Uncertainties in Earthquake Source Spectrum Estimation Using Empirical Green Functions

Germán A. Prieto, Robert L. Parker, Frank L. Vernon, Peter M. Shearer

Institute of Geophysics and Planetary Physics, University of California San Diego, La Jolla, California, USA

David J. Thomson

Department of Mathematics and Statistics, Queen's University, Kingston, Canada

We analyze the problem of reliably estimating uncertainties of the earthquake source spectrum and related source parameters using Empirical Green Functions (EGF). We take advantage of the large dataset available from 10 seismic stations at hypocentral distances (10 km < d < 50 km) to average spectral ratios of the 2001 M5.1 Anza earthquake and 160 nearby aftershocks. We estimate the uncertainty of the average source spectrum of the M5.1 target earthquake by performing propagation of errors, which, due to the large number of EGFs used, is significantly smaller than that obtained using a single EGF. Our approach provides estimates of both the earthquake source spectrum and its uncertainties, plus confidence intervals on related source parameters such as radiated seismic energy or apparent stress, allowing the assessment of statistical significance. This is of paramount importance when comparing different sized earthquakes and analyzing source scaling of the earthquake rupture process. Our best estimate of radiated energy for the target earthquake is 1.24×10^{11} Joules with 95% confidence intervals (0.73×10^{11}, 2.28×10^{11}). The estimated apparent stress of 0.33 (0.19, 0.59) MPa is relatively low compared to previous estimates from smaller earthquakes (1MPa) in the same region.

1. INTRODUCTION

A fundamental problem in seismology is accurate estimation of the radiated seismic energy of an earthquake. When an earthquake occurs, some fraction of the total energy is radiated as seismic waves, providing important information about the earthquake rupture process. Determining the amount of radiated seismic energy can be difficult and large differences in these estimates have been found among different techniques and groups [*Perez-Campos et al.*, 2003].

Earthquakes: Radiated Energy and the Physics of Faulting
Geophysical Monograph Series 170

10.1029/170GM08

Partly the difficulty lies in the need to estimate radiated seismic energy over a large dynamic range and a wide frequency band, from the low frequencies needed to define the seismic moment to well beyond the corner frequency. Another difficulty comes from the need for path and site corrections; these corrections have much larger uncertainties at higher frequencies. Finally, the earthquake source might have more complicated features than first expected, including directivity effects that need to be taken into account in order to avoid biasing the estimation.

Over the last 15 years there have been many studies of the radiated seismic energy of different sized earthquakes in different regions of the Earth [e.g., *Kanamori et al.*, 1993; *Abercrombie*, 1995; *Choy and Boatwright*, 1995; *Mayeda*

and Walter, 1996; *Ide and Beroza*, 2001; *Venkataraman et al.*, 2002; *Mori et al.*, 2003; *Prieto et al.*, 2004 and many others]. What is often missing in these studies is a measure of the uncertainty of each individual estimate. The question of the uncertainty of estimates is of key importance for describing the significance of one measurement compared to another. In particular, how do the errors in the assumed attenuation model *propagate* into the uncertainty of the source spectrum?

The major unknown in the system is the transfer function between source and receiver. We will focus in this paper on developing a technique to use Empirical Green Functions and estimating and reducing the uncertainties by averaging over a set of suitable aftershocks. By using many EGFs we effectively are randomly sampling the errors in the path effects and averaging over the propagation space. As an example we will present results for the October 31 2001 M5.1 Anza earthquake, using data from local stations.

2. DATA PROCESSING

A M_L 5.1 earthquake occurred on 31 October 2001 in the Anza region in southern California (hypocenter 33.5081°N, 116.5143°W, depth 15.2 km, M_W 4.7). The earthquake exhibited thrust motion on a vertical fault striking N35°E [*Hauksson et al.*, 2002]. The data comes from Anza broadband velocity sensors and IDA strong motion sensors at station PFO and both broadband and strong motion sensors from TriNet.

All stations are within 10 to 50 km from the source region. The choice of using local stations to obtain the source spectrum is motivated by the need to obtain reliable estimates of the spectrum at high frequencies and also the need to have good signal-to-noise ratios for the smaller aftershocks. The epicentral distance between the mainshock and all the aftershocks used is less than 2 km (see map in Figure 1 and Table S1 on the CD-ROM accompanying this volume).

Since *S*-*P* times for the closer stations are small, we take a 12 second window that includes both *P* and *S* waves, starting 1 second before the *P* wave pick (e.g., *Venkataraman et al.,* 2002). We also select a noise window of the same length, immediately preceding the signal window. A signal-to-noise ratio (SNR) is obtained by the ratio of the signal and noise spectra. The spectra of the waveforms are estimated using the multitaper technique [*Thomson*, 1982], which not only allows a variance reduction of the spectra within a certain bandwidth, but also provides an estimate of the uncertainties at each frequency bin using a jackknife approach [*Vernon*, 1989; *Thomson and Chave*, 2001]. For each waveform at each station we obtain the spectrum and the 5–95% confidence interval.

Since we are interested in the source spectrum of the *M*5.1 earthquake, we perform spectral division between the mainshock and all the aftershocks recorded at each station. Following the theory of propagation of errors [*Taylor*, 1997], dividing two random variables results in a new random variable $R_i(f)$ (the spectral ratio), where the *i* term represents the *i*th aftershock used for deconvolution, with uncertainties being a function of the

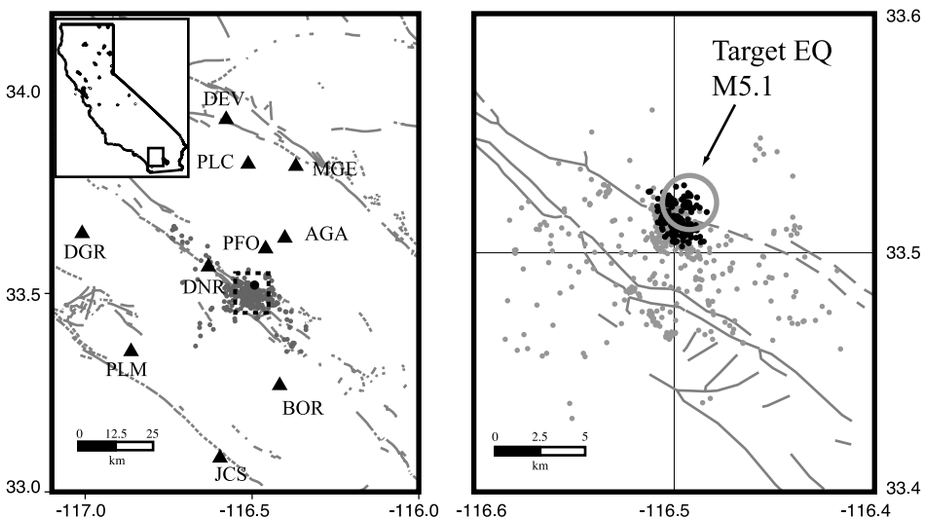

Figure 1. Maps of the study area. The left map covers the entire study region, showing the seismic stations (black triangles), the *M*5.1 target Anza earthquake (black circle) and aftershocks M2.9 and lower (gray circles). The inset shows the location of the study area and the state of California. A close-up region (dashed box) is shown in the right hand map. The target earthquake is shown as an open gray circle of radius proportional to a 1MPa stress drop event. Aftershocks used in this study (black circles) and general aftershock seismicity of the region (gray circles) are shown.

uncertainties of the spectra of the mainshock and aftershock. As shown in Figure 2, the relative uncertainties in *R(f)* are larger than the individual components in *X(f)* and *Y(f)* , due to the instability of the deconvolution process, or simply because we are dividing two noisy spectra. The gray shaded area is larger in Figure 2 for the spectral ratio.

As noted by *Aki and Richards* [1980] (Volume II, Chapter 11.5.5) simple spectral density estimates have only two degrees of freedom and are distributed as χ^2_2; consequently, their ratio is distributed statistically as $F_{2,2}$, a distribution so broad it has infinite variance. We improve on this situation by making use of multitapers: each spectral estimate is made by weighted average over 7 tapers (and time-bandwidth product 4). Then the number of degrees of freedom increases to about 10 (less than the expected 14 because of the weighting).

We take the log of the spectral ratios and stack our data in 15 bins divided by the magnitude of the EGF used (ΔM_L = 0.2). Given that variations in spectra are generally observed to be log-normally distributed, throughout the paper we will average and stack spectra in the log domain, which is equivalent to using a geometric average rather than the arithmetic mean. This also ensures that low amplitude spectra contribute equally to average spectral shapes compared to high amplitude

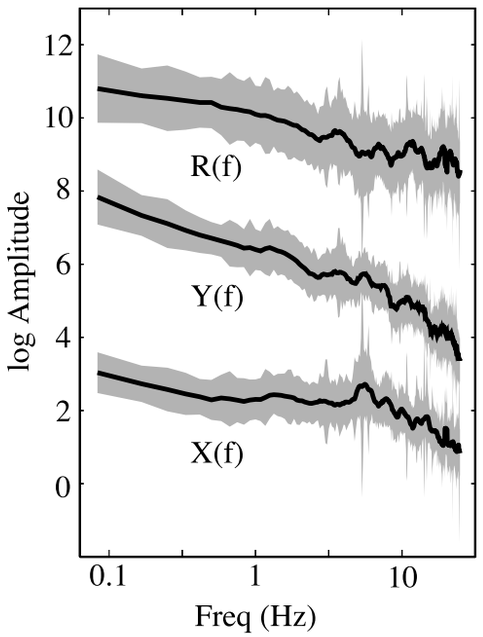

Figure 2. Estimates of the spectrum for ground motion associated with the M5.1 earthquake $Y(f)$ at station AGA and the spectrum of the largest aftershock $M2.9$ $X(f)$ at the same station and corresponding spectral ratio $R(f) = Y(f)/X(f)$ (offset for comparison purposes). The gray area represents the 95% confidence interval estimated using the jackknife approach and propagation of errors. Note how uncertainties grow on $R(f)$ after spectral division.

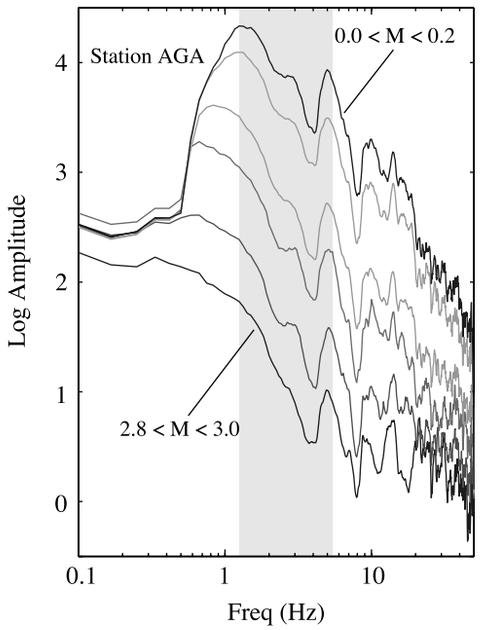

Figure 3. Selected bins of spectral ratios (ΔM_L = 0.2) at station AGA. The effect of the corner frequency of the 2.8< M <3.0 earthquakes (lower line) flattens the spectral ratio at high frequencies, while the low SNR affects lower frequencies of the smaller EGFs, but gives reasonable spectral fall-off at high frequencies. The gray shaded area shows the band used to estimate a scaling factor for combining the spectral ratios.

spectra. Figure 3 shows a set of typical spectral ratios $R_i(f)$ for EGFs with a range of earthquake magnitudes. Note the effect of the corner frequency of the larger aftershocks and the effect of the low SNR at low frequencies for the smaller aftershocks. The uncertainties (not shown in Figure 3) for each particular bin are again estimated using the method of propagation of errors (*Taylor*, 1997). For each station a scaling factor for the spectral ratios is determined over a narrow band where the shapes are consistent (gray area in Figure 3).

3. THE COMBINED EMPIRICAL GREEN FUNCTION

In order to take advantage of the large wealth of data available from the local stations, including very small aftershocks (M 0–0.5), we will average together the different estimates of spectral ratios. This will enable us to bring down the variance or uncertainties of the source spectrum.

The spectral ratio *R(f)* contains two sources of error, (a) the variance due to the intrinsic spectral estimation process, which we take from the multitaper algorithm and (b) systematic errors (which we will call bias) due to the effect of the corner frequency of the EGF. As described by *Ide et al.* [2003], the spectral ratio *R(f)* of two earthquakes

located close to each other, assuming the same focal mechanisms and path effects, can be expressed as

$$R(f) = \frac{Y(f)}{X(f)} = \frac{S(f)}{G(f)} \tag{1}$$

where *Y(f)*, *X(f)* represent the spectrum of the ground motions of the mainshock and aftershock and *S(f)*, *G(f)* represent the source spectrum of the mainshock and the aftershock respectively.

An approximate form of the source spectrum [*Brune*, 1970] is

$$G(f) = \frac{M_0}{1 + (f/f_c)^2} \tag{2}$$

where M_0 is the seismic moment and f_c is the corner frequency of the EGF. An ideal case for an EGF would be to use an earthquake whose corner frequency f_c was very large, giving then in the log domain

$$\log R(f) = \log S(f) - \log M_0 \tag{3}$$

resulting in a scaled version of the source spectrum, without changing its shape.

The bias of the log spectrum that is created by a finite corner frequency is then

$$b(f) = \log(1 + (f/f_c)^2) \tag{4}$$

which clearly shows that as the frequency f grows and approaches the corner frequency of the EGF, the bias of the source spectrum increases.

We do not know the true corner frequency of the EGF and the relatively low sampling rate of the stations in the network (100 sps) is not enough to estimate it from spectral ratios (as in Hough, 1997; *Ide et al.*, 2003). Instead, we use a simple scaling relation to get an approximate corner frequency. We assume that f_c can be approximated [*Venkataraman et al.*, 2002] from $f_c = 0.49\beta(\Delta\tau/M_0)^{1/3}$ where $\Delta\tau$ is stress drop, and we use a value of 1 MPa. This choice of stress drop is rather arbitrary but certainly within the average in southern California and the study region (*Vernon*, 1989; *Shearer et al.*, 2006). We chose a rather low stress drop, as a conservative value to obtain small f_c's for the EGFs and limit the potential bias at high frequencies. We estimate the seismic moment following the procedure from *Prieto et al.* [2004], that is, we assume local magnitude $M_L = M_w$ for the small earthquakes and use the *Kanamori* [1977] relation to obtain M_0. One could argue that this approximation is not accurate, but as it turns out, even allowing the corner frequency of the EGF to change by 20–30% does not substantially affect the results, changing the radiated energy estimate of the mainshock by less than 3% in our example.

A common technique for dealing with estimates that contain variance and bias as sources of error is the *mean-square error* (MSE) [*Rice*, 1988]. At a given station we construct the source spectrum of the target earthquake *S(f)* from a linear combination of the spectral ratios

$$\log S(f) = \sum_{i=1}^{N} w_i(f) \log R_i(f) \tag{5}$$

where the index i in the sum runs over the events, w_i are the weights for each spectral ratio, and N is the number of EGF available at a particular station. As explained earlier, the idea is to create a weighted average of the spectral ratios.

The MSE is the sum of the variance and the bias squared of the estimate. Applying this idea to our linear combination of spectra

$$mse^2 = \sum_{i=1}^{N} w_i^2 \sigma_i^2 + \left(\sum_{i=1}^{N} w_i b_i \right)^2 - \lambda \sum_{i=1}^{N} w_i \tag{6}$$

where the first term represents the variance (σ_i^2), the second term is the bias squared and we added a normalization constraint as a Lagrange multiplier λ.

Taking the derivative of (6) with respect to the unknowns w_i and λ and minimizing, we obtain two sets of linear equations

$$\sum_{j=1}^{N} \left[\sigma_i^2 \, \delta_{ij} + b_i b_j \right] w_j - \lambda = 0 \tag{7}$$

$$\sum_{j=1}^{N} w_j = 1 \tag{8}$$

We solve the linear set of equations for each individual frequency f with a non-negative least squares approach [see *Lawson and Hanson*, 1974], to obtain only positive weights. We find that requiring a SNR of 5 or larger for a particular $R_i(f)$ leads to better results. This approach will discard most of the spectra below 0.8 Hz due to low SNR for the smaller magnitude EGFs, and will only use the set of larger EGFs. At higher frequencies more spectral ratios have good SNR, but the larger earthquakes have either larger bias terms or the variance is much larger than for the smaller EGF (since the network records many small aftershocks, thus the variance of a particular spectral ratio bin is decreased), so that the weights prefer the smaller EGF. This means we keep the part of each spectrum that has good SNR and discard only the frequency points with low SNR.

Figure 3 shows $R_i(f)$ for a range of EGF earthquake magnitudes at station AGA. We scale the spectral ratios before combining the estimates in (5), since the absolute amplitudes depend on the seismic moment of the EGF used. The scaling factor is obtained from a narrow band (shaded area in Figure 3) where the spectral ratios are consistent. We tested the effect of

the scaling factor by changing the band width used for shifting the spectra. Because the resultant energy estimate does not vary significantly, we choose to ignore this effect.

Solving for the weights in (6) and applying the result in (5) we obtain a relative source spectrum for each station in the network. Following *Prieto et al.* [2004] we calibrate the relative source spectrum at each station to the seismic moment of the target earthquake. We express the mean source spectrum by averaging over the different seismic stations.

$$\log \bar{S}(f) = \frac{1}{K}\sum_{k=1}^{K} \log S_k(f) \tag{9}$$

where K are the number of stations and S_k is the outcome of (5) for station k. The processing and averaging is done one frequency at a time, at individual stations, and the mean source spectrum $\bar{S}$ represents a station average.

Figure 4 shows the source spectrum $\bar{S}(f)$ for our target event, with 5–95% confidence intervals at each frequency point. Note how for the lower frequencies the uncertainties are larger, since the weights are non-zero only for the larger aftershocks, while at higher frequencies many events are available, an average over many independent estimates, increasing the number of degrees of freedom and decreasing the variance.

4. RESULTS FOR RADIATED SEISMIC ENERGY

From the mean source spectrum $\bar{S}(f)$ and the uncertainties, it is possible now to obtain an estimate of the radiated seismic energy [*Vassiliou and Kanamori*, 1982] and its uncertainties for our target earthquake:

$$E_R = \frac{4\pi}{5\rho}\left[\frac{1}{3\alpha^5}+\frac{1}{2\beta^5}\right]\int_0^\infty f^2 |\bar{S}(f)|^2 df \tag{10}$$

where ρ is the density, and α and β are the P and S wave velocity at the source. We set ρ = 2700kg/m^3, and the velocities are taken from *Scott et al.* [1994], with α = 6000m/s and β = 3465m/s. The apparent stress is

$$\tau_a = \mu E_R / M_0 \tag{11}$$

where we use μ= 3.24 × 10^4 MPa.

Since the S wave contains more than 90% of the energy we only use the second term of (10). The radiated energy is calculated by integrating (numerically) the mean source spectrum up to the highest frequency possible (30 Hz). We also extrapolate at the higher frequencies assuming a fall-off rate of ω^{-1} in velocity, which contains less than 5% of the total energy. The 95% bounds are integrated to obtain energy uncertainties. The value of radiated energy obtained is 1.24 × 10^{11} (0.73 × 10^{11},2.28 × 10^{11}) Joules, with the 95% confidence interval in parenthesis. We also compute a value of the apparent stress (proportional to ER/M_0) and obtain 0.33 (0.19, 0.59) MPa, with the 95% confidence interval.

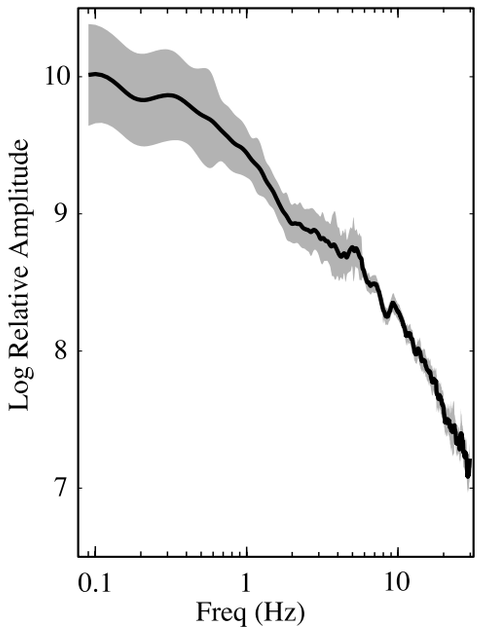

Figure 4. Mean source spectrum over 8 stations with 95% confidence intervals (gray area) for the M5.1 target earthquake. Compared to Figure 2, the uncertainties have been decreased due to averaging over different stations and by the combined EGF, especially at high frequencies where many small aftershocks can be used to perform spectral division.

The apparent stress obtained is within the values of previous results for similar sized earthquakes, and a little lower than smaller events in the same region [*Prieto et al.*, 2004]. But note in this case we have not only obtained the measure of radiated energy, but also the uncertainties, which because we have used over 100 EGFs, have been reduced significantly compared to what one would obtain using just one EGF (i.e., compare Figures 2 and 4). Our analysis focuses on random variations in our input data (spectral ratios) rather than uncertainties in model parameters. Therefore we do not attempt to propagate errors associated with parameters such as wave speed or density, even though these errors might be important. From equation (10) it would be straight forward to propagate such errors, if known, following Taylor (1997).

5. DISCUSSION

As explained before, using spectral ratios will increase the variance and a larger number of source models (corner frequencies, fall-off rates) are to be allowed within the uncertainties. Since stress drop varies as f_c^3, the uncertainties of the

stress drop as estimated from f_c will grow considerably, making the inference of scaling features between different earthquakes more difficult. Our goal is to show a consistent way of estimating the uncertainties of the earthquake source spectrum and some source parameters using the method of propagation of errors and the variance of the spectrum estimation procedure. A source spectrum is then constructed from a weighted set of spectral ratios in order to reduce the variance.

As pointed out by *Sonley and Abercrombie* (2006, this volume) it is good practice to check whether a realistic source pulse is obtained after deconvolution, by inverse FFT of the multitaper eigencoefficients. We compute source pulses for our target M5.1 earthquake and obtain results similar to those of *McGuire* (pers. com., 2005) for the same event. However it is not possible to check all our EGFs since SNR limitations at low frequencies affect the very small EGFs and it is not always possible to recover source pulses.

In our data set we find that there is no single EGF with good SNR and appropriate bias reduction on the entire frequency band of interest (about 0.1–30 Hz) and it is necessary to use multiple EGFs. Even if an ideal EGF is found, the uncertainties of the radiated energy estimate would be considerably larger than presented here and should be taken into account when comparing different results or looking for scaling of energy with earthquake magnitude.

A possible major source of bias is if many EGFs used in this study (especially the small ones) have consistently different focal mechanisms that are not accounted for. We believe that by using many EGFs we are sampling a wide variety of focal mechanisms. For the target event we obtain realistic source pulses, suggesting similar mechanisms. The aim of this study is to show a way of estimating and reducing uncertainties of source spectra based on the target and EGF spectra. It is not intended to completely remove all possible biases associated with the choice of events and model parameterization.

Acknowledgments. We thank Glen Offield for field operations and the Anza group at UCSD who picked and archived the seismograms. We thank M. Hellweg, V. Oye and R. Abercrombie (editor) for thoughtful comments. Funding for this research was provided by NSF Grant number EAR0417983. This research was also supported by the Southern California Earthquake Center. SCEC is funded by NSF Cooperative Agreement EAR-0106924 and USGS Cooperative Agreement 02HQAG0008. The SCEC contribution number for this paper is 933.

REFERENCES

Abercrombie, R. E., Earthquake source scaling relationships from -1 to 5 M_L using seismogram recorded at 2.5-km depth, *J. Geophys. Res.*, *100*, 24,015–24,036, 1995.

Aki, K., and Richards, P., *Quantitative Seismology*, W.H. Freeman, New York., 932 pp., 1980.

Brune, J. N., Tectonic stress and seismic shear waves from earthquakes, *J. Geophys. Res.*, , 4997–5009, 1970.

Choy, G. L. and Boatwright, J. L., Global patterns of radiated seismic energy and apparent stress, *J. Geophys. Res.*, , 18,205–18,228, 1995.

Hauksson, E., Jones, L. M., Perry S., and Hutton, K., Emerging from the Stress shadow of the 1992 M_w 7.3 Landers southern California earthquake? A preliminary Assessment, *Seismol. Res. Lett.*, *73*, 33–38, 2002.

Ide, S., and Beroza, G. C., Does apparent stress vary with earthquake size?, *Geophys. Res. Lett.*, *28*, 3349–3352, 2001.

Ide, S., Beroza, G. C., Prejean, S. G., and Ellworth, W. L., Apparent break in earthquake scaling due to path and site effects on deep borehole recordings, *J. Geophys. Res.*, *108*, B5, 2271, doi:10.1029/2001JB001617, 2003.

Kanamori, H., The energy release in great earthquakes, *J. Geophys. Res.*, *82*, 2981–2988, 1977.

Kanamori, H., Mori, J., Hauksson, E., Heaton, T. H., Hutton, L. K., and Jones, L., Determination of earthquake energy release and M_L using TERRAscope, *Bull. Seismol. Soc. Am.*, *83*, 330–346, 1993.

Lawson, C. L., and Hanson, R. J., *Solving Least Squares Problems*, Prentice-Hall, Englewood Cliffs, NJ., 1974.

Mayeda, K., Walter, W. R., Moment, energy, stress drop, and source spectra of western United States earthquakes from regional coda envelopes, *J. Geophys. Res.*, *101*, 11,195–11,208, 1996.

Mori, J., Abercrombie, R. E., and Kanamori, H., Stress drops and radiated energies of aftershocks of the 1994 Northridge, California, earthquake, *J. Geophys. Res.*, *108*, B11, 2545, doi:10.1029/2001JB000474, 2003.

Perez-Campos, X., Singh, S. K., and Beroza, G. C., Reconciling teleseismic and regional estimates of seismic energy, *Bull. Seismol. Soc. Am.*, *93*, 2123–2130, 2003.

Prieto, G. A., Shearer, P. M., Vernon, F. L., and Kilb, D., Earthquake source scaling and self-similarity estimation from stacking *P* and *S* spectra, *J. Geophys. Res.*, *109*, B08310, doi:10.1029/2004JB003084, 2004.

Rice, J. A., *Mathematical Statistics and Data Analysis*, Brooks-Cole Pub. Co., Monterey, CA, 1988.

Scott, J. S., Masters, T. G., and Vernon, F. L., 3-D velocity structure of the San Jacinto fault zone near Anza, California—I. P waves, *Geophys. J. Int.*, *119*, 611–626, 1994.

Shearer, P. M., Prieto, G. A., Hauksson, E., Comprehensive analysis of earthquake source spectra in southern California, *J. Geophys. Res.*, *111*, B06303, doi:10.1029/ 2005JB003979, 2006.

Sonley, E., Abercrombie, R. E., Variability introduced by methods of attenuation correction: the effect on source parameter determination, this volume, 2006.

Taylor, J. R., *An Introduction to Error Analysis*, University Science Books, 2nd Edition, 1997.

Thomson, D. J., Spectrum estimation and harmonic analysis, *Proc. of IEEE*, *70*, 1055–1096, 1982.

Thomson, D. J., and Chave, A. D, *Jackknife error estimates for spectra, coherences, and transfer functions*, In S. Haykin, editor, *Advances in Spectrum Analysis and Array Processing*, Vol. 1 Chapter 2. Pages 58–113, Prentice Hall, 1991.

Vassiliou, M. S., and Kanamori, H., The energy release in earthquakes, *Bull. Seismol. Soc. Am.*, *72*, 371–387, 1982.

Venkataraman, A., Rivera, L., and Kanamori, H., Radiated energy from the 16 October 1999 Hector Mine earthquake: Regional and teleseismic estimates, *Bull. Seismol. Soc. Am.*, *92*, 1256–1265, 2002.

Vernon, F. L., *Analysis of data recorded on the ANZA seismic network*, *PhD Thesis*. University of California San Diego, 1989.

G.A. Prieto, Institute of Geophysics and Planetary Physics, University of California San Diego, La Jolla, CA 92093-0225. (e-mail: gprieto@ucsd.edu)