13th World Conference on Earthquake Engineering
Vancouver, B.C., Canada
August 1-6, 2004
Paper No. 243

QUANTITATIVE MEASURE OF THE GOODNESS-OF-FIT OF SYNTHETIC SEISMOGRAMS

John G. ANDERSON[1]

ABSTRACT

To develop credibility of synthetic seismograms for engineering applications, there is a need for a quantitative score that can be used to characterize the how well the synthetic matches the statistical characteristics of observed records. Recognizing that strong motion is a very complex time series and any measure that relies on a single parameter for the comparison is seriously incomplete, this paper examines use of a suite of measurements. To be specific, we score seismograms that have been filtered into up to ten narrow pass-bands. Each frequency band is scored on ten different characteristics. The characteristics scored are the peak acceleration, peak velocity, peak displacement, Arias intensity, the integral of velocity squared, Fourier spectrum and acceleration response spectrum on a frequency-by-frequency basis, the shape of the normalized integrals of acceleration and velocity squared, and the cross correlation. Each characteristic is compared on a scale from 0 to 10, with 10 giving perfect agreement. Scores for each parameter are averaged to yield an overall quality of fit. A score below 4 is a poor fit, a score of 4-6 is a fair fit, a score of 6 to 8 is a good fit, and a score over 8 is an excellent fit. One horizontal component of an actual seismogram typically fits the other horizontal component in the "good" range. The method is applied to a blind prediction of ground motions at a station 3 km from the fault in the M7.9 Denali Fault, Alaska, earthquake of November 3, 2002.

[1] Nevada Seismological Laboratory and Department of Geological Sciences, Mackay School of Earth Sciences and Engineering, College of Science, University of Nevada, Reno, Nevada 89557. Phone: 775-784-4265. Email: jga@seismo.unr.edu

INTRODUCTION

One goal of strong motion seismology is to achieve the ability to generate synthetic strong motion accelerograms that are sufficiently realistic that they are useful for engineering applications. Numerous methods to accomplish this have been proposed, including stochastic models: e.g. Boore [1]; models that use empirical Green's functions: e.g. Hartzell [2], Irikura [3,4]; and models that start with the representation theorem, e.g. Hadley [5], Somerville [6], and Zeng [7]. More recent studies have incorporated an increasing amount of physics into the process.

It is somewhat difficult to evaluate how satisfactory the results are for engineering applications. A difficulty is that there is no consensus on how to evaluate the quality of the fit. Every attempt to characterize strong ground motion with a single parameter, such as peak acceleration, is certain to fail. The strong motion accelerogram is variable in its duration, spectral shape, spectral amplitudes, and when, during the seismogram, the energy arrives. Seismograms that are similar on three of these criteria may differ on the fourth and consequently have very different appearance.

This paper proposes a similarity score based on averages of the quality of fit measuring all of these characteristics. Since the criterion is broadly based, it is not obvious what constitutes a good or a poor score. To answer this question, the paper finds the score for the similarity of multiple realizations of synthetic strong motion records generated by the same random process, and also the similarity of the two horizontal components of observed strong motion accelerograms.

PROPOSED CRITERION

The proposed goodness of fit score is described in detail by Anderson [8]. It is derived from ten individual quality measures, each applied to seismograms in ten frequency bands. The quality measures are listed in Table 1. It is convenient to discuss the different goodness of fit criteria in an order different from that in Table 1. For all comparisons, consider two accelerograms, $a_1(t)$ and $a_2(t)$. Let $v_i(t)$ and $d_i(t)$ represent the velocity and displacement obtained from $a_i(t)$. Let $SA_i(f_j)$ be the acceleration response spectrum (5% damping) at frequency f_j, and let $FS_i(f_j)$ be the Fourier amplitude spectrum of the i^{th} accelerogram at frequency f_j. The Arias integral is:

$$I_{Ai}(t) = \frac{\pi}{2g}\int_0^t a_i^2(\tau)d\tau \qquad (1)$$

Arias [9] defined the Arias intensity as $IA = I_A(T_d)$ where T_d is the complete duration of the accelerogram. The energy associated with the ground motion is related to the integral of the velocity, rather than the integral of the acceleration. Therefore, I define the energy integral:

$$I_{Ei}(t) = \int_0^t v_i^2(\tau) d\tau \tag{2}$$

The total energy carried by the seismic waves is proportional to $I_E(T_d)$, where the complete relationship incorporates the velocity of the waves and the density of the ground.

It is assumed that these accelerograms have the same duration and the same origin time. Numerically, it is assumed that they are both sampled with the same sample rate, and thus are represented by matrices of the same length. Differences in duration can, of course, be handled by padding the shorter accelerogram with zeros. The assumption of the same origin time is perhaps the most difficult to assure as in some cases strong motion records do not have known start times, and in any case there may be some uncertainty over the origin time of the earthquake. The scoreing parameters C1, C2, and C10 are sensitive to the relative start times, but rest are not.

The order of the presentation in Table 1 goes from integral measures, which are potentially the easiest to fit, to peak values, then to spectral amplitudes, and finally to the cross correlation, which is by far the most difficult of all of the parameters to achieve a high score. However, it is easier to explain the various measures when they are taken in a different order.

First consider the parameter Spga (C5), the goodness of fit of the peak acceleration. Let $A_i = \max|a_i(t)|$. Then we define $Spga = S(A_1, A_2)$ where

$$S(p_1, p_2) = 10 \exp\left\{ -\left[\frac{(p_1 - p_2)}{\min(p_1, p_2)} \right]^2 \right\} \tag{3}$$

The factor of 10 multiplying the familiar function $\exp(-z^2)$ puts the score into a comfortable range of values. Equation 1 has several advantageous features. First, it is monotonically decreasing as the difference between the parameters increases. Optimization programs such as genetic algorithm programs that aim to maximize the similarity can find this a convenient feature. Second, it is symmetrical, in that it gives the same score regardless of whether the accelerogram x or y has the larger peak value. Third, small differences are not penalized too severely. When the difference is only half the value of the parameter itself, the score is about 8 or larger, out of a peak of 10. On the other hand, it is not sensitivie to differences of more than a factor of 2.5; improving the fit from a difference of a factor of 10 to a difference of only a factor of 3 does not yield a significant improvement in the score. This reflects a certain judgement that differences greater than a factor of 2 or 3 should not be considered useful fits for engineering applications.

Table 1. Goodness of Fit Measurements

Number	Symbol	Name: Similarity of	Definition[1]
C1	SDa	Arias duration	$10\left[1-\max\left(F^{IA}(t)\right)\right]$ where $F^{IA}(t)=\left\|N_1^{IA}(t)-N_2^{IA}(t)\right\|$ and $N^{IA}(t)=\frac{I_A(t)}{IA}$
C2	SDe	Energy duration	$10\left[1-\max\left(F^{E}(t)\right)\right]$ where $F^{E}(t)=\left\|N_1^{E}(t)-N_2^{E}(t)\right\|$ and $N^{IE}(t)=\frac{I_E(t)}{IE}$
C3	SIa	Arias intensity	$S(IA_1, IA_2)$ where $IA_i = I_{Ai}(T_d)$
C4	SIv	Energy integral	$S(IE_1, IE_2)$ where $IE_i = I_{Ei}(T_d)$
C5	Spga	Peak Acceleration	$S(A_1, A_2)$ where $A_i = \max\left\|a_i(t)\right\|$
C6	Spgv	Peak Velocity	$S(V_1, V_2)$ where $V_i = \max\left\|v_i(t)\right\|$
C7	Spgd	Peak Displacement	$S(D_1, D_2)$ where $D_i = \max\left\|d_i(t)\right\|$
C8	Ssa	Response Spectra	$mean\left[S\left(SA_1(f_j), SA_2(f_j)\right)\right]$, where the average is over all frequencies at which *SA* is computed in the frequency band being considered.
C9	Sfs	Fourier Spectra	$mean\left[S\left(FS_1(f_j), FS_2(f_j)\right)\right]$, where where the average is over all frequencies at which *FS* is computed by the fast Fourier transform in the frequency band being considered.
C10	C*	Cross Correlation	$C^* = 10\max\left[C(a_1(t), a_2(t)), 0\right]$ where $C(a_1, a_2)=\frac{\int a_1(t)a_2(t)dt}{\left[\int a_1^2(t)dt\right]^{1/2}\left[\int a_2^2(t)dt\right]^{1/2}}$

[1]*Definitions of functions*:

$a_2(t)$, $v_i(t)$ and $d_i(t)$ are an accelerogram and corresponding velocity and displacement, defined for $0 \le t \le T_d$. T_d is the total duration of strong ground motions.	$S(p_1, p_2)=10\exp\left\{-\left[\frac{(p_1-p_2)}{\min(p_1, p_2)}\right]^2\right\}$
$I_{Ai}(t)=\frac{\pi}{2g}\int_0^t a_i^2(\tau)d\tau$	$I_{Ei}(t)=\int_0^t v_i^2(\tau)d\tau$

The score applied to the peak velocity, peak displacement, and the Arias intensity and Energy integral uses the same rationale as the score for peak acceleration. The Arias integral and the energy integral can be used to measure the duration of the accelerograms. Trifunac [10] defined duration as the time it takes the Arias integral to go from 5% to 95% of its final value. However, a more sensitive measure of the similarity of the duration of the seismograms (C1 and C2) is to look at the differences between the normalized integrals, as defined in Table 1. As the example, $N^{IA}(t)$ is monotonically increasing from 0 at the start of the record to 1.0 at the end. If the time series have energy coming in at different times, but the total duration (in Trifunac's [10] sense) is the same, then $N_1^{IA}(t)$ will differ from $N_2^{IA}(t)$, and C1 will have a score less than 10.

The next two criteria are spectral domain criteria. The two spectra considered are the acceleration response (SA) with 5% damping and the Fourier amplitude (FS). Although the two are similar in some ways, they are conceptually completely different. The SA spectrum has asymptotic properties at low and high frequencies that are not shared by FS, and SA tends to be smoother. Thus it is appropriate to develop criteria based on both. As functions of frequency, there is the potential for the spectra from different seismograms to be compared at every one of the frequencies where the spectra are computed. The approach used here, although not unique, is to compare the spectra at every frequency where they are computed, and to find the mean of the scores for each of those individual comparisons for the score for the spectral shape. The response spectrum can be computed at an arbitrary selection of frequencies, and we use a typical selection of frequencies. The Fourier spectra, coming out of an FFT routine, are sampled with uniform frequency spacing. Since reconstruction of the seismogram requires that the Fourier amplitude must be correct at every one of these frequencies, score Sfs (C9) is testing every frequency in the seismogram. This is quite a stringent test, and for seismograms where only the statistical properties are supposed to agree, the Sfs score tends to be lower than others. The final criterion is based on the cross correlation, except that negative values are not allowed.

The scores defined by Table 1 are applicable to any two acceleration time series. However, it is common that data or synthetics are not available over all frequency bands. Data may be limited by instrument response or noise levels, and synthetics may be limited by the methods used to generate them. The scoring system needs a way to accommodate these characteristics. This can be done by two methods: comparing "broadband" accelerograms filtered to allow the entire frequency band of valid data, or comparing in narrower frequency bands. This study considered both.

Table 2 shows the frequency pass bands used for the narrower frequency bands considered. The spacing of the frequency bands in Table 2 are approximately logarithmic, reflecting a weighting by interest and importance. This contrasts with the Fourier spectrum, where spacing is linear. My preferred score, S1, averages the scores of seismograms filtered in each of the valid frequency bands individually, as well as the broadband seismogram. This increases the weight on the lower frequencies, appropriately, as these frequencies are more amenable to waveform fitting and are particularly important for response of large structures. An alternative score, S2, is

obtained by averaging the scores on the ten individual criteria in Table 1 for only the accelerogram filtered to allow all frequencies to pass. Numerical studies find that S1 more fairly represents the quality of the comparison of two seismograms.

Table 2. Frequency bands used

Band	Frequency limits (Hz)
B1	0.05-0.1
B2	0.1-0.2
B3	0.2-0.5
B4	0.5-1.0
B5	1.0-2.0
B6	2.0-5.0
B7	5.0-10.0
B8	10.0-20.0
B9	20.0-50.0
B10	0.05-50.0

CALIBRATION

It is desirable to determine what the different scores mean. The purpose of this calibration is to develop that information. These similarity scores were calibrated in two different ways. The first was with synthetic seismograms. The second was to compare the two horizontal components of recorded accelerograms. When generating synthetic seismograms with the sole purpose of matching the statistical characteristics of data, one would like to do as well as these approaches. When attempting to fit waveforms, especially at lower frequencies, one would like to do better.

Synthetic seismograms

The synthetic seismograms were created using a new implementation of the method of Boore [1]. Pairs of synthetics were generated starting with random numbers, and shaped in the time and frequency domains, within the Matlab programing environment. Each pair was compared as described above. The averages of the 100 different quality of fit parameters for the synthetics are tabulated for later use.

Figure 1 considers the statistics of score S1, based on 1000 pairs of synthetics generated from the same statistical process to approximate a magnitude 8 accelerogram. The distribution of S1 has a mean of 7.35 and standard deviation of 0.40. The tails of the empirical distribution are not well approximated by the normal distribution. Figure 1 also shows the distribution for score S1 for magnitude 7, 6, and 5 synthetics. The mean values for these distributions are 7.16, 6.73, and 6.30 respectively, and the corresponding standard deviations are 0.44, 0.55, and 0.73. The distributions have similar upper limits, but the smaller magnitudes show greater numbers of smaller scores and greater standard

deviations. These may be due to shorter durations of shaking resulting in less stable statistics of peak values.

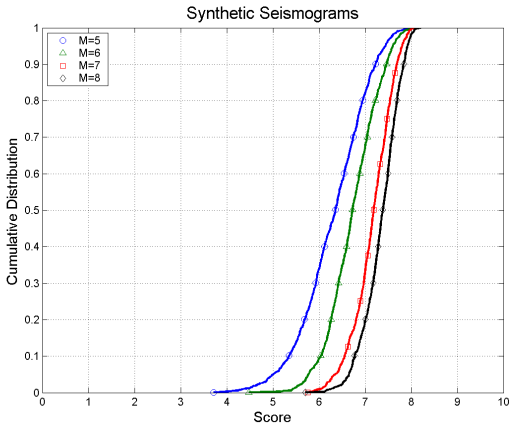

Figure 1. Cumulative distribution of similarity score S1 for pairs of synthetic seismograms generated by a process similar to Boore [1]. Each curve is generated with 800 or more realizations of synthetic seismogram pairs.

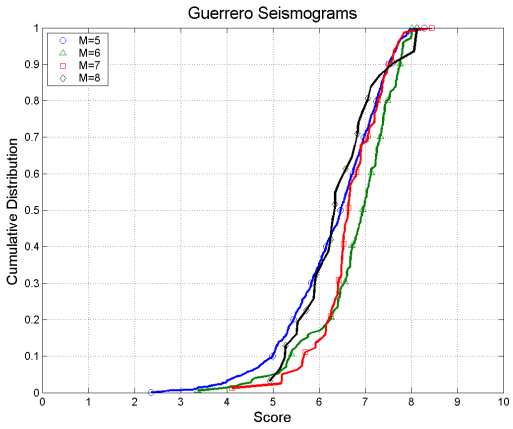

Figure 2. Cumulative distribution of similarity score S1 for east and north components of strong motion accelerograms obtained on rock stations in Guerrero, Mexico (Anderson [12]). These distributions curves are based on 1165 records of magnitude 4.5 or larger.

Guerrero

The Guerrero accelerograph network has been operating since 1985 in a seismic gap in Guerrero, *e.g.* Anderson [11,12]. For this study, we compared the east with the north component of 1165 3-component accelerograms. Figures 2 shows the equivalent of Figure 1, based on the similarity of east and north components for earthquakes in Guerrero. There are 31 comparisons for magnitude 7.5-8.1 (M8 in Figure 2), 81 comparisons for magnitude 6.5-7.49 (M7), 171 comparisons for magnitudes 5.5-6.49 (M6), and 882 comparisons for 4.5-5.49 (M5). For the M8, M7, M6, and M5 categories, the mean values are 6.4, 6.7, 6.8, and 6.3, respectively, and the standard deviations are 0.86, 0.71. 0.90, and 1.04, respectively. For the M6 and M5 categories, the estimated means of the distributions are about the same as for the synthetic data. It appears that there is a weaker and less systematic magnitude dependence than with the synthetic examples, although this conclusion must be moderated because the uncertainties on the estimated distribution parameters are substantial for the large magnitude categories.

From comparing Figures 1 and 2, in each magnitude range the width of the empirical cumulative distribution functions are increased in the Guerrero records. Thus the variability of the horizontal components in actual data is greater than the variability of multiple realizations of synthetics using a common stochastic process. All of the

distributions have upper limits at scores of about 8.0, so with the Guerrero records the increase in the standard deviation of the scores is accompanied by a shift of the mean value to smaller numbers. Examination of the individual criteria shows that the criteria based on peak values have much lower scores in the Guerrero data, but the cross correlation tends to be higher especially in the lower frequency bands, resulting from the physical relationship that is present due to the physics of wave propagation.

Application

Figures 3 and 4 show how these calibrations might be used, for a single station recording the M8.1 Michoacan earthquake. The seismograms compared (Figure 3) are the east and north components from the Caleta de Campos station for the 1985 Michoacan, Mexico, earthquake, Mw=8.1, (Anderson [11]). In Figure 4, each of the criteria is plotted in its own subplot. The comparison of the two seismograms is given by the data points. The averages from the magnitude 8 synthetics are superimposed with the solid line, and the averages from the M8 group of Guerrero seismograms are shown by the dashed line. Study of Figure 4 allows one to identify which parameters fit well, and which fit poorly.

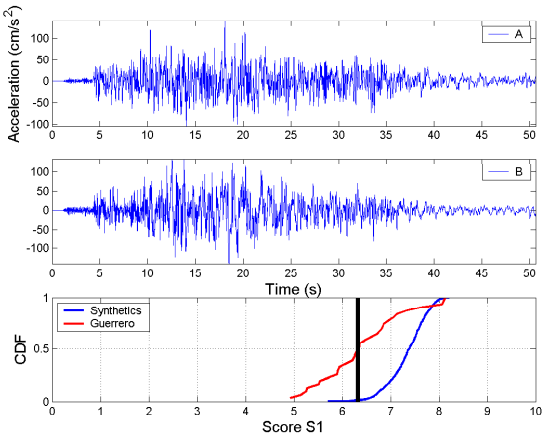

Figure 3. East (A) and north (B) components of strong motion recorded at Caleta de Campos on Sept. 19, 1985, during the Mw8.1 Michoacan, Mexico, earthquake (Anderson [11]). Lower frame shows the comparison of the similarity score S1 for the accelerograms in with the distribution of scores from both synthetics and horizontal data components. These records have an average score compared with other pairs of horizontal components. However, the chances of two synthetics with the same duration being more different from each other is only about 2%.

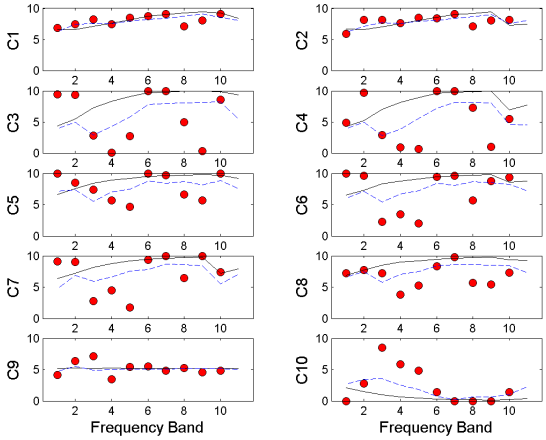

Figure 4. Similarity of the two accelerograms in Figure 3. Each subplot is for a different similarity criterion (Table 1); frequency bands are identified in Table 2. The final score S1 is the mean of all of these individual scores. Solid lines show score of the synthetic seismograms by the same criterion. The dashed line is the equivalent mean score for the similarity of horizontal components of Guerrero accelerograms in the M8 group.

The final score for these two components is S1=6.3. Figure 3 shows the cumulative distribution function of the scores for the two magnitude 8 calibrations. It also shows a vertical line at the score for this accelerogram pair. The conclusion is that these two horizontal components agree about as well as an average pair of horizontal components in the Guerrero data (of which they are a member). However, pairs of synthetics with the same target spectrum typically are more similar than these two; only about 2% have a lower S1 score.

When comparing data and synthetics, in particular, it is important to know not only how closely the two seismograms match, but when they do not match whether the synthetic is predicting higher or lower amplitudes. Therefore, I define the bias in the Fourier spectrum and in the response spectra for each of the frequency bands. For the Fourier spectrum, the bias is defined for each frequency f_i as:

$$b_i = \ln\left(\frac{FS_2(f_i)}{FS_1(f_i)}\right) \tag{4}$$

where FS_1 is the data, and FS_2 is the model. The bias for a frequency band is taken as the mean of b_i over all frequencies in that band. Frequencies outside of the corner frequencies of the filter are not included in the average. The same procedure is used for the acceleration response spectrum. One might expect, on first thought, that a low bias should correspond to a high score. The scoring system could be set up to do that, but that is not how it was done here. Consider for instance a case where the mean bias is zero. Criteria 8 or 9 may still be small if the small bias is the result of model spectra being say three times larger than the data at some of the individual frequencies, and one third of the data at an equal number of other individual frequencies in the frequency band.

BLIND PREDICTION

The Mw 7.9 Alaska earthquake on Nov. 3, 2002, provided a timely opportunity to test predictions of strong ground motion recordings obtained very close to a large magnitude crustal earthquake. The results of this blind prediction exercise are described in detail in a paper by Anderson [13]. Several strong motion sites are located along the route of the Alaska Pipeline which crosses roughly perpendicular to the fault rupture about 85 km east of the epicenter. The closest site is PS10, located about 3 km from the fault. Prior to the release of these data, we conducted a blind prediction experiment to estimate the ground motion waveforms at this closest recording site. Ground motions are computed using, along with other methods, realizations of the stochastic composite source simulation methodology of Zeng [7] with some relatively minor modifications to the source time function. Since the source is stochastic, Zeng prepared 10 realizations of the source model. Predictions were made without accurate knowledge of site conditions or fault-station geometry. The predictions were well publicized before the data were released. The predictions were particularly interesting due to interest in capabilities to predict directivity pulses using synthetic methods.

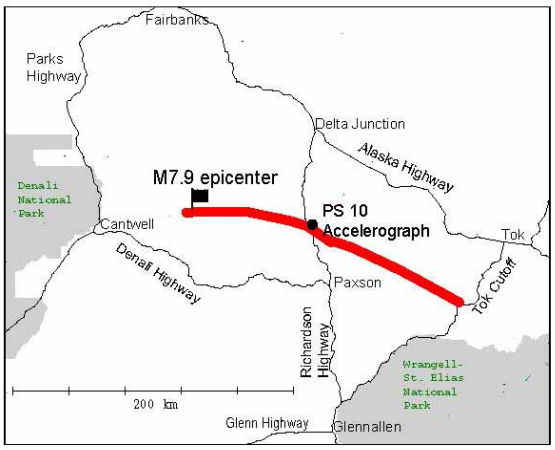

Figure 5. Approximate fault geometry distributed by John Anderson to potential ground motion predictors. Actual rupture proceeded farther to the east, but that is so far from the accelerograph that it does not affect the blind predictions.

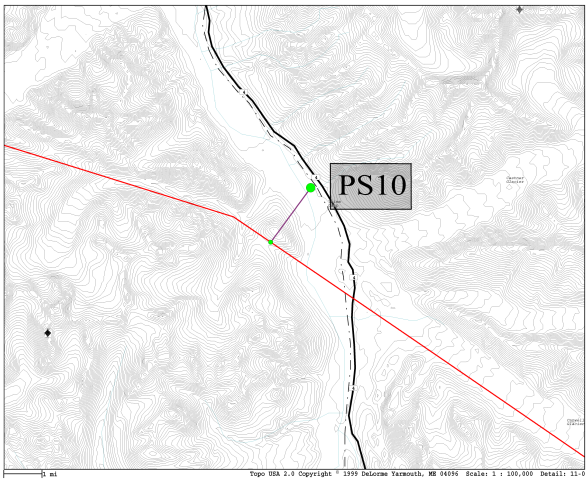

Figure 6. Detail of fault geometry distributed by John Anderson for the area near PS10. The dashed line is the approximate route of the Trans-Alaska Pipeline, which parallels the Richardson Highway. The fault's nearest approach to accelerograph station PS10 (~3 km, as shown) is about 4 km northwest of where the fault crosses the pipeline and the highway. Base map from DeLorme Topo USA.

Figure 5 shows the fault geometry on a large scale, and Figure 6 shows the detail of the location of the fault and the station. Figures 7 to 9 each show, for one of the Zeng models, one of the three components of data, the synthetic velocity, and the comparison of score S1 with the same parameter from synthetics and data used to generate Figures 1 and 2. Data and synthetics are band-pass filetered from 0.1 to 2.0 Hz. The highest scoring of Zeng's ten models is intentionally selected for Figures 7-9. These models do not attempt to match any teleseismic slip functions. The range of scores among the multiple realizations for each component is a little more than one unit of the abscissa. Figure 10-12 shows the fitting scores, equivalent to Figure 4, but in this case the points are the average score of the 10 realizations. Figure 13 shows the mean bias of the ten realizations.

Figures 7-9 shows that the velocity at PS10 is dominated by a strong pulse at the time that the rupture passes the station. The model by Zeng is able to predict that the data would show a strong, compact pulse with high velocities, and to predict the peak velocities within about 30%. The model seems to overestimate the period of the actual pulse slightly, and in detail the observed pulse has a different polarization. The scoring system indicates that on the fault normal and vertical components, the model is as similar to the data as one can expect for the similarity of two randomly-oriented components.

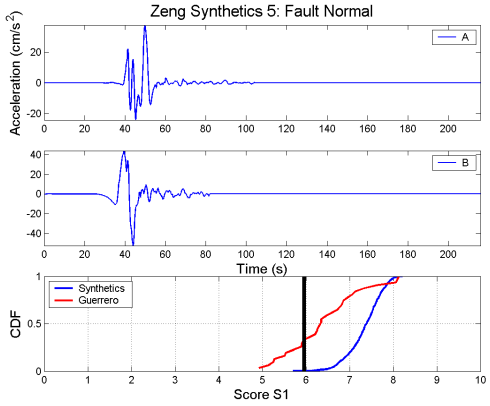

Figure 7. Top, filtered data, rotated into the fault normal orientation and integrated to velocity. Center, prediction by Zeng. Bottom, compares the S1 similarity score of the Zeng synthetic with data to cumulative distribution functions from Figures 1 and 2.

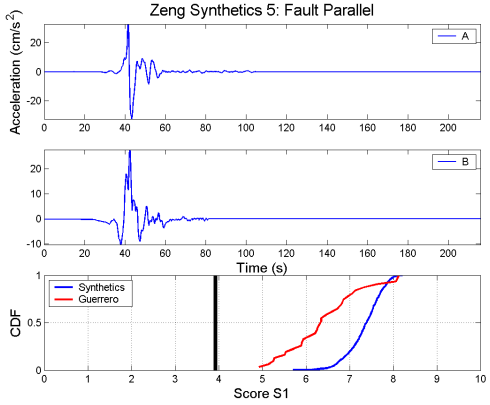

Figure 8. Equivalent to 7 for the fault parallel component of ground motions.

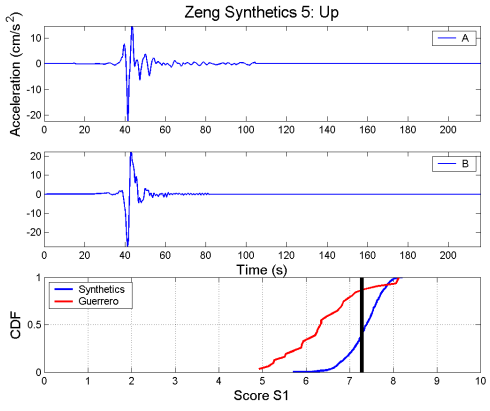

Figure 9. Equivalent to 7 for the vertical component of motion.

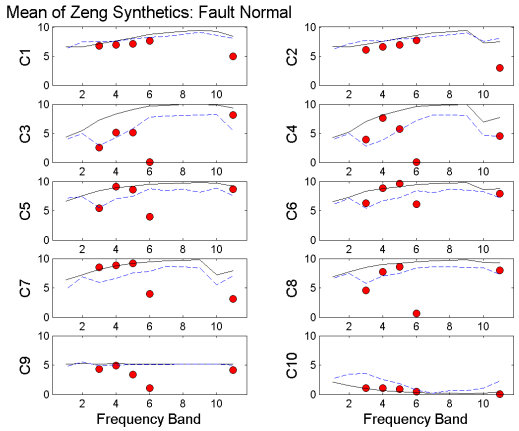

Figure 10. Breakdown of mean similarity scores for the fault-normal component, from 10 realizations of synthetic seismograms. Fig. 7 shows one of the ten. Symbols are as in Figure 4. Frequency band 11 is 0.1 to 2.0 Hz.

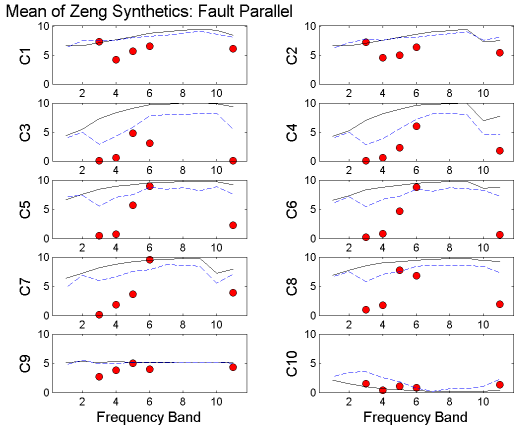

Figure 11. Equivalent of Figure 10 for the fault parallel component.

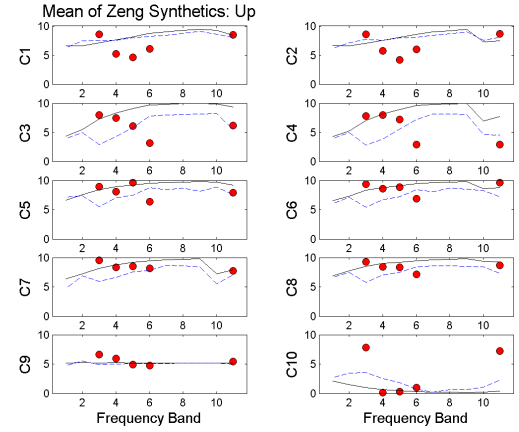

Figure 12. Equivalent of Figure 10 for the vertical component.

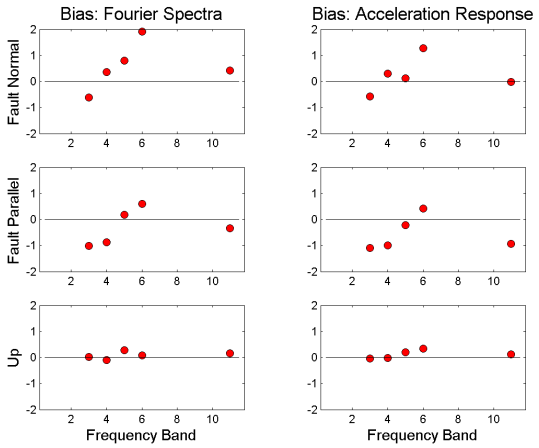

Figure 13. Bias of the mean of the ten Zeng realizations relative to the data.

CONCLUSIONS

The establishment of a widely-accepted scoring system for the comparison of synthetic seismograms with observations will be a major benefit for the engineering seismology community. Once accepted, it will give us a means to increase the credibility, and to document our improved capabilities, in the process of generating realistic synthetic ground motions for engineering applications.

Independence and efficiency are desirable characteristics of a set of parameters to characterize ground motions. The characteristics are related. Independent parameters are not inherently correlated with each other. An efficient set would be a minimum set of parameters that provides the same information as the full set. The scoring system here has not optimized for either of these characteristics. It is in fact easy to recognize that the parameters used here are not independent, and thus efficiency is also sacrificed. In the context of this general approach, such optimization would be possible, and should perhaps be done. On the other hand, it is convenient to be able to look at figures like Figures 10-12 and see at a glance which characteristics of the seismograms do match well, and which do not.

If this score system were to be adopted for synthetic seismogram testing, then an important question is how good is good enough? This will depend on the application. For the generation of synthetics for some engineering applications, where the phase of the seismograms is not critical, then a similarity as good as the two horizontal components is probably good enough. On the other hand, if the phase is critical, as for instance in fault-normal pulses associated with rupture passing the station, then one would hope to achieve high correlation coefficients at the low frequencies in addition to average scores in other categories.

ACKNOWLEDGEMENTS

Part of this research was completed while the author was on a sabbatical leave as a visiting scientist at the Lamont-Doherty Earth Observatory. The support and warm welcome from the staff at Lamont is gratefully acknowledged. Thanks also to Alyeska Pipelines Services for making the strong motion data available to the engineering seismology community, and to Jim Roddick for serving as a host for me to see accelerograph stations at Pump Stations 9 & 10, and the fault crossing. This research was also in part 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 767. This material is also based in part upon work supported by the National Science Foundation under Grant No. 0000050, which partly supports the Guerrero accelerograph network and associated research on data from that network. The Guerrero accelerograph network also receives support from the Instituto de Ingenieria at Universidad Nacional Autonoma de Mexico, Mexico City.

REFERENCES

1. Boore, D. M. (1983). Stochastic simulation of high-frequency ground motions based on seismological models of the radiated spectra. *Bull. Seism. Soc. Am.* **73**, 1865-1894.
2. Hartzell, S. H. (1978). Earthquake aftershocks as Green's functions. *Geophys. Res. Lett*. **5**, 1-4.
3. Irikura, K. (1983). Semi-empirical estimation of strong ground motions during large earthquakes. *Bull. Disaster Prevention Res. Inst. (Kyoto Univ.)* **33**, 63-104.
4. Irikura, K. (1986). Prediction of strong acceleration motions using empirical Green's functions. *Proc. Seventh Jap. Earthq. Eng. Symp.*, 151-156.
5. Hadley, D. M. and D. V. Helmberger (1980). Simulation of strong ground motions. *Bull. Seism. Soc. Am.* **70**, 617-630.
6. Somerville, P., M. Sen and B. Cohee (1991). Simulation of strong ground motions recorded during the 1985 Michoacan, Mexico and Valparaiso, Chile earthquakes. *Bull. Seism. Soc. Am.* **81**, 1-27.
7. Zeng, Y., J. G. Anderson and G. Yu (1994). A composite source model for computing realistic synthetic strong ground motions. *Geophys. Res. Lett*. **21**, 725-728.
8. Anderson, J. G. (2004). A Criterion For Evaluating Goodness Of Fit Of Accelerograms, in preparation.
9. Arias, A. (1970). A measure of earthquake intensity, in R. J. Hansen, ed. Seismic Design for Nuclear Power Plants, MIT Press, Cambridge, Massachusetts, pp438-483.
10. Trifunac, M. D. and A. G. Brady (1975). A study of the duration of strong earthquake ground motion, *Bull. Seism. Soc. Am.* **65**, 581-626.
11. Anderson, J. G., P. Bodin, J. Brune, J. Prince, and S. K. Singh (1986). Strong ground motion and source mechanism of the Mexico earthquake of September 19, 1985. *Science* **233**, 1043-1049.

12. Anderson, J. G., J. N. Brune, J. Prince, R. Quaas, S. K. Singh, D. Almora, P. Bodin, M. Onate, R. Vazquez and J. M. Velasco (1994). The Guerrero accelerograph network, Geofisica Internacional 33, No. 3, 341-371.
13. Anderson, J. G., Robert Graves, Yuehua Zeng, and Paul Somerville (2004). Blind Prediction of Near-Fault Strong Ground Motions, in preparation.