearly interpolated to yield estimates at periods that are multiples of $2^{1/8}$ and then converted to decibels. To make the output independent of sampling rate, the multiples are always with respect to a period of 1.0 seconds. This smoothed version is always plotted as an overlay onto the N(f).

SACPSD is written in C and then easily transformable to other languages (e.g., JAVA or MATLAB). SACPSD is lean, easily adapted to field use and requires only that the acceleration sensitivity of the instrument be known. It is fast on laptops and could easily be used in the field for rapid estimation of noise power spectral densities (e.g., on noise studies for array siting).

Processing scripts have been developed to use SACPSD for a variety of different applications. We have used the method to examine errors and inconsistencies in metadata that accompany downloaded data from seismic networks. Similar analysis is performed routinely at NEIC and at IRIS to focus on individual channel performance with time. The purposes of the current implementation of SACPSD is a) to focus on individual network performance, since the actual noise field at certain periods should be the same for a network covering a relatively small geographical area and b) to note inconsistencies that may have to be addressed by the data/metadata generator or by the data/metadata servers. Figure 4 compares the noise PSDs for 568 vertical component traces. The overlay plot shows the high and low noise model limits (Peterson, 1993) and the individual station PSDs.

The color coding provides a means for identifying a curve. There are obviously a few outliers which can be examined in more detail by analyzing the SACPSD results of individual sub-networks.

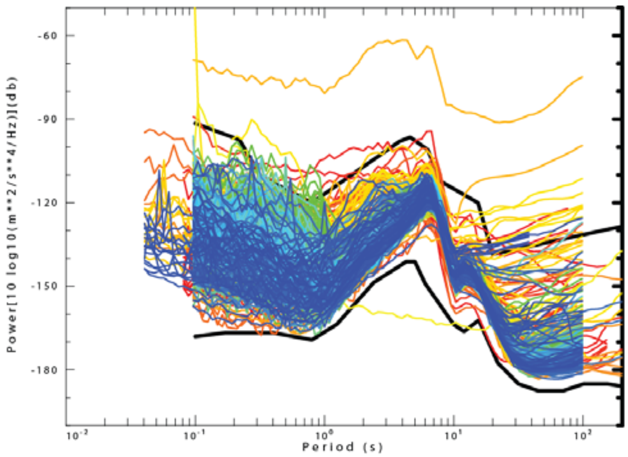

Figure 4. An overlay of 568 individual noise PSD's of USGS Network data based on the one-hour segment starting at 2010/02/17-01:00:00. The colors are used to distinguish individual estimates. The two black lines represent the new low and high noise models (Peterson, 1993). Note that the comparisons are a reflection solely of the waveform data at the USGS and of the existing metadata on the day that this comparison was run, e.g., February 21, 2010.

Figure 5 shows the result of running SACPSD for the New Madrid (NM) network in the central United States. For the NM network operators, the positive aspect of this plot is that all systems, except PBMO, show the same noise levels between 6 and 20 seconds. Clearly PBMO behaves differently having a gain of about 6db (a factor of 2) higher than the others. At long periods, the PBMO response differs greatly, indicating a sensor problem.

The short-period noise at PVMO (neon green in Figure 5) is very high, as is expected since this is a deep sediment site, just a short distance (200 meters) from a railroad. A visual examination of the hour-long segment shows that there is a significant noise burst, possibly from a train.

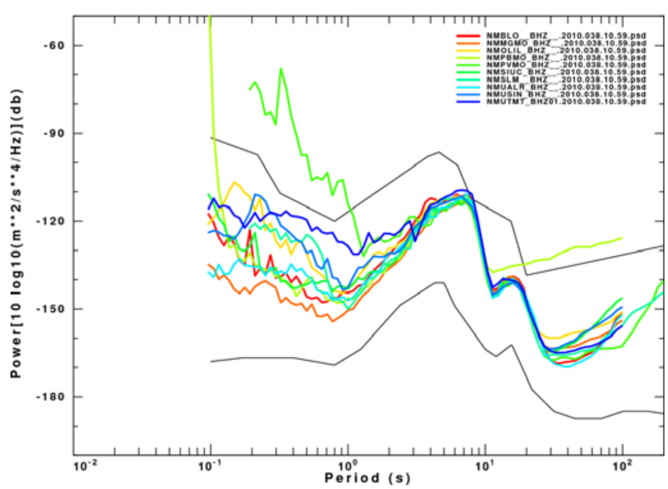

Figure 5. Result of running the SACPSD program for the BHZ components of the New Madrid network. The reason that different components cover different period ranges is due to the variable sample rates. PVMO has a rate of 20 Hz while the others have a rate of 40 Hz.

At a period of 0.1 sec, the curve for PBMO (light green in Figure 5) is very high. We know that this station has a sample rate of 40 Hz, but this behavior is indicative of using the response for a 20 Hz sampling interval which is indicated in the RESP file. This was not expected since the station was upgraded in November, 2009 to 40 Hz. We then verified that the correct metadata had not yet been loaded into the NEIC metadata server.

Results of applying SACPSD to the Cal Tech network data for February 21, 2010 is shown in Figure 6. Interestingly, the top five stations differ from the final four by ~12 db, which would translate into a factor of 4. A station map shows that this difference is not due to proximity to the ocean. We have yet to determine the reason behind this difference, but one suggestion is that the data are being sampled with Quanterra HR digitizers with a metadata response file for a non-HR digitizer.

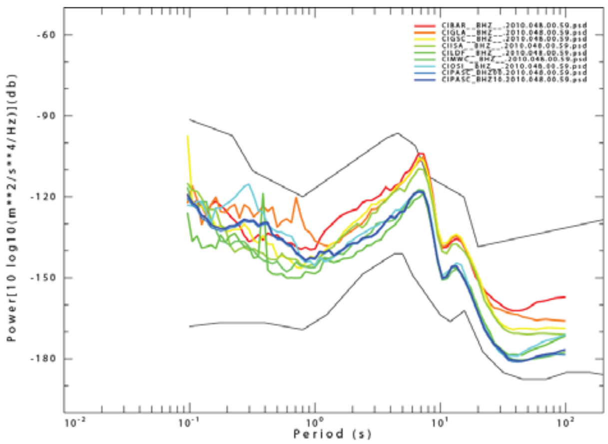

Figure 6. Result of running the SACPSD program for the BHZ components of the Cal Tech (USA) network showing two distinct sets of noise curves separated by ~12 db.

NOISETRAN: Noise Transfer Functions for Pseudo-Response Estimation

Important calibration datasets may be obtained in which no metadata accompany the waveforms. We are examining methods that attempt to determine a rough order-of-magnitude (ROM) estimate of the seismometer-digitizer response by considering the background seismic noise at a nearby seismic station with a known response (e.g., the "reference" station). We refer to this method as NOISETRAN, as we attempt to estimate a *transfer function* $T(\omega)$ to convert the background seismic *noise* at a station ($R_1(\omega)$) with unknown response (henceforth referred to as the "candidate" station) to the background noise levels at a reference station $R_2(\omega)$. The transfer function can be estimated using:

$$R_1(\omega)=R_2(\omega)*T(\omega). \quad (7)$$

Once the ROM system response $R_1(\omega)$ has been recovered, we can then use waveform modeling techniques to further refine the system response.

Our assumption for attempting the NOISETRAN technique is that the actual noise field at certain periods should be the same or similar for a network covering a relatively small geographical area. McNamara and Buland (2004) showed the geographical distribution of noise in three different period bands (e.g., 0.125-0.0625 sec, 8-4 sec, and 64-32 sec) in North America. They observed the strongest geographical variations at periods > 1 second due to cultural noise, especially near large population centers. Proximity to oceans dominates the variations in the microseism band (8-40 sec). They note that at longer periods (64-32 sec), the geographic variations decrease, and that the responses are more a function of the installation characteristics (e.g., borehole vs. surface installation) than the actual seismic noise characteristics. These results suggest that our best results of NOISETRAN will most likely occur at periods longer than 1 sec. We did not correct for possible noise sources during the Phase I project, but believe that our results suggest improvements could be made during the Phase II research based on noise prediction for the candidate stations.

NOISETRAN is initiated by manually selecting a reference station. The considerations for choosing a reference station should include: 1) a well-established background noise model for the reference stations (e.g., such as those provided by Berger *et al*. 2004; McNamara and Buland, 2004, and others), 2) proximity to the stations of interest, and 3) whenever possible, a similar noise background level between the reference and unknown stations.

The choice of a pre-determined background level at the reference station based on a yearly average was based on a Phase I objective that the data at the reference station and unknown station need not overlap. However, the noise stationarity becomes an issue in this case, because the noise PSD varies depending on the time of day, year, weather, etc. For example, the noise in the microseism band has significant annual variations (15-20 dB) according to McNamara and Buland (2004). In order to find an asymptotically unbiased estimate of the PSD one has to use very long time intervals (> 1 year).

If the shorter time interval is used, the PSD can be expressed as a sum of a mean value m_{PSD} and some quasi-periodic function d_{PSD}. If the seasonal variations are similar between the reference station and the unknown station, it may be possible to estimate the PSD for both stations during the same time intervals, which might minimize the error associated with these long-period variations. For the Phase II project, we will determine if NOISETRAN can be improved by calculating SACPSDs on data recorded from the same date and time at the reference and unknown stations; however, this procedure requires accurate knowledge of the system response at the reference station.

For the examples presented in the Phase I research, we chose to recover the ROM system response of a 19-station proprietary dataset from the Middle East obtained in 2008 by Weston Geophysical Corporation. Due to the sensitivity of these data, we do not provide a map of the candidate stations or the reference station, which was selected because it was the only open Global Seismographic Network (GSN) station with both a well-known seismometer-digitizer response and published background noise levels within 10 degrees of the proprietary seismic network. The reference station was located at distances between 183 and 612 km from the candidate stations. No metadata accompanied the initial delivery of the candidate waveform database to Weston Geophysical Corp, thus we did not know if the data were in counts or volts, velocity or acceleration. We chose to develop the NOISETRAN functions for this unknown network by comparing it to the 50^{th} percentile background noise levels at the reference station as estimated by Berger et al. (2004).

For the next stage of NOISETRAN, the SACPSDs for the raw data recorded by the candidate stations are estimated and compared to the noise PSD of the reference station (Figure 7). For this case, we have chosen the Berger et al. (2004) supplied noise curve for the reference station, which are typically provided at frequencies between 0.01 to 10 Hz. The spectral division between the candidate and the reference station yields the noise-based transfer function, which can be used to convert the unknown units to the units of the reference station. Because seismologists are often more familiar with the velocity response for many broadband instruments (see Figure 3), we convert the transfer functions to velocity and plot each NOISETRAN in Figure 7. Each of these curves can be used to convert the raw data to “approximate” units of velocity in m/sec. We emphasize that NOISETRAN will not provide the true ground velocity unless the noise between the reference and station of interest are exactly the same. We maintain that NOISETRAN can provide seismologists with a broadband ROM estimate of the response that may be improved with additional tools to be developed during the Phase II project.

We did know that the instruments that comprised this proprietary dataset were broadband instruments all from the same manufacturer (Guralp), thus the response for each should be very similar with only small differences in the calibration values. The NOISETRAN results provide somewhat conflicting information. At frequencies between 0.05 and 0.8 Hz, the scatter between the curves is reduced, and it would be possible to fit these curves reasonably well with a single instrument response. At frequencies above 1 Hz and below 0.05 Hz, the scatter between the transfer functions is increased and a single-response would seem implausible. These frequency ranges are somewhat in agreement with the geographical variations of noise as a function of frequency suggested by McNamara and Buland (2004). NOISETRAN seems to be providing reliable system-response recovery in the microseism bandwidth for these stations, which are all located at least 500 km from the nearest ocean. It is outside this band, where cultural noise and site effects are providing the increased scatter in the NOISETRAN method.

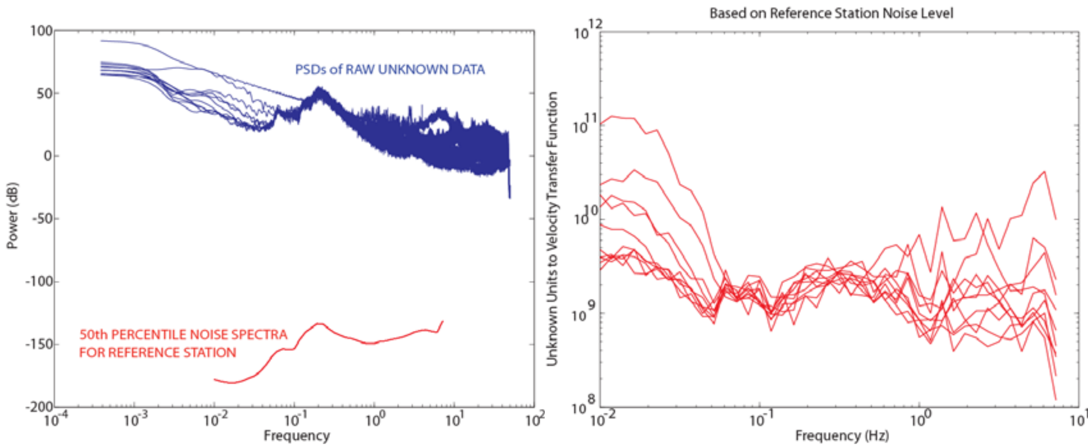

Figure 7. (Left) SACPSDs of raw data (blue) from a 2008 dataset that originally had no accompanying metadata. The red line is the 50% percentile noise estimate (Berger et al., 2004) at a reference station located between 183 and 612 km from the stations candidate stations. The reference data PSD are in units of acceleration while we did not know the units for the candidate stations. (Right) NOISETRAN-estimated transfer function that transforms the candidate data from unknown units to either acceleration or (in the case above) velocity (m/sec) by correcting the data to the 50% noise level at the reference station.

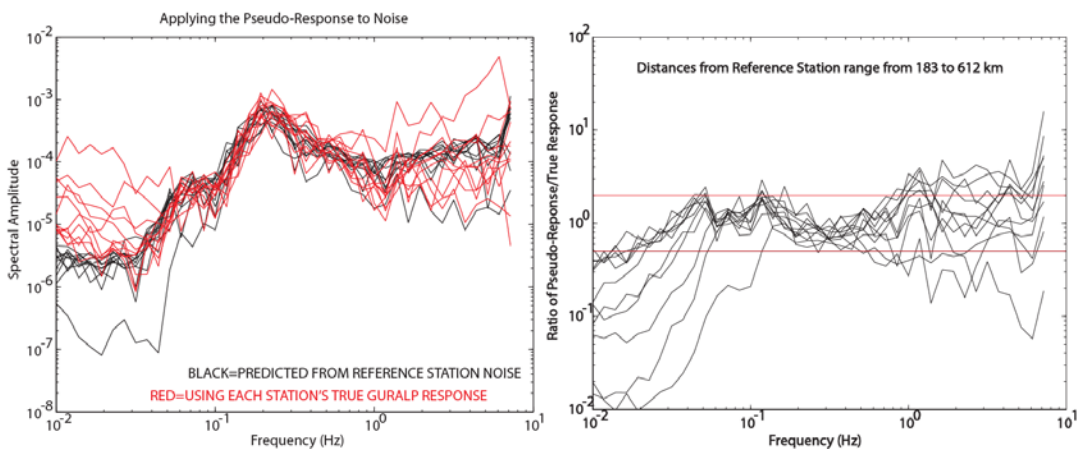

Figure 8. (Left) Comparison of our pseudo-response corrected data (black) to the true response calculations (red). We eventually obtained some metadata, including company-supplied PZ files, for the stations in this dataset. We corrected the data for the instrument response and then compared to our NOISETRAN-estimated responses obtained from noise levels at the reference station. The spectral ratios for the individual stations (pseudo-response/true-response) are shown at right.

Eventually, we were able to obtain nominal instrument response data for the candidate sensors from the manufacturer. We generated poles and zeros (PZ) files for each station and corrected the data for the true instrument response using the *transfer* command in GSAC (Herrmann, 2008). Figure 9 provides a comparison of background noise that has been converted to spectral amplitudes using NOISETRAN (black) and the true response (red). As suspected from Figure 7, NOISETRAN provides reliable estimates of the background noise at frequencies between 0.05 and 0.8 Hz and less reliable estimates outside this band. When spectral ratios between the NOISETRAN and true response estimates are formed (Figure 8), we note that most of the spectral ratios are within a factor of ±2 of an exact fit between the true response and NOISETRAN method across the entire frequency band. It is surprising and worth additional study as to why the worst fit is at frequencies less than 0.05 Hz. One thing to note is that the scatter in these spectral ratios would have been reduced more had we used the 90th percentile noise estimate for the reference station in the initial steps of NOISETRAN. In other words, it was a quiet day at the reference station at the time the candidate PSDs were estimated. This led us to postulate that the more appropriate method might be to use data from the reference station at the same time as the data from the candidate stations when forming the transfer functions. We hope to examine this further in Phase II.

We present Figure 9 to show how well NOISETRAN works when applied to actual earthquake signals recorded on the proprietary network. For regional recordings of this earthquake, we applied NOISETRAN and the true seismometer-digitizer responses to the signals. We then estimated a surface wave magnitude using data at periods between 8 and 40 seconds. We note only a difference of ~0.1 magnitude unit between the NOISETRAN-corrected and known pole-zero-corrected waveforms (e.g., 5.65 vs. 5.56). These results indicate technical feasibility and warrant further analysis and possible refinement during the Phase II project.

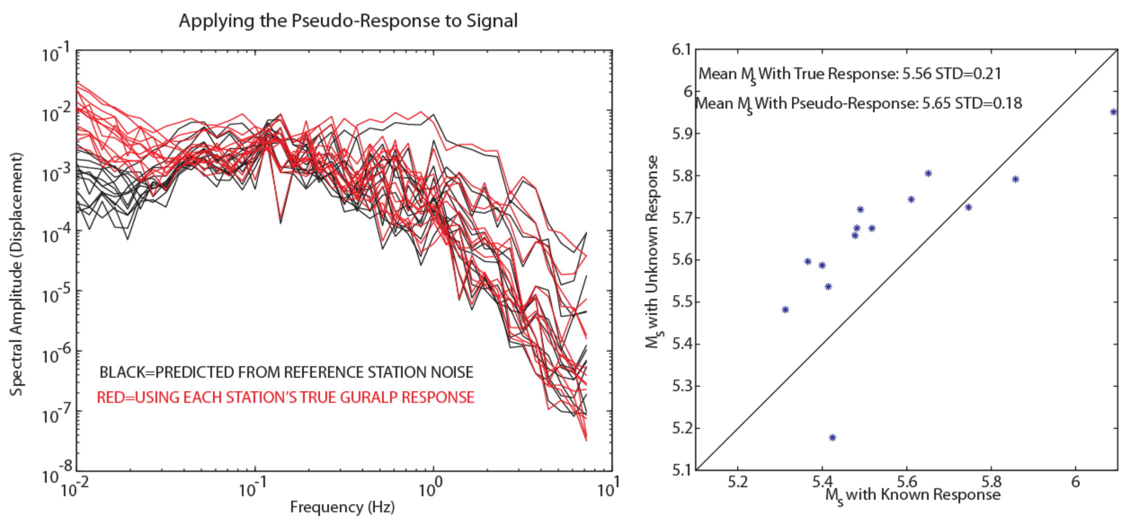

Figure 9. Results of applying the NOISETRAN response functions to an earthquake recorded on the proprietary network. The surface wave magnitude estimated using the NOISETRAN-estimated response is approximately 0.1 m.u. larger than the estimate from the true response.

Graphical User Interface

To facilitate the procedures required for an effective system for monitoring system response, Weston Geophysical Corp. has implemented an easy-to-use GUI in MATLAB. The Phase I toolbox currently includes 1) a database of known seismometer and digitizer responses and 2) methods for estimating response "transfer" functions between stations with and without known responses. The program is written in Matlab™ and uses SAC files as an input. The features implemented in the software package during the Phase I project include:

Seismometer-Digitizer Database. Figure 10 shows the GUI layout highlighting access to the instrument response database. The pull-down menu on the right shows a list of the instruments currently in the database. In this example, the pole-zero diagram is presented on the lower right of the GUI and shows the poles (red crosses) and zeros (blue circles) for a Guralp 3T seismometer. The amplitude response for the systems is shown to the left.

NOISETRAN. This feature provides a GUI for calculating the NOISETRAN functions. Figure 11 shows the response transfer function is obtained for station ANMO (Albuquerque, New Mexico, United States) using the TUC (Tucson, Arizona, United States) noise floor.

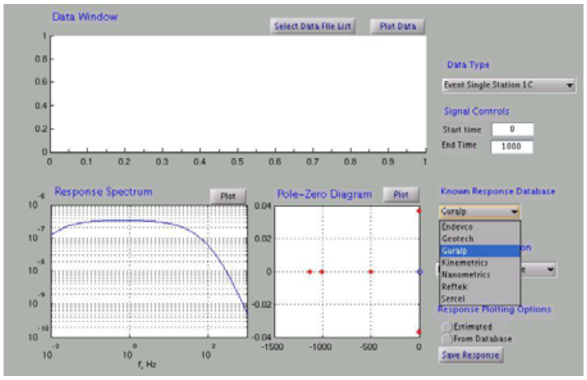

Figure 10. GUI layout with shown access to the database of known responses. In this example the pole-zero diagram (bottom right) corresponds to the response function (bottom left) for Guralp 3T as an example.

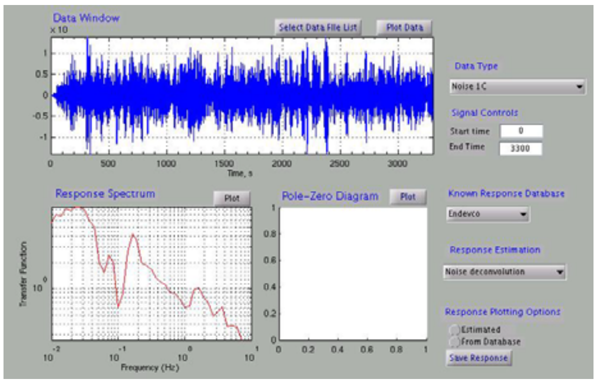

Figure 11. Acceleration NOISETRAN for station ANMO obtained by dividing the spectrum of the noise shown in the window by the noise floor computed for TUC (e.g., Berger et al, 2004).

CONCLUSIONS AND RECOMMENDATIONS

Under the DOE SBIR/STTR FY 2009 Phase I program, we have developed a set of software tools that can recover a ROM sensor/digitizer response function from raw data when the response is either unknown or incorrectly known. Our results indicate that our Phase I project demonstrated both technical feasibility and a high potential to achieve the GNEM R&D goals for system response recovery of important seismic datasets. For the Phase II project, we plan to further enhance the software tools initiated during the Phase I. We will focus on the development of an easy-to-use interface (either using GUIs or scripts) for the seismologist to apply these software tools to their dataset of interest. We will Alpha test the updated version of the system response recovery toolbox on several proprietary datasets recently acquired from the Middle East by Weston Geophysical that have response and metadata inconsistencies. The final component of the Phase II project would include Beta testing of the software toolbox at another data center.

REFERENCES

Berger, J., P. Davis, and G. Ekstrom (2004). Ambient earth noise: a survey of the Global Seismographic Network, *J. Geophys. Res.* 109: B11307.

Ifeachor E. C. and B. W. Jervis (1993). *Digital Signal Processing - A Practical Approach*. Addison-Wesley Publishing Company, Wokingham, England, 492–505.

Herrmann, R. B. (2008). *Computer Programs in Seismology*. St. Louis University.

McNamara, D. E. and R. P Buland (2004). Ambient noise levels in the continental United States, *Bull. Seism. Soc. Am.* 94**:** 1517–1527.

Peterson, J. (1993). Observation and modeling of seismic background noise, *U. S. Geol. Surv. Tech. Rept. 93-322,* 1-95.