QUANTIFYING CISN SHAKEMAP UNCERTAINTY

Kuo-wan Lin[1], David J. Wald[2], Bruce Worden[3], and Anthony F. Shakal[1]

[1]California Geological Survey, Sacramento, CA
[2]U. S. Geological Survey, Golden, Colorado, CO
[3]U. S. Geological Survey, Pasadena, California, CO

Abstract

Efforts underway to quantify uncertainties associated with ShakeMap ground motions through efforts by the California Integrated Seismic Network (CISN) ShakeMap Working Group are discussed. There are multiple sources of uncertainty in producing a ShakeMap, including sparse ground motion measurements, approximate representation of fault finiteness and directivity, empirical ground motion predictions, numerical interpolation, and site corrections. These ground motion uncertainty measures are critical for evaluating the range of possible losses and allow users to associate appropriate levels of confidence when using rapidly produced ShakeMaps as part of their post-earthquake critical decision making process. We quantify the uncertainties of the maps on a point-by-point basis, by combining the separate, but related, contributions of uncertainty for each ShakeMap parameter as a function of location on the map. Finally, we show examples of results of estimates of uncertainty for ShakeMap for earthquakes in California with/without defined fault traces. We discuss future developments and plans for integration of these uncertainty measures, both quantitative and qualitative, into the online system and user interfaces of ShakeMap.

Introduction

The accuracy of a given ShakeMap varies spatially over the map area and depends on a number of contributing factors. However, it is usually dominated by two aspects: 1) spatial variability of peak ground motions near recording stations (and thus, station density), and 2) the aleatory uncertainty associated with empirical ground motion estimation relationships used to estimate the shaking between stations.

In this study, we consider these two sources of variability in estimating ShakeMap uncertainty. Spatial variability of peak ground motions can be generalized in the form of a rapidly increasing variability with increasing distance from the nearest station. Aleatory variability, in contrast, is more complicated and becomes more significant as the earthquake fault dimensions get larger (about M5.5 and greater), particularly when the fault location and dimensions are not yet ascertained. Without an accurate representation of the fault rupture geometry, the appropriate distance to a particular location—which is needed when using a forward ground motion prediction equation—is poorly constrained. Not knowing the true distance to the fault rupture contributes significant uncertainty, particularly in the near-fault region.

Our goal in quantifying ShakeMap uncertainty is to produce a grid of latitude and longitude pairs which contains not only the various peak ground motion parameters at each point, but also contains the variance at that point for each ground motion parameter. This grid could also be converted to an overall *qualitative* assignment of ShakeMap accuracy, an issue we will be addressing in the near future.

In the meantime, methods of estimating uncertainty in three end member cases are described for generating a map of ShakeMap uncertainty values at each grid point.

Case 1. Small To Moderate Sized Earthquake, Suitable For A Point Source Representation

When a grid point is near to a station (~10 km or less) uncertainty is controlled by proximity to that station and the variability can be quantified by the model of Boore et al. (2003):

$$\sigma^2_{\Delta \log Y} = \sigma^2_{indobs}(1+\frac{1}{N})F(\Delta)^2 \text{ and } F = 1 - e^{-\sqrt{0.6\Delta}}, \qquad (1)$$

where $\sigma_{\Delta \log Y}$ is the standard deviation of differences in the logarithm of the peak motion Y (e.g., acceleration), σ_{indobs} is the standard deviation of an individual observation about a regression, and N is the number of recordings used in averaging the peak motion of a group of recordings in a small area. $F(\Delta)$ is a function that models the spatial correlation of the motion, where Δ is the distance between the two points. For this study we assume that N is large enough so that the *1/N* term can be neglected. Thus, the predicted spatial variability in ground motion reduces to zero as the distance between a grid point and the nearest station decreases to zero as shown in Figure 1 (from Boore et al., 2003). With a large grid point to station distance, the spatial variability in ground motion approaches the standard deviation of the regression model. The cut-off distance for computing spatial variability in ground motion is set at 10 km in this study.

For greater distances than 10 km, we use the Boore et al. (1997) ground motion prediction equation's total aleatory uncertainty ($\sigma_{Aleatory}$).

$$\sigma_{Aleatory} = \sqrt{\sigma^2_{Interevent} + \sigma^2_{Intraevent}} \qquad (2)$$

With several ShakeMap data points (i.e., station amplitudes), we can remove a bias term between the ground motion predictions and the data, thereby removing the inter-event term. However, when no data are available, no event-specific bias correction can be made and both the intra- and inter-event terms contribute (Table 1).

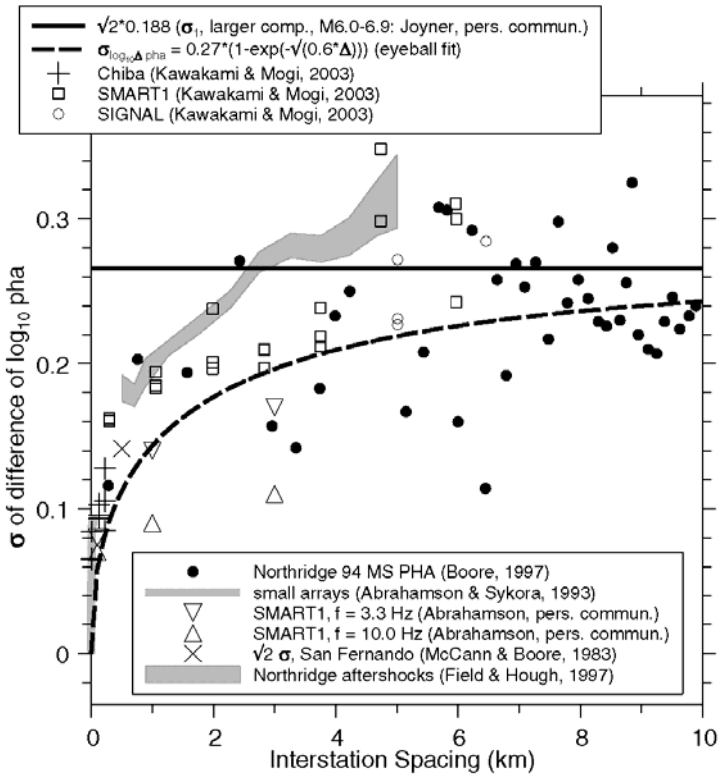


Figure 1. Standard deviation of difference of the largest peak horizontal acceleration as a function of interstation spacing. F(Δ) given in equation 1 is shown by the curve. Data include Northridge earthquake strong motions (Boore, 1997) and previous studies as indicated (figure from Boore et al., 2003).

Case 2. Large Earthquake, And Fault Rupture Geometry And Dimensions Are Not Known

For earthquakes of magnitude 5.5 and larger, the fault dimension affects the measure of distance from the fault to the site of interest. When employing the Joyner-Boore distance measure used for forward ground motion estimation, the fault rupture dimension must be known, since the Joyner-Boore distance is defined as the closest distance from a site to the surface projection of the fault rupture.

If necessary, initial ShakeMaps are produced without knowledge of the rupture dimensions. Again, the uncertainty is generally low near the seismic stations, but at some distance from the stations it is constrained only by the predictions using a ground motion attenuation relationship. In this case, distance adjustments are made to convert the point source (epicentral) distance to the appropriate Joyner-Boore distance for the ground motion attenuation model used. We also must adjust the aleatory uncertainty. We adopt the results and the approach defined in EPRI (2003), in which the distance adjustment is determined for the case where the rupture orientation is assumed to be uniformly distributed in azimuth from 0 to 360 degrees and for a mixture of strike-slip and reverse ruptures using random epicenters. For each simulated rupture, EPRI (2003)

i) Computed the appropriate distance measure and corresponding median ground motion parameter,
ii) Considered the geometric mean of all these simulation values to be the median ground motion for that epicentral distance and magnitude,
iii) Inverted the median ground motion to find the distance that corresponds to that median ground motion value,
iv) Determined a distance adjustment factor for each epicentral distance, magnitude, and ground motion parameter, and
v) Fit these distance adjustment factors with a functional form, and provided the necessary coefficients in a series of look up tables.

Using the distance correction factor then simply entails employing these distance adjustment relationships (EPRI, 2003) that translate epicentral to the equivalent Joyner-Boore distance:

$$r_{Joyner-Boore} = r_{Epicentral} \times \{1 - 1/\cosh(C_1 + C_2(M-6) + C_3 \ln(r'))\}, \quad (3)$$

where $r' = \sqrt{r_{Epicentral}^2 + h^2}$ (4), and $h = e^{C_4 + C_5(M-6)}$, (5)

$r_{Joyner-Boore}$ is the Joyner-Boore distance, $r_{Epicentral}$ is the epicentral distance, M is the magnitude of the earthquake, and C_1 to C_5 are model coefficients (which vary by ground motion model and seismic frequency) given in Table 2.

Hence, when the fault geometry and orientation is not known, a mean value of ground motion at each point is provided rather than the simple epicentral distance-based estimation. While the latter approach is currently used for ShakeMap, it tends to underestimate ground motions near a finite fault (since it is the maximum possible source-station distance) rather than providing a mean value based on random fault geometry and epicenter. Hence, we will be adopting these distance adjusted ground motions for ShakeMap calculations in the near future.

The variability associated with this approach is also derived in EPRI (2003). The variability in the median ground motion due to the randomness in epicenter location and rupture orientation was used to compute a ground motion standard deviation, and we employ their equations to compute the additional component of aleatory uncertainty:

$$\sigma_{AdditionalPointSource} = e^{C1+C2(M-6)+C3(M-6)^2} \times [1 - 1/\cosh(f_a)] \times 1/\cosh(f_b), \quad (6)$$

$$f_a = e^{C4+C5(M-6)} + e^{C6+C7(M-6)} \times r_{Epicentral}, \quad (7)$$

$$f_b = e^{C8+C9(M-6)} \times \ln(r'/h), \quad (8)$$

$$r' = \sqrt{r_{Epicentral}^2 + h^2}, \; h = e^{C10+C11(M-6)}, \quad (9)$$

where $\sigma_{AdditionalPointSource}$ is the point source aleatory, $r_{Epicentral}$ is the epicentral distance, M is the magnitude of the earthquake, and C_1 to C_{11} are the model coefficients (see Table 3).

We can then combine this additional point source variability (Equation 6) with that associated with the prediction equation (Equation 2):

$$\sigma_{Total} = \sqrt{\sigma^2_{Aleatory} + \sigma^2_{AdditionalPoint Source}} \tag{10}$$

Again, if at any time a grid point is closer to a station than 10 km, the variability associated with that grid to station distance controls the uncertainty; at greater distances, the above relation is employed (Figure 2).

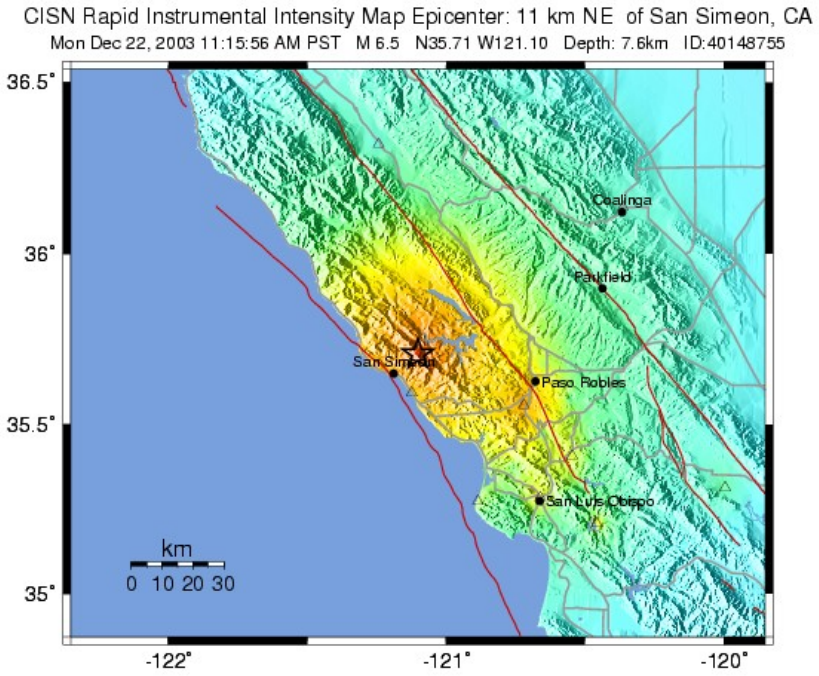
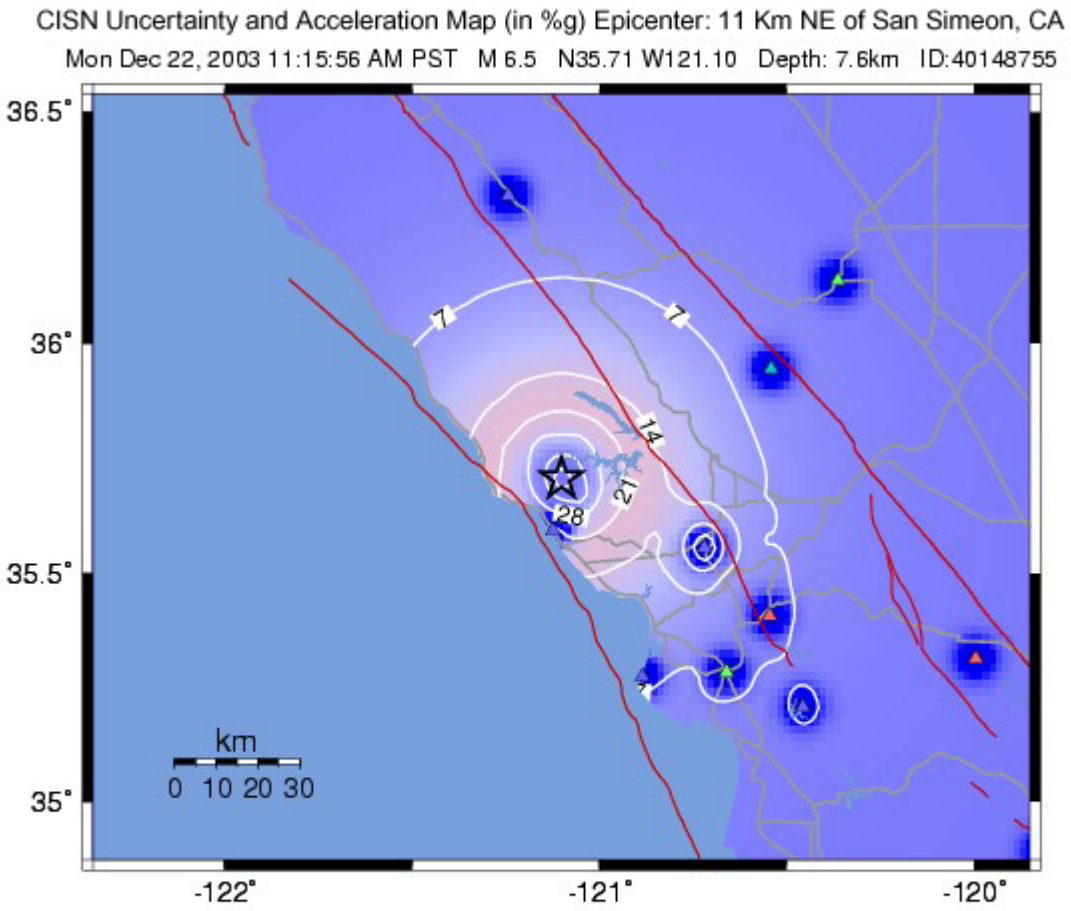


Figure 2a. ShakeMap for the magnitude 6.5 San Simeon, California earthquake of 2003, Case 2. Note there are few seismic stations.



Quality	A	B	C	D	E
Accuracy	Very High	High	Moderate	Low	Very Low
Uncertainty (Sigma) Log_{10} Values	0.0	0.1	0.2	0.3	0.5

Figure 2b. Uncertainty map, prior to adding finiteness, indicating linear units of standard error above the aleatory uncertainty (red areas) and below (dark blue, near stations). The ring of large uncertainty around the epicenter is a result of early uncertainty of the fault location and thus the uncertainty for computing distances to the fault.

Case 3. Large Earthquake, And Fault Finiteness Is Constrained

After a description of the source becomes available—usually from some combination of source modeling, aftershock patterns, or surface rupture—we can fall back on the approach of Case 1, since the appropriate Joyner-Boore distance measure from the fault location can be determined. As in case 1, uncertainty is determined by the grid-to-station proximity uncertainty (Equation 1) or if there are no nearby observations, the uncertainty associated with the predictive relationship (Equation 2). Typically by this time, enough seismic stations are available that a bias between the data and the ground motion estimates can be computed and thus the inter-event uncertainty term is dropped (Figure 3).

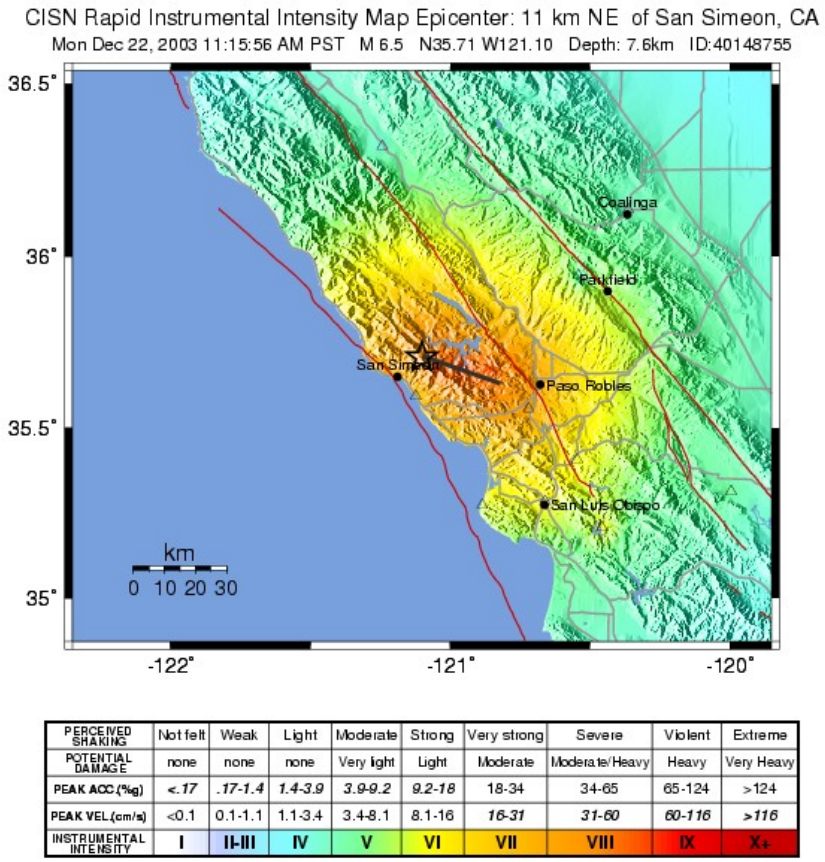


Figure 3a. ShakeMap for the magnitude 6.5 San Simeon, California earthquake of 2003 with fault finiteness imposed, Case 3 (line source shown as a grey line).

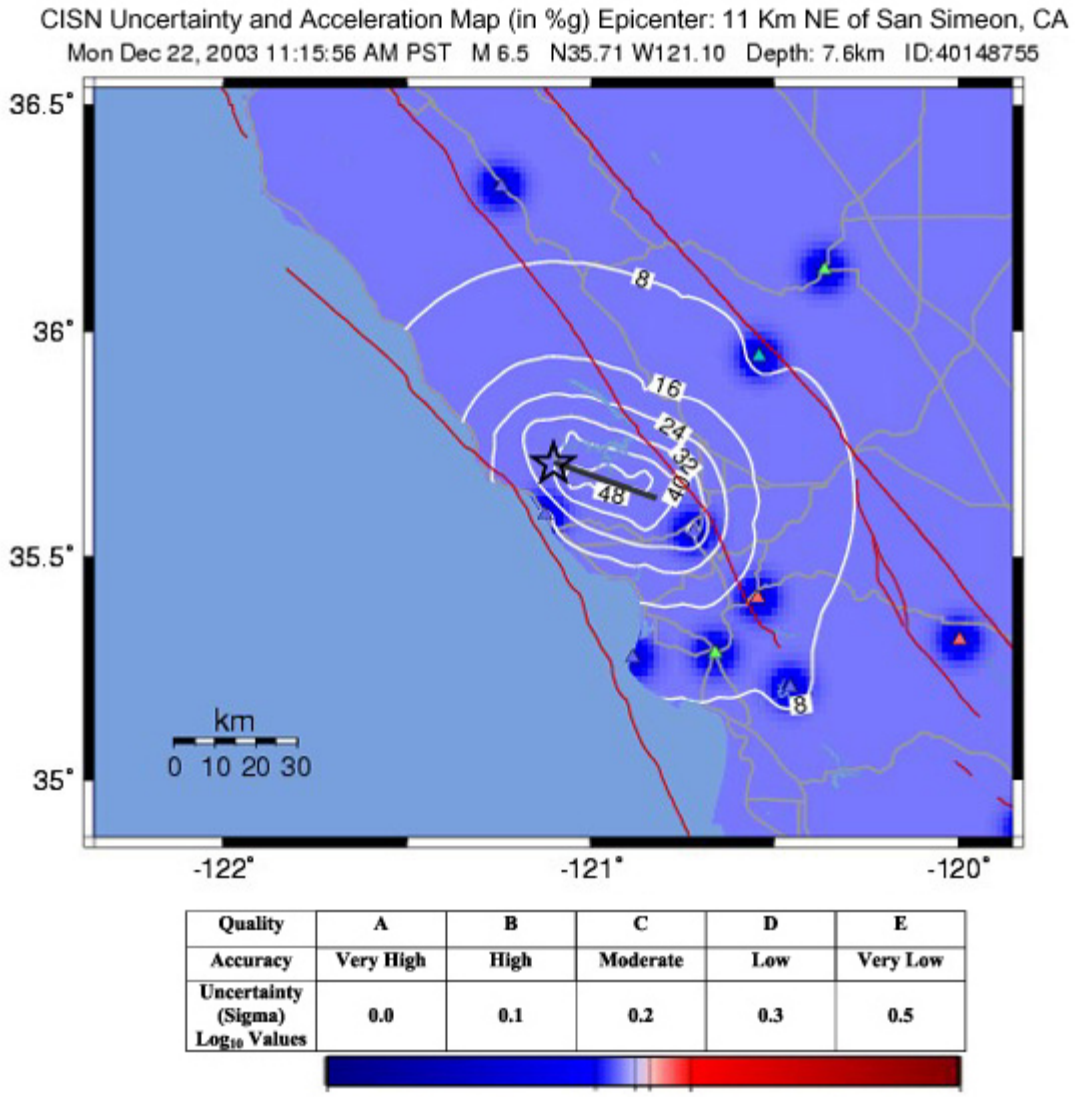


Figure 3b. Uncertainty map, after adding fault finiteness, indicating linear units of standard error above the aleatory uncertainty (red areas) and below (dark blue, near stations).

Analysis of Spatial Variability of PGA for the M6 Parkfield Earthquake

In order to validate the effectiveness of the method for quantifying CISN ShakeMap uncertainty, we analyzed the spatial variability of PGA data from the 28 September 2004 Parkfield earthquake and compared differences between observed and estimated PGA values. The Parkfield array data recorded by the CISN during the M6 earthquake provided the highest density of recording stations in the near-fault region of any earthquake recorded to date (e.g., Langbein et al., 2005; Shakal et al., 2005). A total of 56 stations were located within 20 km of

the fault; 48 were within 10 km of the fault, more than for many other earthquakes combined. The distances between stations ranged from <1 to 40 km. It is of interest to study the spatial variability of peak ground motions across the Parkfield array and to examine dependence of variability on nearest distance-to-fault and on sensor orientation. To do this, we have followed the analysis approach of Boore et al. (2003).

The distances between all possible stations pairs were calculated and sorted in increasing order. These sorted pairs were grouped into bins with 15 station pairs per bin, and over station-pair spacings of up to 20 km. For each bin we computed mean station spacing Δ and standard deviation $\sigma(\Delta)$ of differences in logarithm of PGA. The spatial variability of standard deviation for the Parkfield earthquake (Figure 4) is consistently higher at most station spacings than that for both the Northridge and San Simeon earthquakes, and higher (by a factor of ~1.3) than the standard deviation for the regression model. The standard deviation value increases rapidly from ~0.2 at 1.2 km station spacing, to ~0.3 at 2.0 km station separation.

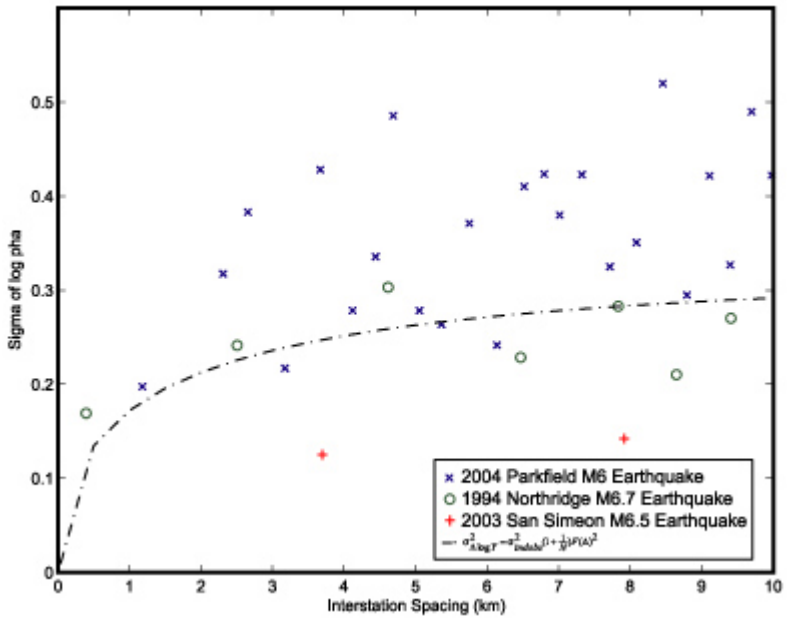


Figure 4. Standard deviation of difference of the largest peak acceleration as a function of interstation spacing. $F(\Delta)$ given in equation 1 is shown by the curve. Data include strong motion records from the Parkfield, Northridge, and San Simeon earthquakes.

To examine any dependence of the variability on sensor orientation, we divided the data to contain only East-West or North-South components and repeated the same procedure. For dependence on nearest distance-to-fault, the data was divided into near-fault/away-from-fault stations with a cut-off nearest distance-to-fault of 12 km. Figure 5 shows the results for spatial variability in standard deviations for both analyses.

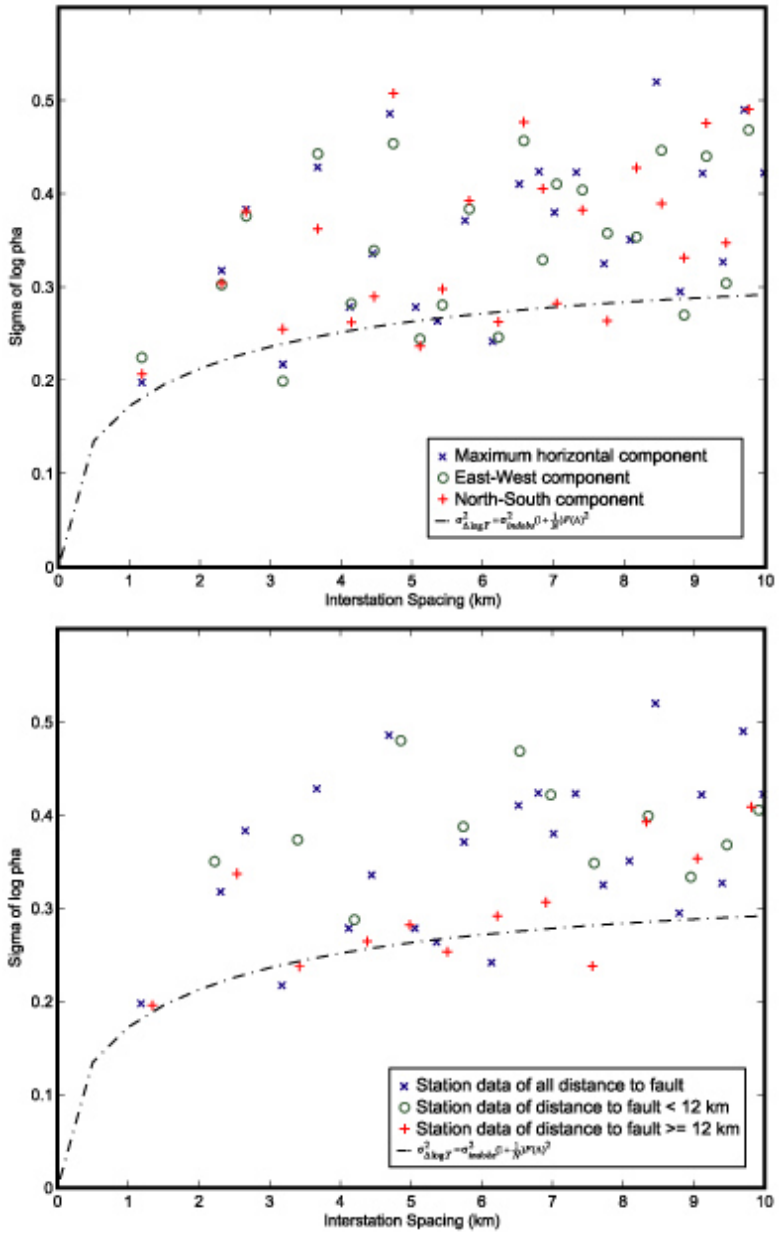


Figure 5. Standard deviation of difference of log of the peak acceleration as a function of interstation spacing for the Parkfield earthquake. (Top) Standard deviation for data with either East-West of North-South component of the strong motion data. (Bottom) Standard deviation for data either within or beyond a distance of 12 km to nearest fault.

We found that spatial variability for the Parkfield data with respect to the orientation of recording components (Figure 5) shows similar results to standard deviations obtained with analysis using the larger of the horizontal components. Contrary to direction insensitivity of spatial variability, we observed better correlation of peak ground motions for distant station pairs than for the close station pairs (Figure 5). The estimated variability of peak ground motions for distant station pairs approximates the standard deviation of the regression model.

The results of spatial variability for the Parkfield array data represent a special case compared with other earthquake data sets because Parkfield has many near-fault stations and relatively few at distance. The highly variable standard deviations in the near-fault data can be attributed to the complexity of near-fault ground motions, which are not accounted for by the regression model and will require further investigation.

We also compared observed peak values of strong ground motion for the Parkfield earthquake with estimated peak values to analyze the extent of data misfit for the two regression models currently used by the CISN ShakeMap, the HazusPGV (Southern California) and the Large_Seg (Northern California). (Note that the peak values of model estimates were not adjusted for site condition and do not coincide with the results of the CISN ShakeMap.) Both regression models use forms similar to the regression model of Boore et al., 1997 (BJF97). The models generate PGA, PGV, and SA for a given magnitude and distance.

Figure 6 shows results of comparison of the observed PGA recorded by the CISN network and the estimated values, based on the Southern California and Northern California regression models, for the 2004 Parkfield earthquake. The estimated PGA values from both regression models are consistently higher than the observed values at low PGA range (less than ~0.2g). However, the estimated values are lower than the observed values at high PGA range (greater than ~0.5g).

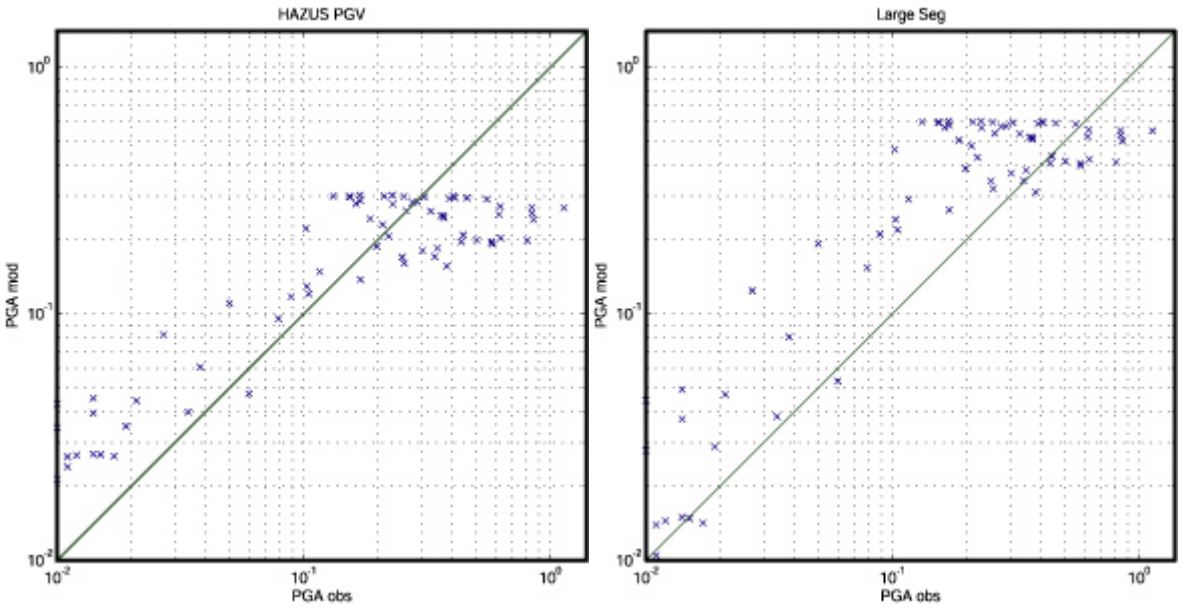


Figure 6. Comparison of the PGA recorded by the CISN network and the estimated values based on either (left) Southern California or (right) Northern California regression model for the 2004 Parkfield earthquake.

Summary

Quantifying ShakeMap uncertainty is an ongoing development at CISN as growing expectations of the use of the maps continue. The next version of ShakeMap (V3.1) is nearing

release, and should be available in mid-2005. The new features in the release include the underlying code for quantification of uncertainty, for testing and feedback, and the implementation of an XML file for the distribution of the gridded ShakeMap data.

The three end member cases described here will be used in quantifying ShakeMap uncertainty: (1) small earthquake of M5.5 or less as point source representation, (2) large earthquake without known fault rupture geometry and dimensions, and (3) large earthquake with fault finiteness. Using one of these methods, the goal is to produce a grid of latitude and longitude pairs containing peak ground motion parameters and uncertainty at each point.

Results from a study of spatial variability of the Parkfield earthquake data indicate that the variability is generally direction insensitive and approximately corresponds with model predictions for distant stations. In the near-fault area the variability is complex and cannot be accounted for by the standard deviation from the regression model.

References

Boore, D. M. (1997). Estimates of average spectral amplitudes at FOAKE sites, Appendix C in An evaluation of methodology for seismic qualification of equipment, cable trays, and ducts in ALWR plants by use of experience data, K. K. Bandyopadhyay, D. D. Kana, R. P. Kennedy,and A. J. Schiff (Editors), U.S. Nuclear Regulatory Commission NUREG/CR-6464 and Brookhaven National Lab BNL-NUREG-52500, C-1-C-69.

Boore, D. M., W. B. Joyner, and T. E. Fumal, (1997). Equations for Estimating Horizontal Response Spectra and Peak Acceleration from Western North American Earthquakes: A Summary of Recent Work, *Seismological Research Letters*, Volume 68, Number 1, January/February 1997.

Boore, D M., Gibbs, J. F., Joyner, W. B., Tinsley, J. C., and Ponti, D. J (2003). Estimated Ground Motion From the 1994 Northridge, California, Earthquake at the Site of the Interstate 10 and La Cienega Boulevard Bridge Collapse, West Los Angeles, California, *Bull. Seism. Soc. Am*, 93, 6, 2737-2751.

EPRI (2003). CEUS Ground Motion Project: Model Development and Results, EPRI Report 1008910, EPRI, Palo Alto, CA, 105 pp.

Field, E. H., and S. E. Hough, (1997). The variability of PSV response spectra across a dense array deployed during the Northridge aftershock sequence, *Earthquake Spectra*, **13**, 243-257.

Hok, S., and D. J. Wald, (2003). Spatial Variability of Peak Strong Ground Motions: Implications for ShakeMap Interpolations, *EOS. Trans. AGU*, 84(46), F1121.

Langbein, J., R. Borcherdt, D. Dreger, J. Fletcher, J.L. Hardebeck, M. Hellweg, C. Ji, M. Johnston, J.R. Murray, R. Nadeau, M.J. Rymer, and J.A. Treiman, (2005). Preliminary report on the 28 September 2004, M 6.0 Parkfield, California earthquake, *Seism. Res. Lett.*, **76**, 1, 10-26.

Shakal, A., V. Graizer, M. Huang, R. Borcherdt, H. Haddadi, K. Lin, C. Stephens, and P. Roffers, (2005). Preliminary analysis of strong-motion recordings from the 28 September 2004 Parkfield, California earthquake, *Seism. Res. Lett.*, **76**, 1, 27-39.

	PGA	PGV	PSA03	PSA10	PSA30
HazusPGV (SC)	0.2258	0.2423	0.2267	0.2662	0.2918
Large_Seg (NC)	0.3606	0.3286	undef	undef	undef

Table 1. Standard deviation (aleatory) values (log10) for both Northern and Southern California regression models for the CISN ShakeMap.

Frequency (Hz)	C_1	C_2	C_3	C_4	C_5
0.5	-0.4098	-1.394	1.003	1.235	1.421
1.0	-0.4060	-1.394	1.003	1.237	1.424
2.5	-0.4066	-1.394	1.003	1.235	1.426
PGA	-0.4517	-1.394	1.003	1.239	1.431

Table 2. Distance adjustment from epicentral to Joyner-Boore distance coefficients for equations 3 and 5 (random epicenters) [from Table 3-38 of EPRI (2003)].

Frequency (Hz)	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	C_{11}
0.5	-1.502	0.5506	-0.03874	-0.8330	-0.01935	-1.341	-0.6375	-0.1008	0.3328	1.564	1.635
1.0	-1.604	0.6415	-0.05674	-0.8626	-0.01209	-1.177	-0.7274	-0.1472	0.4290	1.722	1.635
2.5	-1.430	0.5386	-0.03777	-0.7968	-0.04394	-1.378	-0.6413	-0.1241	0.3472	1.607	1.630
PGA	-1.407	0.5926	-0.05345	-0.8708	-0.001605	-1.305	-0.7161	-0.1846	0.3675	1.599	1.629

Table 3. Additional aleatory variability (random epicenters) coefficients for equations 6-9 [from Table 3-42 of EPRI (2003)].