2011 SCEC Final Report

EMPIRICAL GREEN'S FUNCTION (EGF) METHODS: IDENTIFYING HOW TO 'GET IT RIGHT' BY STUDYING HOW WE 'GET IT WRONG'

Debi Kilb
Scripps Institution of Oceanography
La Jolla, California 92093-0225
Phone: (858) 822-4607
Email: dkilb@epicenter.ucsd.edu

Abstract

To constrain the source properties of a mainshock earthquake one can use another earthquake as an empirical Green's function (EGF) to approximate the path- and site-effect contributions to the seismic waveform. An ideal EGF earthquake is smaller in magnitude than the mainshock and has a similar focal mechanism and hypocentral location to the mainshock. Here, we quantitatively define optimal EGF event selection using data from the spatially complex San Jacinto Fault Zone (SJFZ) in southern California. The high seismicity rate within the SJFZ allows us to test the EGF method for 51 target M>3 mainshock events using a large range of potential EGFs (>200 for all mainshocks; >1000 for 29 mainshocks). We estimate a mainshock corner frequency from the spectral ratio of each mainshock/EGF waveform pair. The suitability of an EGF event is determined by quantifying the variability of these corner frequencies across the network, assuming an ideal EGF event produces similar corner frequencies at every station. We find optimal EGF events are at least one unit of magnitude smaller than the mainshock, but that selecting EGFs with larger than one magnitude differentials does not produce significantly better results. Negligible differences are found in our results for EGF events located 2-14 km from the mainshock, suggesting an EGF event ~2 km from the mainshock may be as poor a choice as an EGF event ~14 km from the mainshock. We suggest also limiting mainshock/EGF pairs to those with high waveform cross-correlations at all stations (median >0.5 for this dataset).

1.0 Introduction

In applying the EGF technique to a given earthquake (hereafter referred to as the 'mainshock'), the path- and site-effect contributions to the seismic waveform are approximated by the ground motion of a smaller earthquake sharing a similar source location and focal mechanism (*e.g.,* Hartzell, 1978; Mueller, 1985; Hutchings and Wu, 1990a; Hutchings and Wu, 1990b; Mori and Frankel, 1990; Dreger, 1994; Velasco *et al.,* 1994; Hough, 1997; Plicka and Zahradnik, 2002; Vallee, 2004; Prieto *et al.,* 2006; Iglesias and Singh, 2007; Causse *et al.,* 2008; Viegas *et al.,* 2010; Baltay *et al.,* 2011; Chen & Shearer, 2011; Kane *et al.,* 2011). This approach assumes that the smaller earthquake (hereafter referred to as the 'EGF earthquake') approximates a point-source relative to the larger earthquake, and that the waveform of the smaller earthquake primarily represents the propagation path and site effect signals. The mainshock source can therefore be obtained by deconvolving the EGF earthquake signal from the mainshock signal. This technique assumes that the EGF earthquake: (1) is sufficiently small to be approximated as a point-source relative to the mainshock earthquake; (2) is close enough to the mainshock earthquake such that the propagation paths between both events and a given station are identical; and (3) has a focal mechanism solution that mimics the mainshock and therefore the two events share an identical radiation pattern.

2.0 Study Area and Data

We use a mainshock database of 51 M > 3 earthquakes recorded by the ANZA seismic network (Berger *et al.,* 1984; Vernon, 1989). For each mainshock, we create a catalog of potential EGF earthquakes (>200 events for each mainshock) that locate within a 0.2° by 0.2° box centered on the mainshock epicenter. We limit event selection to those in the Lin *et al.* (2007a and 2007b) LSH relocated catalog. For our 51 mainshocks, our selection criteria net 56,196 trial EGF events and 183,953 P-wave velocity waveforms recorded at 100 sps by 11 stations in the ANZA network. For each vertical component waveform, we compute the spectrum using a one-second time window starting 0.25 seconds prior to the P-wave arrival. We also compute a noise spectrum using one second of data prior to the start of the P-wave. These spectra are calculated using the multitaper method (Thomson, 1982) with a time-bandwidth parameter of 3.5. We restrict each EGF record to have a minimum signal to noise ratio (SNR) of 3 within the 2 to 5 Hz band, and we additionally require that a minimum of half of the frequency points examined also have a minimum SNR of 3.

3.0 Source spectra and source parameter estimates

The ground motion recorded at a given station can be represented as a convolution of signals from the source, $s(t)$, the propagation path between source hypocenter and recording station, $p(t)$, and any near-site surface scattering effects and contributions from the recording instrument, $i(t)$. For the mainshock earthquake (indicated by subscript m), we formulate this as:

$$u_m(t) = s_m(t) * p_m(t) * i_m(t) \tag{1}$$

We represent the EGF earthquake similarly using a subscript of *egf*:

$$u_{egf}(t) = s_{egf}(t) * p_{egf}(t) * i_{egf}(t) \tag{2}$$

In accordance with the EGF method, we treat the propagation paths and instrument effects for both the mainshock and EGF earthquakes as identical (*e.g.,* $p_m(t) = p_{egf}(t)$ and $i_m(t) = i_{egf}(t)$). We also assume that the EGF earthquake source is a point-source in time, and hence s_{egf} is a scalar value rather than a function of time. These assumptions allow us to rewrite equation 1 as:

$$u_m(t) = s_{rel}(t) * u_{egf}(t) = \frac{s_m(t)}{s_{egf}} * u_{egf}(t) \tag{3}$$

Here, $s_{rel}(t)$ is the relative source time function obtained by scaling the mainshock source by the amplitude of the EGF source. The waveforms $u_m(t)$ and $u_{egf}(t)$ are the recorded ground motions for the mainshock and EGF earthquake, respectively. Obtaining $s_{rel}(t)$ requires inverting the recorded ground motion equation to solve for the source term.

In practice, it is common to approach this inversion in the frequency domain because the convolution can be simply represented as multiplication:

$$U_m(f) = S(f)U_{egf}(f) \tag{4}$$

Then the relative source spectrum, $S(f)$, is determined through spectral division:

$$S(f) = U_m(f)/U_{egf}(f) \tag{5}$$

In this study, we use the Brune (1970, 1971) source spectrum to fit the relative seismic moment, M_0, and the corner frequency, f_c:

$$S(f) = \frac{M_o}{1 + \left(\frac{f}{f_c}\right)^2} \quad (6)$$

Here, the corner frequency is inversely related to the source rupture duration.

We compute spectral ratios and identify the preferred corner frequency for all possible mainshock/EGF event pairs meeting our signal to noise ratio criteria (67,498 pairs examined). We assume that an ideal choice of mainshock and EGF event pairing will generate similar corner frequency estimates at all stations in the array. In this simple approach, we overlook possible azimuthal dependence in corner frequency measurements generated by effects such as rupture directivity. To gauge the appropriateness of each EGF earthquake, we measure the standard deviation of the corner frequency estimates at all stations in the array. We take the logarithm of the corner frequency estimates prior to measuring this standard deviation in order to normalize the effects of fitting the spectrum in the log domain (*e.g.,* the difference in source spectrum fit for a shift in corner frequency from 10 to 15 Hz is significantly larger than the fit for a shift in corner frequency from 30 to 35 Hz). We define this measured quantity as the 'corner frequency variability'.

4.0 EGF selection criteria

A key aspect of our analysis is that we use a large suite of possible EGF earthquakes representing a broad range of event choice characteristics. This large dataset of trial EGF earthquakes ranges from those within the common limits for EGF selection to events falling far outside the typically accepted limits. The purpose in allowing such a wide range is to identify the transition separating these two populations. We assume the preferred EGF event will produce similar corner frequency estimate at every station and we gauge the suitability of the EGF using the variability in corner frequency estimates for each mainshock and EGF earthquake pairing across the network. Low variability indicates appropriate EGFs and high variability indicates non-appropriate EGFs.

4.1 Importance of EGF/mainshock Magnitude Differential

We begin our study by examining how the magnitude differential between the mainshock and the EGF earthquake can influence source parameter estimates. For each of our 51 mainshocks, we create a catalog of possible EGF events that includes all smaller earthquakes within 3 km (hypocentral distance) of the mainshock. This nets 3134 mainshock/EGF pairs, recorded at up to 11 stations. For each of these mainshock/EGF waveform pairs we compute corner frequencies, as defined above, and track the variability between the mainshock and EGF corner estimates at each station. We analyze the corner frequency variability by looking at the mean values in differential magnitude bins of 0.5 magnitude units. Overall, no strong correlation between corner frequency variability and magnitude differential is observed. The lack of a trend in this relationship suggests it is sufficient to require the EGF earthquake to be at least one unit of magnitude unit smaller than the mainshock, and that magnitude differentials that exceed this one magnitude unit threshold do not necessarily produce significantly better results.

4.2 Importance of Mainshock and EGF Hypocentral Separation Distance

We next assess how the separation distances between the mainshock and EGF hypocenters alter our results. We select a magnitude 3.6 mainshock and a representative sample of EGF events, purposely selecting potential EGF events at hypocentral separation distances out to 20 km, many of which clearly do not satisfy the collocation requirement. We find the signatures of the EGF velocity waveforms vary considerably from station to station, but exhibit similar characteristics between mainshock and EGF records at individual stations when interevent spacing is small. Such similarity would traditionally confirm the validity of the EGF assumptions and the acceptability of these mainshock/EGF pairing. We expect as the spacing between mainshock and EGF hypocenters increases, the mainshock/EGF waveform pairs would become increasingly different in shape at each station because the assumption of identical path propagation effects is less appropriate. However, if asked to identify the waveform from the most distant EGF event for each station one is hard-pressed to select the correct answer. Similarly, a demarcation between waveforms from 'far' and 'close' EGF events is not clearly apparent.

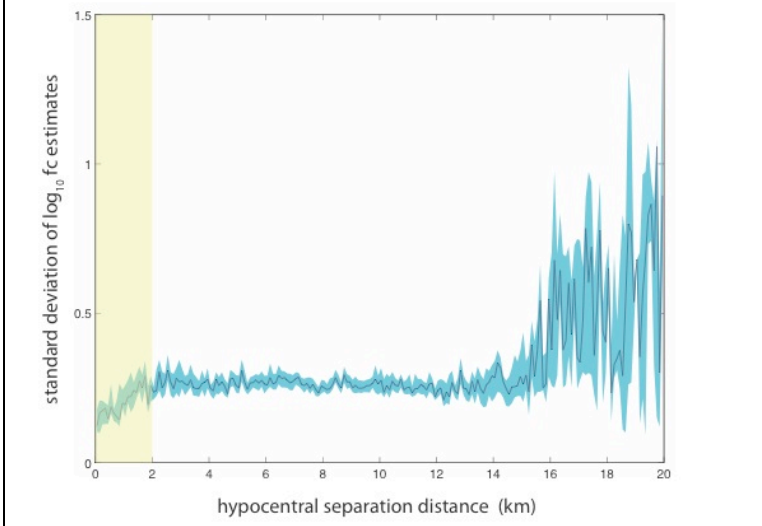

Figure 1: Median corner frequency variability versus hypocentral separation distance using distance bins of 0.1 km width. For reference the shaded region marks 2 km hypocentral separation distance that we assume should be used as a limiting value in EGF selection because it corresponds to the lower standard deviations in corner frequency estimates.

To conduct a more quantitative study, we first limit the EGFs we consider to those with magnitudes ≥1.5 units smaller than the mainshock magnitudes in order to remove any magnitude differential effects. We measure the corner frequency variability (as defined above) for these event pairs and then compute the median of these estimates over 0.1 km separation distance bins of the hypocentral locations (Figure 1). The resulting trend of the median corner frequency variability with changing interevent separation distance has two interesting features. First, the corner frequency variability is smallest at the closest separation distances and increases approximately linearly with separation distance up to ~2 km. This result validates the expectation that the closest events will produce the best estimates and provides an ideal upper bound of 2 km on the maximum separation between the EGF and mainshock. The second interesting feature is the relatively constant median corner frequency variability with separation distance from ~2-14 km. This relatively low variability and the minimal change with distance demonstrates that the mainshock and EGF waveforms have fairly similar characteristics even at larger interevent spacing, or put another way, that the waveforms are not unique enough to discriminate between their different features.

4.3 Importance of Mainshock and EGF Waveform Similarity

In regions of relatively homogeneous focal mechanism distributions, it is reasonable to assume that a mainshock and EGF event pair chosen by location proximity will likely satisfy the requirement of similar focal mechanism solutions. In the SJFZ, however, where the distribution of focal mechanism orientations is heterogeneous (Bailey *et al.,* 2010) relative to other regions in California (*e.g.*, the San Andreas Fault at Parkfield), this criterion is less likely to be satisfied. Because focal mechanism solutions can be prone to substantial uncertainties and smaller magnitude events, such as those typically used as EGF earthquakes (*e.g.,* M<2), often do not have reliable solutions (Hardebeck and Shearer, 2003) instead of using focal mechanism solutions to assist in our EGF selection we prefer using waveform similarity.

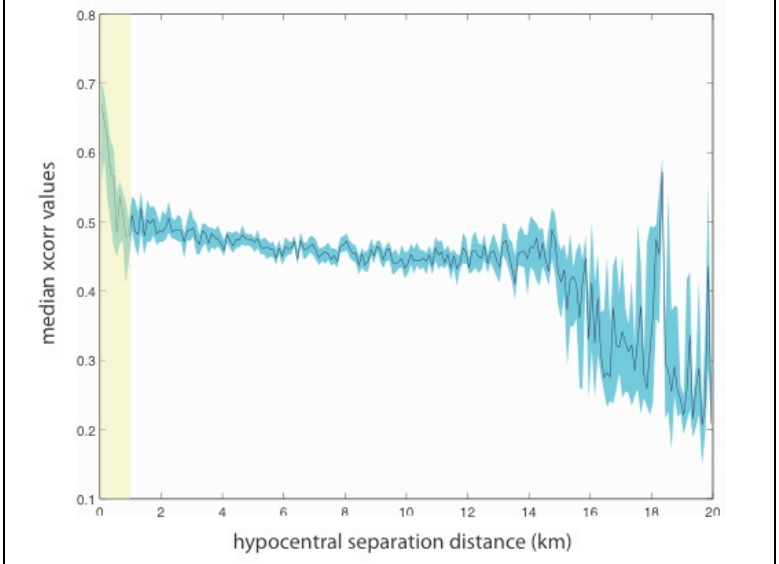

Figure 2: Median peak cross-correlation values versus hypocentral separation distance. These results are shown using distance bins of 0.1 km width. For reference, the shaded region marks 1 km hypocentral separation distance that we assume should be used as a limiting value in EGF selection because it corresponds to the higher cross-correlation values.

For mainshock/EGF event pairs with small separation distances we assume waveform similarity can be used as a proxy for focal mechanism similarity. We compute peak cross-correlation coefficients for each mainshock and EGF waveform pair at each station using 0.25 seconds of data following the P-wave arrival. This time window excludes the noise prior to the P-wave arrival while retaining several cycles of P-wave ground motion. As expected, the EGF events with the highest peak cross-correlation coefficients between EGF and mainshock waveforms indeed exhibit waveforms very similar to the mainshock. We order the waveforms from highest peak cross-correlation to lowest at station KNW. The EGF waveforms for stations FRD and SND follow the same ordering from KNW, yet they do not follow the similar high-to-low cross-correlations arrangement. In fact, the EGF with the highest cross-correlation for station KNW (0.71) produces the lowest cross-correlation at station FRD (0.19). This non-sequential ordering demonstrates that a high correlation value at one station does not assure high values at all other stations, indicating more than one station should always be tested if possible.

To determine if there exists a threshold cross-correlation value that can be used to select optimal EGF events we compute the binned corner frequency variability over cross-correlation coefficient bins of 0.05 width, using event pairs with EGF magnitudes ≥1.5 units smaller than the mainshocks. Here, we consider EGFs at all distances. Because we have a cross-correlation estimate for each station in the array, rather than a single value for the event pair as in our previous analyses, we track the results using the mean, maximum, and median values of cross-

correlation coefficients for each mainshock/EGF event pair at all stations. For each of the three quantities we consider (mean, median and maximum cross-correlation coefficient) all exhibit low corner frequency variability at high levels of correlation, with a transition to increasing corner frequency variability at lower levels of correlation. This observation reinforces the importance of choosing an EGF event with similar waveform characteristics to the mainshock. This also confirms our assumption that similar corner frequencies result across the array when an appropriate EGF is selected and more highly variable corner frequencies results when a poor EGF choice is made. Results of the median cross-correlations as a function of mainshock/EGF hypocentral separation distance (Figure 2) suggest the optimal EGFs are those with median cross-correlation coefficient ≥ 0.5. If waveform cross-correlation information is not available we suggest instead limiting interevent spacings between the mainshock/EGF to $\leq$1 km, which should yield waveform pairs more highly correlated at all stations than spacings exceeding 1 km.

5.0 Conclusions

Based on our analysis of a large amount of data (51 mainshock events; 67,498 mainshock/EGF pairs examined; 11 seismic stations) from the SJFZ we find EGF selection can be greatly improved when using a catalog with precise locations. In agreement with Hutchings and Wu (1990) our results show EGF events should be at least 1 magnitude unit or more smaller than the mainshock, but that magnitude differentials significantly smaller than 1 magnitude unit are not necessarily a better choice (*i.e.,* one magnitude unit is sufficient to satisfy the point-source approximation assumption). We also find the difference in magnitude between the mainshock and EGF event is not as important a factor in EGF selection as hypocentral separation distance. Surprisingly we found a relatively constant median corner frequency variability for mainshock/EGF event pair separation distance from ~2-14 km, suggesting an EGF ~2km from the mainshock may be as inadequate a choice as one ~14 km from the mainshock.

In regions of highly heterogeneous seismic sources like the SJFZ region we suggest that in addition to distance separation and magnitude differential restrictions it is also wise to require high waveform cross-correlation coefficients (median $\geq$0.5 over the array for this dataset). Ideally, the cross-correlation coefficients for any mainshock/EGF pair would be measured to confirm an acceptable EGF choice. We caution that a high correlation value at one station does not assure high values at all other stations, so all stations used in analysis should be tested. Examining median mainshock/EGF waveform cross-correlation coefficients as a function of mainshock/EGF hypocentral separation distance we find a limiting value of <1 km interevent spacings where the waveform pairs are more highly correlated across the array than for larger spacings. From this observation we conclude that separation distances of 1 km should be considered the maximum allowed, at least within regions of high faulting complexity.

In summary, for the SJFZ data we find selecting EGF events based on the hypocentral separation distance between the mainshock and EGF locations is a more important constraint than selection based on magnitude differentials or similarity in waveforms. For regions of highly heterogeneous faulting we recommend selecting EGFs using relocated earthquake catalogs to get accurate separation distances, choosing separation distances of less than 1 km, and confirming waveform similarity between the mainshock and EGF waveforms at each station. Each of these three requirements must be met to obtain robust source parameter estimates and minimize erroneous results.

Intellectual Merit

A key aspect of our analysis is that we use a large suite of possible empirical Green's function EGF earthquakes representing a broad range of event choice characteristics. This large dataset of trial EGF earthquakes ranges from those within the common limits for EGF selection to events falling far outside the typically accepted limits. The purpose in allowing such a wide range is to identify the transition separating these two populations.

Broader Impacts

This project included work by graduate student Deborah Kane.

Publications and presentations partially supported by this grant

Kane D.L., D. L. Kilb, and F.L. Vernon, "Mainshock/Aftershock Sequences Within the Heterogeneous San Jacinto Fault Zone: Assessing if Empirical Green's Function Methods Produce Reliable Results," Abstract S21C-2063, AGU Fall Meeting, 2010.

Kane D.L., D.L. Kilb, and F.L. Vernon, "A Guide to Selecting Empirical Green's Functions in Regions of Fault Complexity: A Study of Data From the San Jacinto Fault Zone, Southern California," SCEC Annual Meeting, Poster B-067, 2011.

Kane D.L., D.L. Kilb, and F.L. Vernon, "Selecting Empirical Green's Functions in Regions of Fault Complexity: A Study of Data From the San Jacinto Fault Zone, Southern California, "Abstract S53B-2290, AGU Fall Meeting, 2011.

References

Bailey, I.W., Y. Ben-Zion, T.W. Becker, and M. Holschneider (2010). Quantifying focal mechanism heterogeneity for fault zones in central and southern California, *Geophys. J. Int.,* doi: 10.1111/j.1365-246X.2010.04745.x

Baltay A., S. Ide G. Prieto and G. Beroza (2011). Variability in earthquake stress drop and apparent stress, *Geophys. Res. Lett.,* V38, L06303, doi:10.1029/2011GL046698.

Berger, J., L. M. Baker, J.N. Brune, J.B. Fletcher, T.C. Hands, and F.L. Vernon III (1984). The Anza array: a high-dynamic-range, broad-band, digitally-radiotelemetered, seismic array, *Bull. Seism. Soc. Am.,* 89, 1469-1481.

Brune, J. (1970). Tectonic stress and the spectra of seismic shear waves from earthquakes, *J. Geophys. Res.*, 75 (26), 4997-5009.

Brune, J. N. (1971). Correction, *J. Geophys. Res.,* 76, 5002.

Causse, M., F. Cotton, C. Cornou, and P.-Y. Bard, (2008). Calibrating Median and Uncertainty Estimates for a Practical Use of Empirical Green's Functions Technique, *Bull. Seism. Soc. Am.,* 98, pp. 344–353, doi: 10.1785/0120070075.

Chen, X. and P.M. Shearer (2011). Comprehensive analysis of earthquake source spectra and swarms in the Salton Trough, California, *J. Geophys. Res.,* V116, B09309, 17 PP., 2011 doi:10.1029/2011JB008263.

Dreger, D.S., (1994). Empirical Green's function study of the January 17, 1994 Northridge, California earthquake, *Geophys. Res. Lett.* 21, 2633–2636.

Hardebeck, J. L. and P. M. Shearer, (2003). Using S/P Amplitude Ratios to Constrain the Focal Mechanisms of Small Earthquakes, *Bull. Seismo. Soc. Am.*, 93, 2434-2444.

Hartzell, S. (1978). Earthquake aftershocks as Green's functions, *Geophys. Res. Lett.*, 5(1).

Hough, S.E. (1997), Empirical Green's function analysis: taking the next step, *J. Geophys. Res.,* Vol. 102, No. B3, pp. 5369-5384.

Hutchings, L. and F. Wu (1990a). Empirical Green's Functions From Small Earthquakes: A Waveform Study of Locally Recorded Aftershocks of the 1971 San Fernando Earthquake, *J. Geophys. Res.*, Vol. 95, No. B2, pp.1187-1214.

Hutchings, L. and F. Wu (1990b). Small earthquakes as empirical Greens functions: a waveform study of locally recorded after-shocks of the 1971 San Fernando earthquakes, *J. Geophys. Res.,* 95, pp. 1187-1214.

Iglesias, A. and Singh, S. K., (2007). Estimation of radiated energy using the EGF technique: What should be the upper limit of integration in the frequency domain?, *Bull. Seismol. Soc. Am.*, 1346-1349.

Kane, D.L., G.A. Prieto, F.L. Vernon, and P.M. Shearer (2011), Quantifying Seismic Source Parameter Uncertainties, *Bull. Seismol. Soc. Am.*, vol. 101 no. 2, 535-543, doi: 10.1785/0120100166.

Lin G. (2007). "LSH Catalog", http://www.rsmas.miami.edu/personal/glin/LSH.html.

Lin, G., P.M. Shearer, and E. Hauksson (2007), Applying a three-dimensional velocity model, waveform cross correlation, and cluster analysis to locate southern California seismicity from 1981 to 2005, *J. Geophys. Res.*, Vol. 112, B12309, doi:10.1029/2007JB004986.

Mori, J. & Frankel, A., 1990. Source parameters for small events associated with the 1986 North Palm Springs, California, earthquake determined using empirical Green functions, *Bull. seism. Soc. Am.,* 80, 278– 295.

Mueller, C. S. (1985). Source pulse enhancement by deconvolution of an empirical Green's

function, *Geophys. Res. Lett.* 22, 33-36.
Plicka V., Zahradnik J. (2002). The EGF method for dissimilar focal mechanisms: the Athens 1999 earthquake. *Tectonophysics,* 359, 81-95.
Prieto, G. A., R. L. Parker, F. L. Vernon, P. M. Shearer and D. J. Thomson (2006). Uncertainties in earthquake source spectrum estimation using empirical Green functions. *Earthquakes: Radiated Energy and the Physics of Faulting*, Abercrombie, McGarr, Kanamori, and di Toro eds. AGU Geophys. Monograph 170, pp. 69-74. doi : 10.1029/170GM08.
Thomson, D.J. (1982), Spectrum estimation and harmonic analysis, *Proceedings of the IEEE*, Vol. 70 (9), pp. 1055-1096.
Vallee, M. (2004). Stabilizing the Empirical Green Function Analysis: Development of the Projected Landweber Method, *Bull. Seism. Soc. Am.*, Vol. 94 (2), pp. 394-409.
Velasco, A.A., C.J. Ammon, and T. Lay (1994). Empirical green function deconvolution of broadband surface waves: Rupture directivity of the 1992 Landers, California (Mw= 7.3), earthquake, *Bull. Seism. Soc. Am.*, Vol. 84, No. 3, pp. 735-750.
Vernon, F.L. III (1989). Analysis of data recorded on the ANZA seismic network *Ph.D. Thesis University of California at San Diego*.
Viegas G., R. E. Abercrombie, and W.-Y. Kim (2010). The 2002 M5 Au Sable Forks, NY, earthquake sequence: Source scaling relationships and energy budget, *J. Geophys. Res.*, V115, B07310, 20 PP., 2010 doi:10.1029/2009JB006799.