Bulletin of the Seismological Society of America, Vol. 95, No. 6, pp. 2040–2057, December 2005, doi: 10.1785/0120050017

Model Uncertainties of the 2002 Update of California Seismic Hazard Maps

by Tianqing Cao, Mark D. Petersen, and Arthur D. Frankel

Abstract In this article we present and explore the source and ground-motion model uncertainty and parametric sensitivity for the 2002 update of the California probabilistic seismic hazard maps. Our approach is to implement a Monte Carlo simulation that allows for independent sampling from fault to fault in each simulation. The source-distance dependent characteristics of the uncertainty maps of seismic hazard are explained by the fundamental uncertainty patterns from four basic test cases, in which the uncertainties from one-fault and two-fault systems are studied in detail. The California coefficient of variation (COV, ratio of the standard deviation to the mean) map for peak ground acceleration (10% of exceedance in 50 years) shows lower values (0.1–0.15) along the San Andreas fault system and other class A faults than along class B faults (0.2–0.3). High COV values (0.4–0.6) are found around the Garlock, Anacapa-Dume, and Palos Verdes faults in southern California and around the Maacama fault and Cascadia subduction zone in northern California.

Introduction

Studying the uncertainty and sensitivity of seismic source parameters is important to better understand and use seismic hazard maps. In this article we discuss the uncertainty associated with the California portion of the U.S. National Seismic Hazard Maps (Frankel *et al.*, 2002; Petersen *et al.*, 1996) that are applied in current building codes, insurance rate structures, and the public policy decision-making process. From both scientific and user points of view, the uncertainty estimates are an inseparable part of the seismic hazards that are as important as the other hazard products and should provide valuable insights for applying the maps. For example, it is useful to provide uncertainty maps to the hazard map developers and users that describe our confidence in the mean hazard estimates. In addition, the uncertainty and sensitivity information can guide the earth science and engineering communities in determining directions of research that will reduce hazard uncertainties. These products have a direct impact on public policy decisions and it is critical that this uncertainty information is available.

In general, we divide the uncertainties into two types: (1) model or epistemic, which is the basis of a logic tree formulation, and (2) random or aleatory, which is included in the hazard models directly. Model uncertainty (Reiter, 1990; Cao *et al.*, 1996) is the knowledge-based variability (Cramer, 2001a) of the seismic hazard model. The model uncertainty is related to how the empirical data are characterized and include, for example, the use of several different attenuation relations and fault rupture area–magnitude relations. A logic tree is a methodology for quantifying the knowledge-based uncertainty by systematically stepping through the decisions needed for developing a hazard map and tracking the results. These decisions are represented by a series of decision points (nodes) where one may choose from several weighted alternatives (branches). If all the potential branches from each node are selected, the terminations of the logic tree branches represent a suite of viable weighted hazard models that depict the uncertainty in hazard from our lack of knowledge regarding the future earthquake sources and associated ground shaking. Random uncertainty is also called aleatory uncertainty and is the natural variability in earthquake sources and ground-shaking parameters. Aleatory uncertainties are considered explicitly in the hazard calculations (Reiter, 1990; Cao *et al.*, 1996) via numerical integrations.

In this uncertainty analysis, the number and nature of the nodes and branches make it difficult to sample the entire logic tree. Therefore, we adopt a Monte Carlo approach of randomly sampling the California seismic hazard logic tree. This sampling method only provides the sample, not the population uncertainty. We can make the inference from sample to population uncertainty through increasing the number of iterations. Monte Carlo random sampling has been shown to yield similar results compared with completely sampling the logic tree (Coppersmith and Youngs, 1986). This approach was used to evaluate the uncertainties for the 1996 U.S. Geological Survey (USGS) National Seismic Hazard Maps (Frankel *et al.*, 1997). This approach was also applied to the uncertainty analysis of the seismic hazard

assessment of Los Angeles, Ventura, and Orange counties (tricounty) (Cramer *et al.*, 1996) and to the New Madrid seismic zone and southern Illinois Basin (Cramer, 2001a, 2001b; Cramer *et al.*, 2002).

This article is organized into six sections after the introduction. In the first section we introduce the logic tree for the 2002 California hazard model and compare the 2002 and 1996 hazard models. In the second section we discuss fundamentals of the Monte Carlo sampling method. In the third section we outline the results of uncertainty analysis for four basic test cases. These test cases are for simple one-fault and two-fault systems, which are designed to explore the fundamental patterns of seismic ground-motion uncertainties and their sensitivities. In the fourth section we present the ground-motion uncertainty maps for the San Francisco Bay region in northern California and the tricounty region in southern California. These are the two most populated and hazardous regions in California. We compare our uncertainty results with the studies by McGuire and Shedlock (1981) for the Bay region and Cramer *et al.* (1996) for the tricounty region. A detailed sensitivity analysis is performed to show the relative contributions of uncertainties associated with each node in the logic tree. In the fifth section we show the COV (coefficient of variation) of peak ground acceleration (PGA) uncertainty map for the whole state. This uncertainty map is analyzed by using the fundamental uncertainty patterns obtained in the third section. In the sixth section we summarize and discuss the results from this study.

2002 California Seismic Hazard Logic Tree

In the 1996 California seismic hazard model, the fault sources are divided into three classes: (1) class A for active faults with rupture histories, (2) class B for active faults with unknown rupture histories, and (3) class C for areas with seismicity but unknown active faults (Petersen *et al.*, 1996; Cramer *et al.*, 1996). The class A faults are modeled as pure characteristic recurrence processes. The class B faults are modeled as a combination of characteristic and Gutenberg-Richter (G-R) recurrence processes. These same classifications are used in the 2002 California model. The 2002 model, however, is more complex than the 1996 model.

Figure 1 is the logic tree for the 2002 California model, which has evolved from that presented by Cramer *et al.* (2001) at a user workshop convened by the Applied Technology Council (ATC) and the USGS in 2001. The top row in Figure 1 lists all the node titles. The symbols in parentheses indicate the seismic sources for which the node title above and the branches below apply. The only symbol not mentioned previously is BG, which refers to the background seismicity (Cao *et al.*, 1996). The historical seismicity in California is smoothed with a Gaussian function to obtain the a-value in the G-R magnitude–frequency relation. The hazard is produced from this gridded seismicity, which we refer to as the background seismicity (Frankel, 1995). Three nodes (fault length, fault width, and shear modulus) in the logic tree of Cramer *et al.* (1996) for the tricounty region, which together contribute to the magnitude uncertainty, are not included in Figure 1. These parameters are omitted because we have implemented a new epistemic magnitude uncertainty node (Fig. 1), which replaces those three nodes. The fault slip rate, which is converted to earthquake occurrence rate in hazard calculations, is assumed to have a normal distribution and truncated at $\pm$ two standard deviations. As indicated by Petersen *et al.* (1996), the cumulative fault slip rates in the 1996 model are consistent with the modern plate tectonic rate (about 48 mm/yr) in northern California, whereas there is a small systematic discrepancy in southern California. We expect that the comparison with the plate tectonic rate would be similar for the 2002 model because only minor modifications of slip rates were implemented in the later model.

Differences exist between the 1996 and 2002 California hazard models at almost every node in Figure 1. The 1996 model was developed with the following characteristics: the tectonic settings between eastern and western California were not distinguished; three attenuation relations (Boore *et al.*, 1993; Campbell and Bozorgnia, 1994; Geomatrix, 1995) were used; only the Wells and Coppersmith (1994) fault area–magnitude relation was used; no epistemic magnitude uncertainty was included; and the moment release ratio between characteristic and G-R recurrence processes was one (same weight). In the 2002 model, there are two branches under the node "fault area-magnitude relations," which are only applied to class B faults. Each of these two branches uses the Wells and Coppersmith (1994) fault area–magnitude relation when rupture areas are smaller than 500 km^2. For rupture areas greater than 500 km^2, one branch uses the Ellsworth formula (2002 Working Group on California Earthquake Probabilities in the San Francisco Bay region [WG02]) and the other uses the Hanks and Bakun (2002) formula. This node was not in the 1996 California hazard model and incorporates extra uncertainty to the class B faults but not to the class A faults. In the 2002 model, we do not give the same weight to the characteristic and G-R models as we did for the 1996 model but instead gave two-thirds weight to the characteristic model and one-third weight to the G-R model. This modification results in a better comparison with the historic seismicity rates observed over the past 100 years (Petersen *et al.*, 2000).

The 2002 hazard model also introduced aleatory magnitude uncertainty to the class A and class B faults (Frankel *et al.*, 2002), which is not shown in the logic tree (Fig. 1). This uncertainty follows a truncated normal distribution with a standard deviation of 0.12 magnitude units. It is truncated at ± 0.15 magnitude units (Frankel *et al.*, 2002). The epistemic magnitude uncertainty is ± 0.2 for class B faults as shown in Figure 1 and ± 0.1 for class A faults, which is not shown in Figure 1. The epistemic and aleatory magnitude uncertainties are applied to both class A and class B faults. In the 2002 model, the standard deviation of magnitude 0.24 (Wells and Coppersmith, 1994) is split equally into episte-

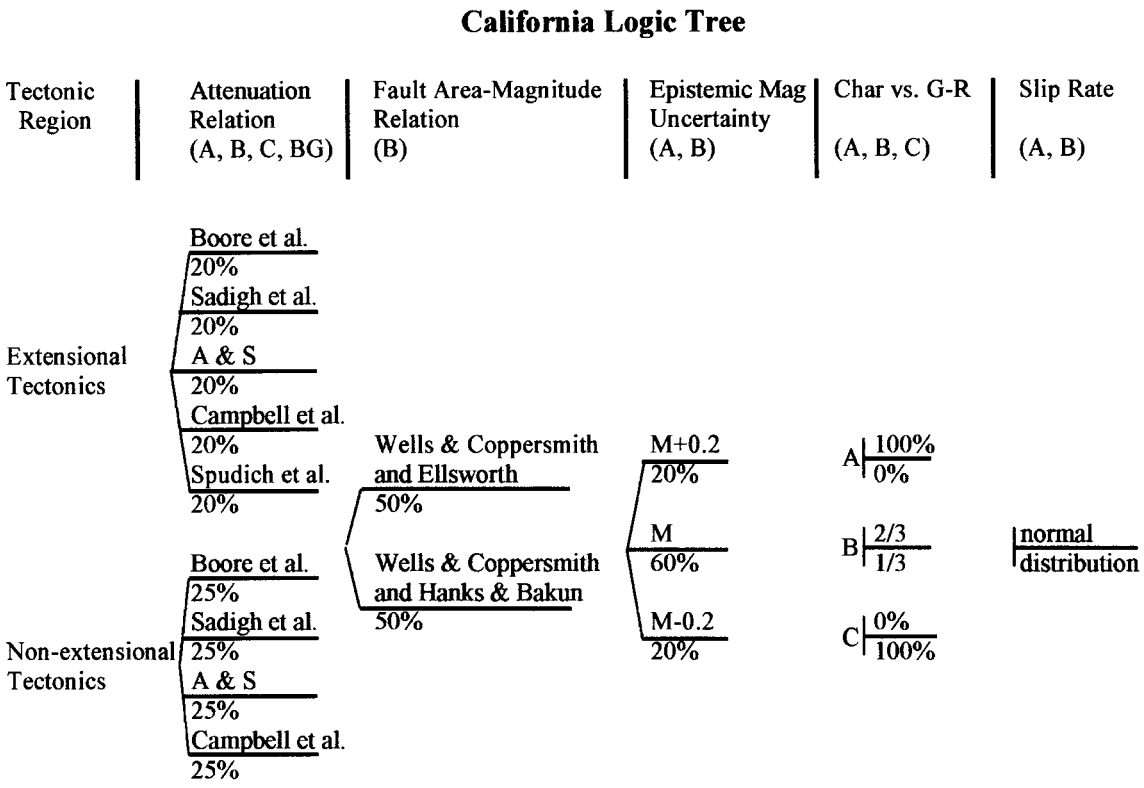

Figure 1. Partial logic tree for the 2002 update of the California seismic hazard model. The subduction zone and some other sources (see text) are not included here. The symbols in parentheses indicate the seismic source types applicable to the nodes above. They are "A" for class A faults, "B" for class B faults, "C" for area seismic zones with no known active faults, and "BG" for background seismicity. The abbreviated node title "Char vs. G-R" is for characteristic versus Gutenberg-Richter modeling of the fault recurrence processes. The epistemic magnitude uncertainty for class A faults is ± 0.1, or half of the value for class B faults shown in this figure. The references are Boore *et al.* (1997), Sadigh *et al.* (1997), Abrahamson and Silva (1997), Campbell and Bozorgnia (2003), Spudich *et al.* (1999), Wells and Coppersmith (1994), Ellsworth (WG02) and Hanks and Bakun (2002).

mic and aleatory uncertainties. Therefore, we expect lower model (epistemic) uncertainties from magnitude than Cramer *et al.* (1996), in which the entire magnitude uncertainty from Wells and Coppersmith (1994) is counted as model (epistemic) uncertainty. In addition, some other important changes also affect the magnitude distribution; these are for localized areas and are not included in the logic tree of Figure 1. These changes include the development of multiple-segment rupture models along the southern San Andreas fault, the modeling of the creeping section near the middle of the San Andreas fault, the characterization of the Brawley seismic zone, the formation of the multiple models for the Cascadia subduction zone in northwest California, and the inclusion of the results of the WG02.

Monte Carlo Sampling

Hazard models represented by different combinations of the branches of a logic tree can be sampled by using a Monte Carlo method. The weight of each model is determined by the weights of the selected branches under all applicable node titles. Let us use five integers, i, j, k, l, and m, to denote the branches selected from the five nodes in Figure 1. Fault slip rate uncertainty is the only node modeled with a continuous distribution (truncated normal distribution). In the following we assume this distribution is discretized into seven bins. For a class B fault modeled by a combination of five branches from five nodes, which are specified by those five integers, it produces a hazard curve for a specific hazard model at a particular site. The hazard curve represents the annual frequency of exceedance (AFE) as a function of PGA or spectral accelerations at a certain period and damping ratio. Each hazard curve for a specific sampled model can be expressed as $f_{i,j,k,l,m}(g)$, where g is PGA or spectral acceleration. The mean hazard curve (it is a population mean) can be expressed as:

$$f_0(g) = \sum_{i,j,k,l,m=1}^{ix,jx,kx,lx,mx} w_i^a w_j^b w_k^c w_l^d w_m^e f_{i,j,k,l,m}(g), \qquad (1)$$

where a, b, c, d, and e denote the five nodes and w_i^a, w_j^b, w_k^c, w_l^d, and w_m^e denote the branch weights for the selected branches under five nodes. The integers *ix, jx, kx, lx*, and *mx* denote the total number of branches for each of the five nodes. They are 4 (or 5), 2, 3, 2, and 7, respectively, in the California logic tree (Fig. 1). If there is more than one fault source, one more summation to sum up all the fault sources is needed in (1). Monte Carlo sampling randomly generates five integers in each Monte Carlo realization. The probability of each integer being generated is equal to the weight in (1) for a particular branch. When the number of simulations is large enough, the sample mean of the sampled hazard curves will approach the population mean (1).

In each Monte Carlo realization of Cramer *et al.* (1996), all the faults share the same set of integers except for three nodes that have continuous uncertainty distributions: fault length, fault width, and slip rate (the first two are eliminated in the current logic tree). This means that if a generated integer happens to select the characteristic model, then all the faults will be modeled as characteristic. From a physical point of view and the 2002 model assumption that each fault releases moment with both characteristic and G-R recurrence processes, we think this type of dependence between faults should be eliminated. The attenuation and fault area–magnitude relations are exceptions because all the attenuation relations used were developed for the western United States and the fault area–magnitude relations for the whole of California. Therefore, in this study we apply a slightly different Monte Carlo sampling, in which only the selections of attenuation and area–magnitude relations are the same for all faults in a particular realization. The other three selections for the branches of epistemic magnitude uncertainty, characteristic or G-R modeling, and slip rate in the logic tree are independently derived for each fault. We tried two ways of sampling, one is more fault dependent (Cramer *et al.*, 1996) and the other is more fault independent, and found that the difference in PGA (10% of exceedance in 50 years) uncertainty is about 10–15%. In the next section we will show that the more independent sampling reduces the uncertainties compared with the more dependent sampling. From the right side of (1) we can see that if we are studying the uncertainties from one or any number of nodes (sensitivity studies), the mean hazard curve is always represented by (1). For example, if we want to study the uncertainties from attenuation relations only, the Monte Carlo sampling is to randomly select i with a probability w_i^a. But the selected hazard curve referring to i is expressed by the right side of (1) without the summation to index i and the factor w_i^a. The Monte Carlo sampling for i brings back the summation to i and the factor w_i^a.

The uncertainties from Monte Carlo simulations are sample uncertainties not population uncertainties. To make the inference from sample uncertainties to population uncertainties, such as from sample COV to population COV, we tested the number of Monte Carlo simulations, or realizations, needed by plotting the 2nd, 16th, 50th, 84th, and 98th percentile values of PGA (10% of exceedance in 50 years) at a site as a function of the number of simulations. To determine the minimum number of hazard modes needed to estimate the uncertainty percentiles, we ran a sequence of models with 50–800 iterations of the logic tree. We confirmed the result of Cramer *et al.* (1996) that a minimum of 100 simulations is needed to obtain estimates of PGA uncertainty within 5% of the value obtained for simulations run with a higher number of iterations. The minimum number of iterations depends also on the unknown population standard deviation, which is related to the complexity of the logic tree. In the 2002 California hazard model, all the faults are modeled independently and, therefore, increasing the number of faults in a system will not increase the minimum number of iterations required. We used 150 iterations for all the regional simulations and 400 for all the test cases in this study. By increasing the number of iterations from 100 to 150, the 5% difference stated previously is reduced to 4% ($= 5\%\sqrt{100}\ /\ \sqrt{150}$), assuming that the population standard deviation does not change.

Ground-Motion Uncertainties of One-Fault and Two-Fault Systems

To study the fundamental patterns of uncertainty, four basic test cases are designed, as illustrated in Figure 2. The purpose of case 1 is to show how ground-motion uncertainty changes with distance from a fault and what the uncertainty contribution is from each node in the logic tree. Case 2 ex-

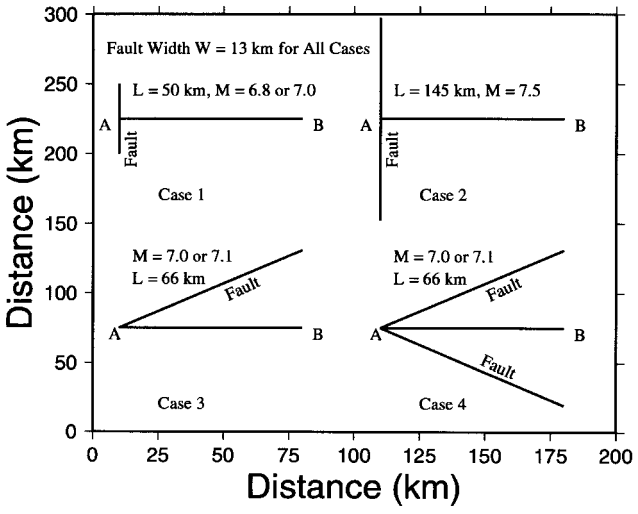

Figure 2. A schematic illustration showing the fault geometries for the four test cases. The fault is class B and fault slip rate is 5 ± 2 mm/yr in all cases. The angle between the fault and line AB is 58° in case 3. It looks smaller because the scales are different in vertical and horizontal axes. The angle between two faults in case 4 is 116° or two times 58°. The magnitudes in cases 1, 3, and 4 are calculated by using different area–magnitude relations (see text for details).

plores how uncertainty and sensitivity change with fault magnitude. Case 3 is designed so that the results for a one-fault system can be compared with the results for a two-fault system in case 4. Case 4 shows why the COV values of ground motion at locations near both faults are lower than the corresponding ones in case 3 for a one-fault system.

Case 1 is a class B fault with moderate to large magnitudes (6.8–7.0) in a nonextensional tectonic region (Fig. 2). The distance is measured from point A on the fault along a line AB perpendicular to the fault. The COV of PGA for a given AFE value of 0.0021 (a return period of 500 years) as a function of distance along line AB is shown in Figure 3. The COV of PGA due to all uncertainties from five nodes in the logic tree (Fig. 1) decreases with distance from the fault to a minimum value at a distance around 45 km and then starts to increase to higher than the values near the fault. A sensitivity analysis in Figure 4 shows that this distance-dependent uncertainty pattern is due to the uncertainty distribution given in the attenuation relations. In Figure 4 all the attenuation relations (M 7.0) used in the 2002 California hazard model for nonextensional tectonic regions are plotted. The strong-motion data used to develop these relations are better constrained at distances at about 30–50 km where most of the data are grouped and this is reflected in Figure 4. McGuire and Shedlock (1981) first found this pattern and also attributed it to the epistemic uncertainties among attenuation relations. At a large distance from the fault the COV value is high, but the mean ground motion and the corresponding standard deviation (S.D.) are all low. If a site is located near to the first fault and far from the second fault, the relatively large contribution of uncertainty from the second fault will be overwhelmed by the larger contribution of uncertainty from the first fault. This is because the ground motion and its uncertainty at the site are dominated by the first fault.

Figure 5 shows a similar analysis as in Figure 3 but for a much longer return period of 2500 years (AFE = 0.0004), which is for a 2% probability of exceedance in 50 years. Every COV curve in Figure 5 (return period of 2500 years)

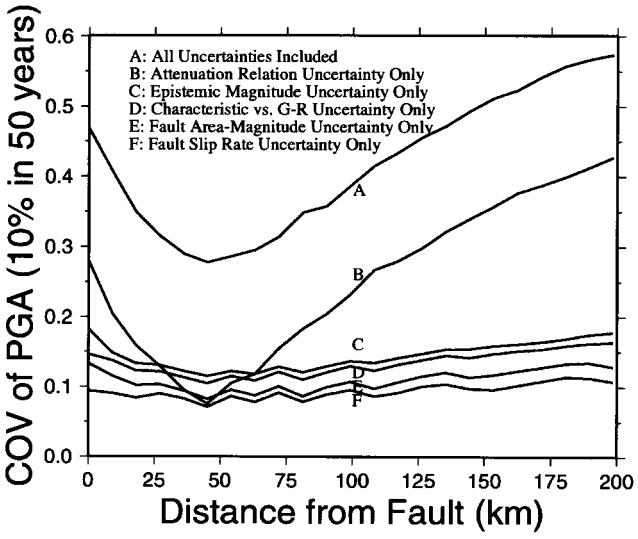

Figure 3. COV of PGA (10% of exceedance in 50 years) as a function of distance for the total model uncertainty associated with five nodes in the logic tree (Fig. 1) and for the uncertainties associated with one node only (sensitivity analysis). This figure is for test case 1 with moderate magnitudes (M ~7.0). The total uncertainty is dominated by the attenuation relations, which have strong distance dependence.

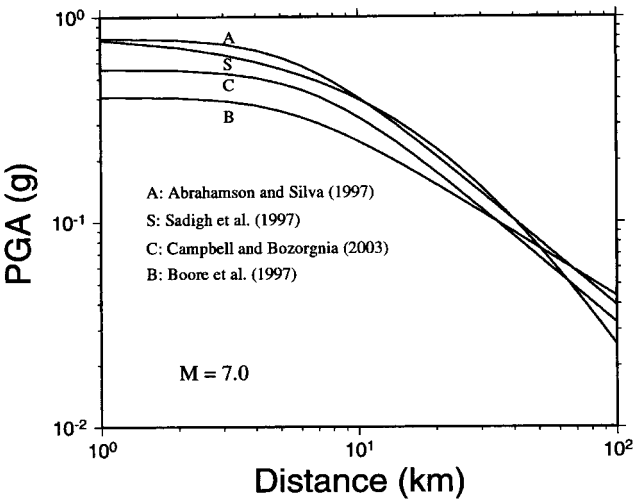

Figure 4. Four attenuation relations used in the 2002 update of the California hazard model (Fig. 1). The strong distance dependence of uncertainties shown in Figure 3 due to the attenuation relations is because of the better data constraint in all relations at distances at about 30–50 km.

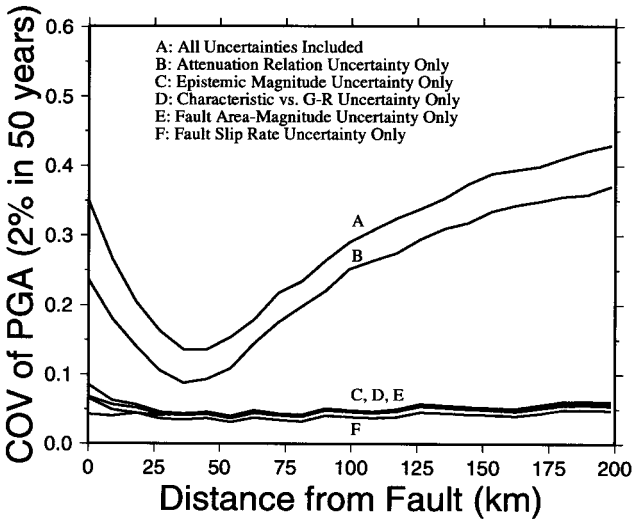

Figure 5. COV of PGA as a function of distance. It is for test case 1 but with a different AFE (2% of exceedance in 50 years) from Figure 3. The total uncertainty, as well as the uncertainties associated with one node only, is lower than the corresponding ones in Figure 3 (10% of exceedance in 50 years).

is lower than the corresponding one in Figure 3 (return period of 500 years). Here it is equivalent to say the COV decreases with decreasing AFE or increasing ground motion. But this COV decrease is following a descending hazard curve and is for the same location but increasing return period. It is not the same as the COV decrease or increase with ground motion on a ground-motion uncertainty map that is for a given return period but varying locations. To avoid any confusion in the following, we will refer the decrease of COV to a longer return period instead of to a higher ground motion when the decrease is for the same location but increasing return period. The result of COV decreasing with increasing return period seems contrary to what might be expected. This surprising result is simple to explain. We know that the mean ground motion increases with increasing return period because of the lognormal distribution of ground motion in all the attenuation relations. But the ground motion differences among all the hazard models of the logic tree (S.D. of ground motion) at different return periods or ground motion levels are determined not only by the model differences in magnitude, occurrence rate, and attenuation relation but also by the shapes of the hazard curves, which are functions of the return period or ground motion. It turns out (see Appendix) that, because of the general shape of hazard curves, which decreases monotonically with increasing return period, the ground motion is more sensitive at shorter return periods than at longer return periods to the model changes in magnitude, occurrence rate, and attenuation relation, which are all independent of the return period. Therefore, S.D. and COV of ground-motion decrease with increasing return period. Currently there is not enough information to assign uncertainty parameters in the logic trees to be return-period dependent. This may limit us in applying the uncertainty analysis to very-long-return periods because some of the uncertainties assigned in the logic tree are from short-term observations.

In the Appendix, we explain why COV decreases with increasing return period. We show that the differences between two models in magnitude, occurrence rate, and attenuation relation are equivalent to a hazard curve translation without rotation, which converts the hazard curve of a model to the hazard curve of another model. This is true in a logarithmic AFE versus logarithmic ground motion plot if the standard deviations of the distributions of logarithmic ground motion in the attenuation relations are assumed to be the same. We will show that COV does not change with increasing return period for a group of hazard curves that differ only by horizontal translations. However, S.D. and COV decrease exponentially with increasing return period for a group of hazard curves that differ by vertical translations only.

Case 2 includes a longer class B fault than in case 1 that ruptures in M 7.5 earthquakes (Fig. 2). Every COV curve for PGA at an AFE value of 0.0021 in this case as shown in Figure 6 is higher than the corresponding one in case 1 (Fig. 3). Among all the uncertainties associated with the five

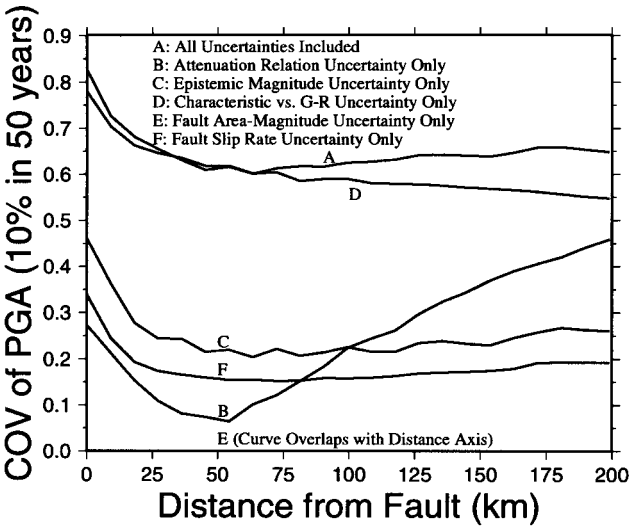

Figure 6. COV of PGA (10% of exceedance in 50 years) as a function of distance for test case 2 with a large magnitude (M ~7.5). Among all the nodes, the uncertainty associated with the node for characteristic versus G-R modeling increases the most and becomes dominant compared with case 1. The pattern that COV decreases with distance from the fault and then increases for faults with moderate magnitudes (M ~7.0) in case 1 is now changed to decreasing with distance from the fault and then leveling off. The uncertainties due to epistemic magnitude uncertainty and fault slip rate uncertainty also increase from case 1.

nodes in the logic tree, the COV for the uncertainty associated with the node for attenuation relations increases the least, and the COV for the uncertainty associated with the node for the characteristic versus G-R modeling increases the most and dominates the total uncertainty. The total uncertainty does not show the typical pattern of distance-dependent COV as shown in case 1. Instead, the COV for the total uncertainty decreases with distance from the fault and then levels off without creating a deep trough. Figure 7 shows the four hazard curves of cases 1 and 2 at site A for characteristic and G-R models, respectively. A longer fault with higher magnitude 7.5 (case 2) has a much longer mean recurrence time or a much lower occurrence rate than a shorter fault with the same slip rate but lower magnitude 7.0 (case 1). So the longer fault has lower hazards than the shorter fault at all AFE values. In both cases the G-R model predicts higher hazards than the characteristic model because the moderate-size events with high occurrence rates in the G-R model generate higher hazards than the large and less frequent characteristic events. Because of this relation, when the ratio of seismic moments to be released by characteristic and G-R models was raised from one for the 1996 California model to two for the 2002 model, the resulting hazard and its uncertainty decreased.

Figure 7 also shows that for an AFE value of 0.0021, the difference in PGA hazard between characteristic and G-R modeling is much larger for the M 7.5 fault than for the

M 7.0 fault. These results can be explained by the occurrence rate differences between the large characteristic events and the moderate events in the G-R model. Such occurrence rate difference is larger for the *M* 7.5 case than for the *M* 7.0 case. This is why the uncertainty due to characteristic versus G-R modeling in case 2 (Fig. 6) becomes dominant. The difference in PGA between characteristic and G-R modeling decreases with decreasing AFE or increasing return period. This is because, when the AFE value decreases, the hazard contribution from the characteristic events with low recurrence rates becomes increasingly important, but the contribution from G-R model tends to saturate. In the 2002 model, class A faults are not modeled with different area–magnitude relations but are modeled with characteristic recurrence only; therefore, the pattern of decreasing and then leveling off with distance for the COV in case 2 does not apply for large-magnitude class A faults. The pattern in case 1 applies to class A faults.

When two faults are close together, the values of COV are reduced for nearby sites because of the independent sampling of the branches of the logic tree. We demonstrate this by comparing cases 3 and 4. Case 3 is not much different from case 1, except the line AB (Fig. 2) is no longer perpendicular to the fault. The angle between line AB and the fault is 58°. This angle appears to be less than 58° in Figure 2 because the scales for vertical and horizontal axes are not equal. In case 4 a mirror image of the fault in case 3 is added so we can study how the hazards and uncertainties are changed from case 3. Figure 8 shows the mean hazard curves (f_1^0 and f_2^0) at site A for cases 3 and 4, respectively. According to a basic rule of statistics, the mean value of a sum of random variables equals the sum of the mean values. Therefore, we should expect the AFE value at a given PGA on curve f_2^0 to be twice the value on curve f_1^0. This simply means that the annual frequency of exceedance in case 4 is doubled compared with case 3 because we have two symmetric faults, each contributing hazard to site A. According to another rule of statistics, the variance (the square of the standard deviation) of a sum is equal to the sum of the variances for independent variables. So we should also expect the variance of case 4 to be doubled from case 3 because the two faults are independent. The COV of ground motion in case 4 is the COV in case 3 multiplied by a factor of $\sqrt{2}/2 = 0.707$, resulting in about a 30% decrease for the two-fault case.

We tested the preceding conclusions numerically. For

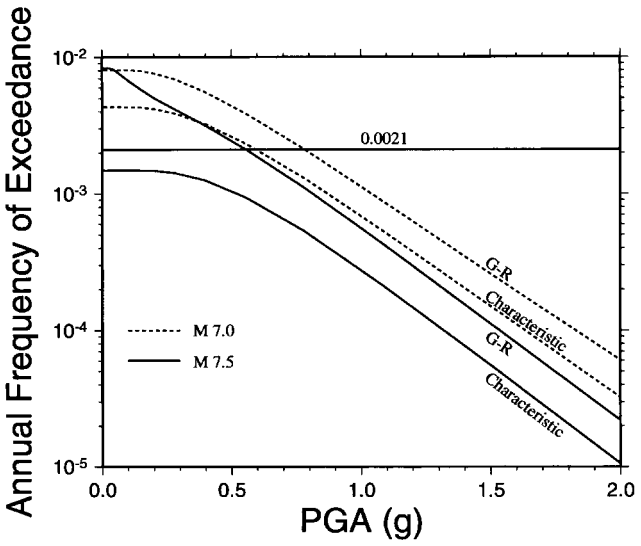

Figure 7. Hazard curves at site A in test cases 1 and 2, assuming the uncertainty from characteristic versus G-R modeling is the only uncertainty considered. The difference in PGA between characteristic model and G-R model is much larger for the large magnitude ($M \sim 7.5$) than for the moderate magnitude ($M \sim 7.0$) at an AFE value of 0.0021. The difference in PGA between characteristic model and G-R model decreases with increasing return period for both magnitudes.

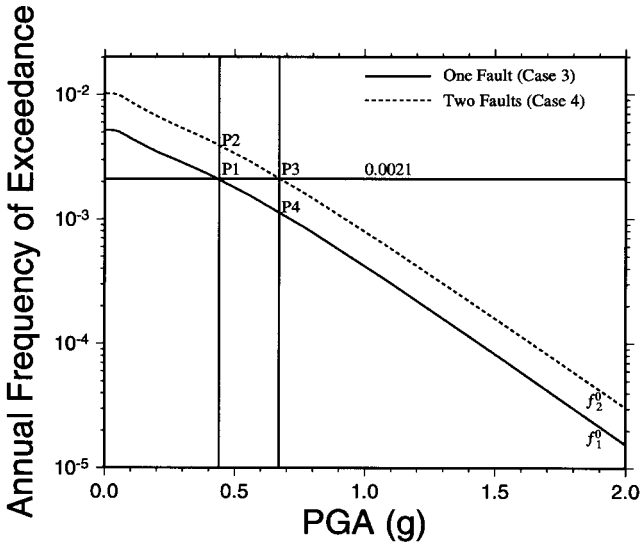

Figure 8. Mean hazard curves (f_1^0 and f_2^0) at site A in test cases 3 and 4. P1 and P3 are the cross points between f_1^0 and f_2^0 and the line for an AFE value of 0.0021. P2 is the cross point between f_2^0 and the vertical line going through point P1; P4 is the cross point between f_1^0 and the vertical line going through P3. The following relations are shown in the text: (1) AFE of f_2^0 is twice f_1^0 at any given PGA for both independent and dependent sampling; (2) COV of PGA at P2 is the same as at P1 for dependent sampling and 30% lower than at P1 for independent sampling; (3) the same relations in (2) for COV apply to points P3 and P4; (4) COV at P4 is lower than at P1 because of the lower AFE value or higher PGA value at P4 (see text for this COV decrease with increasing return period along a mean hazard curve); (5) the same relations in (4) for COV apply to points P2 and P3; and (6) COV at P3 is much lower than at P1 because of the independent sampling and the increase of ground motion from the added independent fault, which requires a descending from P2 to P3 to have the same AFE and causes COV to decrease further.

example, at a PGA value of 0.4g, the mean AFE values are 0.0022 and 0.0043 for cases 3 and 4, respectively; the COV values for ground motion are 0.59 and 0.43 for cases 3 and 4 (a 30% decrease), respectively (Fig. 8). However, if we use the dependent sampling, in which two faults share the same selection of a branch in each node, the COV of ground motion is not decreased from case 3 to case 4. This is because the variance for two dependent faults would now be four times that for one fault (equivalent to a random variable, which is twice as large as the other). In Figure 8, the line for an AFE value of 0.0021 crosses the mean hazard curve, f_1^0 for the one-fault system at P1 at a PGA value of 0.45g. For the same PGA value, the AFE value is doubled at point P2 on the mean hazard curve, f_2^0. But the COV of ground motion at P2 decreases by 30% from P1. This decrease of COV value by adding another independent random variable is due to basic statistics that show the mean value increases more than the standard deviation for an increasing number of random variables in a sum.

We are usually interested in the uncertainty of ground motion at a given AFE value. The COV value decreases from point P1 to point P2 (Fig. 8), but this is not the reduction we will see on an uncertainty map because the AFE value at P2 is different from at P1. What we calculate is the COV value at point P3 (Fig. 8) with AFE value unchanged from point P1. Point P3 is the cross point between the line for an AFE value of 0.0021 and the mean hazard curve, f_2^0 for the two-fault system in case 4. We just showed that the COV value at P3 is 30% lower than at P4 on curve f_1^0 because of the independent sampling. We also showed in case 1 (Fig. 5) that the COV value decreases with increasing return period following the descending mean hazard curve from point P1 to point P4. Therefore, the COV value reduction from point P1 to point P3 can be calculated from the two consecutive reductions from P1 to P4 because of the longer return period and then from P4 to P3 because of independent sampling. For example, if the reduction from P1 to P4 is 10% and the reduction from P4 to P3 is 30%, the reduction from P1 to P3 is 37% ($= 1-(1-0.1)(1-0.3)$).

Figure 9 compares the PGA, S.D., and COV of PGA between cases 4 and 3. We see that the PGA increases from the one-fault system to the two-fault system, but it is less than a factor of 2. The COV and S.D. values both decrease because of independent sampling of the two nearby faults as discussed previously (Fig. 8). The S.D. for the two-fault system is lower than for the one-fault system, assuming the same AFE value or return period. This is due to the combination of the increased rate of ground shaking in the two-fault model (higher hazard curve) and the unchanged AFE value. This combination leads to a S.D. and the corresponding COV reductions, which can be viewed as the reductions along the descending hazard curve from point 2 to point 3 in Figure 8. Without this, the S.D. of PGA for a two-fault system should be higher than for the one-fault system by a factor of $\sqrt{2}$. The COV and S.D. reductions discussed previously for the two-fault system becomes negligible when

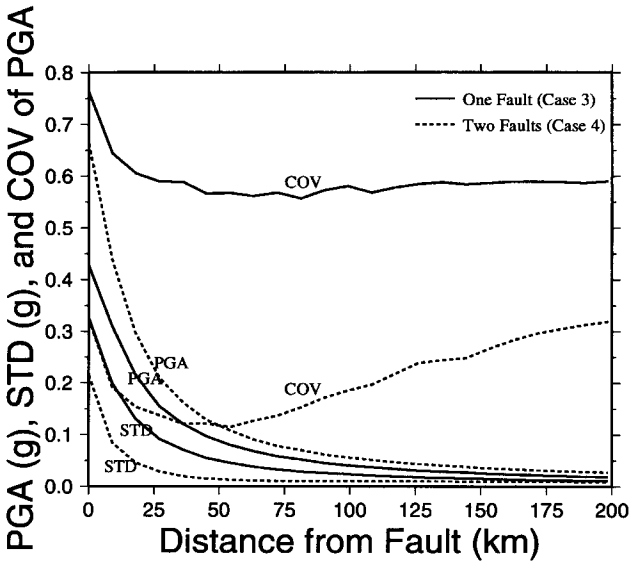

Figure 9. Curves of COV, S.D., and PGA (10% of exceedance in 50 years) from test cases 3 and 4 at site A with independent sampling. The PGA value increases from a one-fault to a two-fault system. But it is less than the increase of AFE value. The AFE value is exactly doubled from a one-fault to a two-fault system. Values of COV and S.D. of PGA decrease significantly from a one-fault to a two-fault system because of independent sampling of the two nearby faults.

faults are far apart from each other. In this case the uncertainty is dominated by a single fault.

In this section we have shown several basic test cases for class B faults with the following results: (1) sources with $M \sim 7$ are characterized by a pattern in which COV of PGA decreases with distance from a fault and then increases with distance beyond 50 km (case 1); (2) sources with large $M \sim 7.5$ magnitudes are characterized by a pattern in which the COV decreases with distance from the fault and then levels off (case 2); (3) a system containing two or more faults that are close to each other is characterized by a pattern in which COV of ground motion at a site near these faults decreases compared with the case of a single fault. These are general and qualitative uncertainty patterns that will help us to analyze the uncertainty maps in the next sections. The quantitative estimation of how much COV decreases or increases with distance from a fault and how much the COV decreases due to adjacent faults depends on the fault geometries, magnitudes, and slip rates.

Uncertainties for the San Francisco and the Southern California Tricounty Regions and Sensitivity Analysis for the Tricounty Region

The San Francisco Bay region and southern California tricounty region that comprises Ventura, Los Angeles, and Orange counties are heavily populated and have very high

seismic hazards. These two regions are also rich in fault types such as the well-defined class A faults and the poorly defined blind thrust faults beneath or adjacent to the Los Angeles and San Francisco basins, the western Transverse Ranges, and the Santa Barbara Channel. We compute uncertainties and sensitivities with a 0.05° spacing in these two regions.

Figure 10 is a fault map for the tricounty region and Figure 11 is the map of COV for PGA (10% exceedance in 50 years) for the same region. The multisegment rupture models for the southern San Andreas fault are not included in the California logic tree (Fig. 1) but are sampled in the Monte Carlo simulations according to the cascade model weights described in Appendix A of Frankel *et al.* (2002). To compare with the early study of Cramer *et al.* (1996), we first made versions of maps (Figs. 11 and 12) that include only fault source hazard but do not include the hazard due to background seismicity. The contour interval is 0.05 in Figure. 11. The COV values along the San Andreas fault range from 0.1 to 0.15. The COV values decrease with distance on both sides of the San Andreas fault to 0.1 and then increase to about 0.2 to 0.3 near class B faults. This pattern of decreasing COV with distance from the fault and then increasing at distance beyond 50 km is typical, as we have shown in case 1. The COV does not keep increasing to higher values, like 0.4 to 0.5. There are three particularly high COV areas, which are centered over the Garlock, Anacapa-Dume, and Palos Verdes faults. These are the poorly defined faults characterized by large magnitudes and relatively high long-term slip rates. They are the same type of faults as in case 2. The high COV values are mostly due to the large magnitudes, which lead to very different ground-motion hazards between characteristic and G-R recurrence models. These faults are also relatively isolated, with few nearby faults that would reduce the COV values.

The general patterns of this COV map (Fig. 11) are similar to Cramer *et al.* (1996) with one exception, which is that the uncertainties along class A faults are lower than along class B faults in this study. From the logic tree (Fig. 1), we except class A faults to have lower ground-motion uncertainties than class B faults. The class A faults do not have uncertainty contributions from area–magnitude relations and characteristic versus G-R modeling (they are 100% characteristic). The epistemic magnitude uncertainty for the San Andreas fault is ± 0.1, which is only half of the value for class B faults (Fig. 1).

The fault slip rate uncertainties (COV) for class A faults are always much lower than for class B faults. These class A faults have higher standard deviations and mean ground motions than class B faults, but the ratio of these two quantities is lower for class A faults than for class B faults. In addition, the multiple cascading models for the southern San Andreas fault, which are sampled independently in this study, also contribute to the lower uncertainties of class A faults compared with the Cramer *et al.* (1996) model.

Another observation to be made in Figure 11 is that the

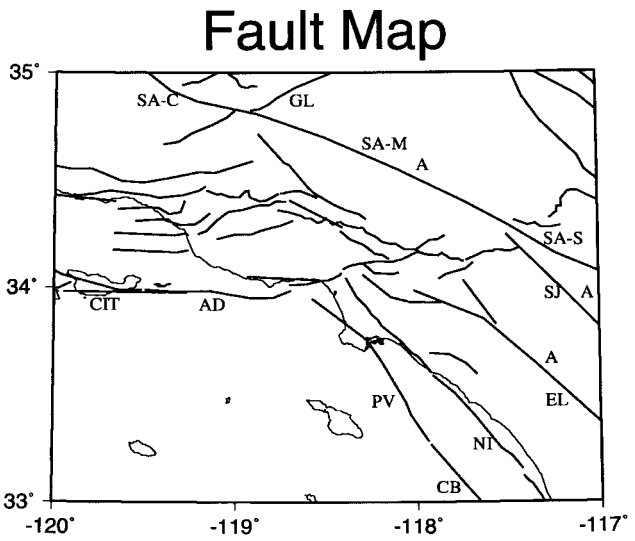

Figure 10. Fault map for the tricounty region. Class A faults are labeled with the symbol "A." The abbreviations are: SA-M, Mojave segment of San Andreas fault; SA-C, Carrizo segment; SA-S, San Bernardino segment; GL, Garlock fault; SJ, San Jacinto fault; EL, Elsinore fault; NI, Newport-Inglewood fault; CB, Coronado Bank fault; PV, Palos Verdes fault; AD, Anacapa-Dume fault; CIT, Channel Islands Thrust fault.

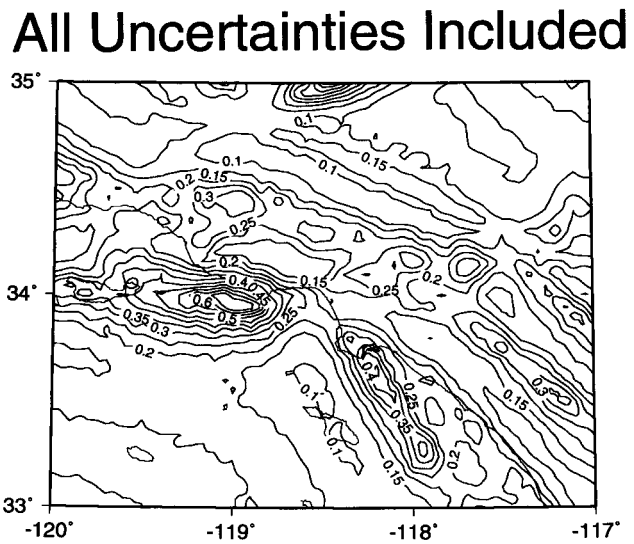

Figure 11. COV of PGA (10% of exceedance in 50 years) map for the tricounty region with all the model uncertainties included except the uncertainty from background seismicity. COV values range from 0.1 to 0.15 along class A faults and 0.2 to 0.3 along most of the B faults. High COV values (0.4–0.6) are observed around the Garlock, Anacapa-Dume, and Palos Verdes faults. The contours make U-turns at both ends of the Mojave segment, where multiple faults or segments are close to each other. These are typical examples of COV decrease due to independent sampling of multiple nearby sources. These can be found at many other places on this map.

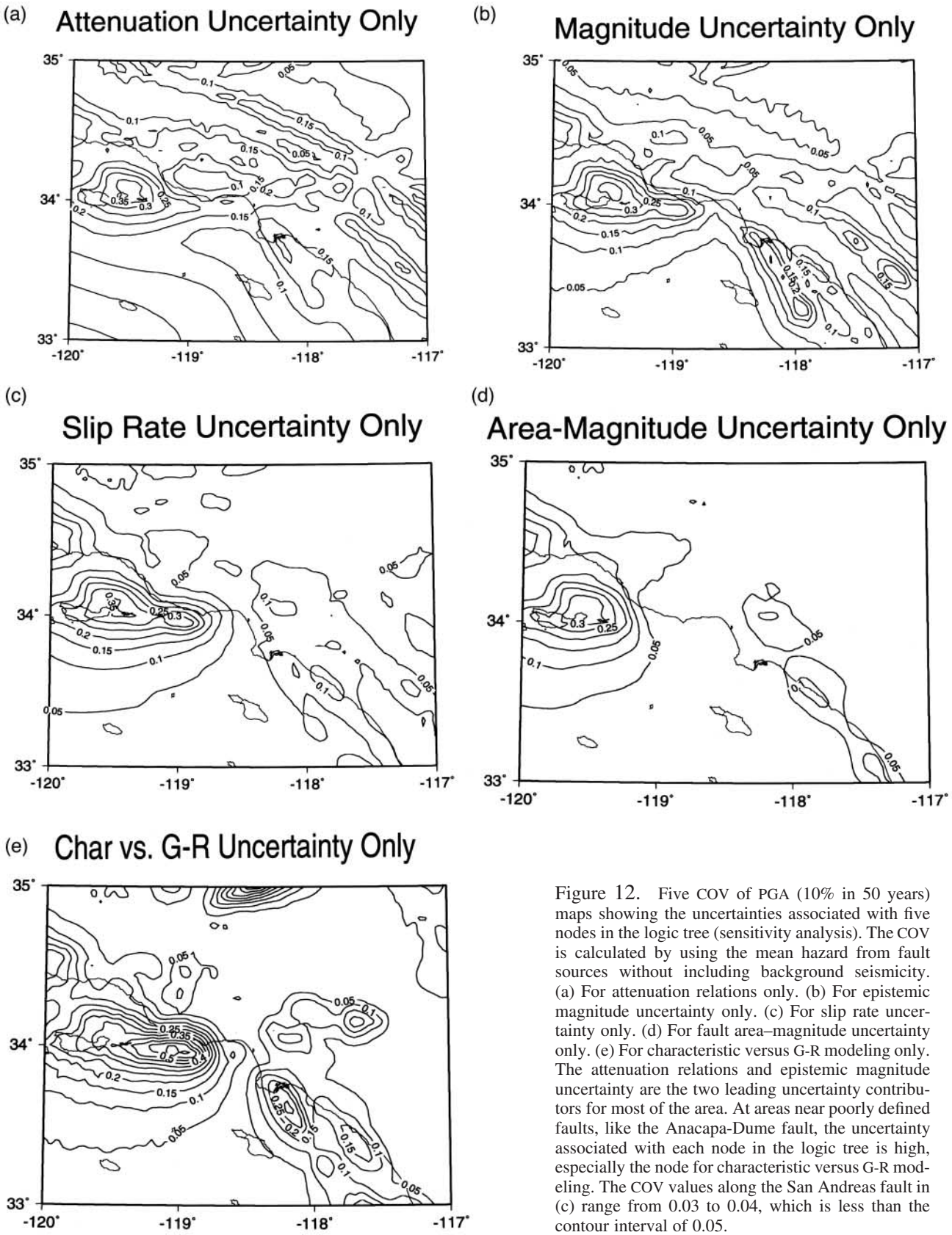

Figure 12. Five COV of PGA (10% in 50 years) maps showing the uncertainties associated with five nodes in the logic tree (sensitivity analysis). The COV is calculated by using the mean hazard from fault sources without including background seismicity. (a) For attenuation relations only. (b) For epistemic magnitude uncertainty only. (c) For slip rate uncertainty only. (d) For fault area–magnitude uncertainty only. (e) For characteristic versus G-R modeling only. The attenuation relations and epistemic magnitude uncertainty are the two leading uncertainty contributors for most of the area. At areas near poorly defined faults, like the Anacapa-Dume fault, the uncertainty associated with each node in the logic tree is high, especially the node for characteristic versus G-R modeling. The COV values along the San Andreas fault in (c) range from 0.03 to 0.04, which is less than the contour interval of 0.05.

contours along the Garlock fault and the Mojave and Carrizo segments of the San Andreas fault all make a U-turn where these faults approach each other. This phenomenon also happens near the southeast end of Mojave segment, where the San Jacinto fault and the San Bernardino and Mojave segments of the San Andreas fault are close to each other. The low COV areas between those contours are due to the independent sampling of the multiple nearby faults as shown in test case 4. The COV values in the central part of the region (Figs. 10 and 11), where many faults are close to each other, are in the range of 0.2 to 0.3 and not 0.4 to 0.5. In general, the isolated faults mentioned previously have very high COV values because they are poorly defined. There are no nearby faults to reduce the uncertainty through independent sampling of these nearby faults. This may also partly explain the high COV values in eastern and central United States, where in many cases only one major fault system dominates the hazards.

Figure 12a–e shows the sensitivity analysis using the COV of PGA (all at 10% of exceedance in 50 years) for each node separately. These figures show how each of the five nodes (Fig. 1) contributes to the total uncertainty shown in Figure 11. The relative contributions among these figures are consistent with the results of cases 1 and 2. The attenuation relations (Fig. 12a) contribute most of the uncertainty as shown in case 1 (Fig. 3), except for regions around the Garlock, Anacapa-Dume, and Palos Verdes fault areas, where the uncertainty is primarily due to characteristic versus G-R modeling (Fig. 12e) as shown in case 2. The uncertainties due to the epistemic magnitude uncertainty (Fig. 12b), fault slip rate uncertainty (Fig. 12c), and area–magnitude relations (Fig. 12d) are all relatively small as shown in case 1 (Fig. 3). The relatively low uncertainty due to fault slip rate for the San Andreas fault in southern California (Fig. 12c) is because of the independent sampling assigned to these multiple cascading models.

The uncertainty from the background seismicity is due to the attenuation relations (Fig. 1) and the incompleteness of seismic catalogs (Cao et al., 1996). The latter is not shown in the logic tree (Fig. 1) but is counted in this study by resampling the catalogs with a simple bootstrap method. Figure 13 shows the map of COV for background seismicity. The mean hazard used to calculate COV here is different from Figures 11 and 12. To compare our results with Cramer *et al.* (1996), the mean hazard used to normalize the standard deviation of ground motion in Figures 11 and 12 is from fault sources only. In Figure 13 the mean hazard is from both fault and background sources. The mean hazard in Figure 13 is about the same as the hazard in the 2002 update of California model within a few percent. Figure 13 shows that the uncertainty due to background seismicity is insignificant compared with the uncertainties from fault sources. The contours in Figure 13 show some influence from the major faults in this region because the background seismicity is mostly distributed off the faults.

Figure 14 is a fault map for the San Francisco Bay re-

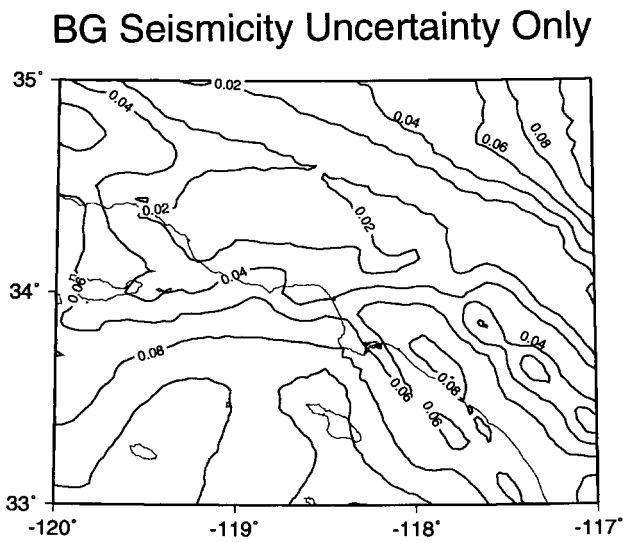

Figure 13. COV of PGA (10% in 50 years) map showing the uncertainty from background seismicity only. The COV is calculated by using the mean hazard from fault and background seismicity sources. The uncertainty contribution from background seismicity is very low compared with the fault sources. The contours still show the distribution of fault traces that the uncertainty is the lowest along the faults. This is consistent with the fact that the uncertainty from background seismicity is due to the seismicity off the faults.

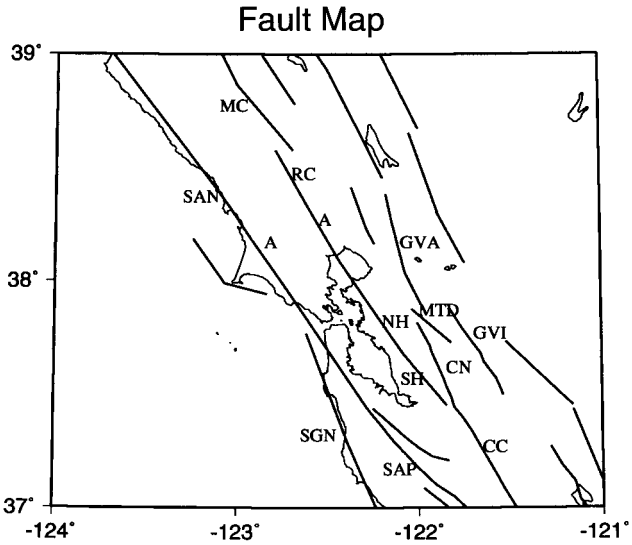

Figure 14. Fault map for the San Francisco Bay region. Class A faults are labeled with the symbol "A." Abbreviations for the fault names: SAN, San Andreas North Coast; SAP, San Andreas Peninsula; SGN, San Gregorio North; CC, Calaveras Central; CN, Calaveras North; SH, Hayward South; NH, Hayward North; RC, Rodgers Creek; MC, Maacama; GVA, Green Valley; GVI, Greenville; MTD, Mt. Diablo.

gion. In the 1996 model (Petersen *et al.*, 1996) only the San Andreas and Hayward faults were classified as class A faults and the other faults were class B faults. In the 2002 update of the California hazards, the results from WG02 were adopted. We may expect lower uncertainties from these faults that have been updated to class A faults because of the studies of WG02. In our uncertainty analysis we also adopted the WG02 results that are based on a logic tree analysis. The attenuation relations used for this region are the same as in the logic tree for nonextensional tectonics (Fig. 1). Figures 15 is the COV map for PGA at 10% of exceedance in 50 years. Because this map extends far enough into the ocean where there are no seismic sources included in the 2002 hazard model, the COV contours show a clear distance-dependent trend of uncertainty that decreases with distance from the faults and then increases to the west. The COV values along the class A faults and those class B faults studied by WG02 range from 0.1 to 0.15, which are similar to the COV values calculated for A faults within the tricounty region of southern California. The very high COV values over the Maacama fault are mostly due to the characteristic versus G-R modeling as shown in case 2. The decrease of COV due to independent sampling of the multiple nearby faults also can be found between Maacama and Rodgers Creek faults, in the area where Green Valley, Concord, Greenville, Calaveras, Mt. Diablo, and Hayward faults are close to each other. The COV values in this region are about half of the values obtained by McGuire and Shedlock (1981). Cramer *et al.* (1996) attributed the higher COV values of McGuire and Shedlock (1981) to the use of discrete distributions of fault parameters with large variances.

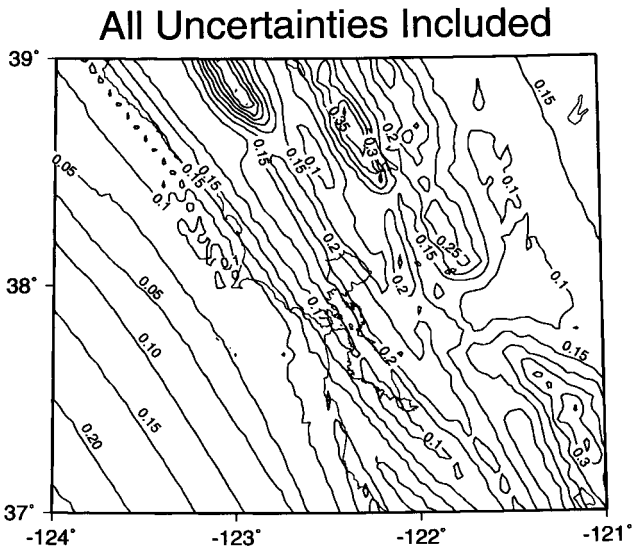

Figure 15. COV of PGA (10% in 50 years) map for the San Francisco Bay region with all the model uncertainties included except the uncertainty from background seismicity. COV values along class A faults and those class B faults studied by the WG02 range from 0.1 to 0.15, which are similar to the tricounty region. COV values range from 0.2 to 0.3 along class B faults in the southeast and north parts of the region. The pattern of COV decreasing with distance from the fault and then increasing (case 1) is very clear in the west from the San Andreas fault far into the ocean. There are no seismic sources to interrupt this pattern to the west. This pattern is overwhelmed by local sources and background seismicity to the east. The decrease of COV due to independent sampling of multiple nearby sources can be found in many localized areas like between Maacama and Rodgers Creek faults and among many faults around the Mt. Diablo fault area.

California Ground-Motion Uncertainty Map

In creating the California probabilistic ground-motion (PGA) uncertainty map (COV of PGA), we have included all the seismic sources that are not included in the logic tree (Fig. 1). These additional sources include the Cascadia subduction zone, the northern California deep earthquake zone (depth >35 km), and the shear zones in California and Nevada. In the 2002 California hazard model, four alternative models are used to characterize the eastern edges of the Cascadia subduction zone. Three of these are based on the work of Flück *et al.* (1997) with weights of 0.1, 0.2, and 0.2 (Frankel *et al.*, 2002). The fourth one is revised from the 1996 model (Frankel *et al.*, 1996; Petersen *et al.*, 1996) with a weight of 0.5. Floating earthquakes with magnitudes 8.3 and 9.0 are placed on these rupture zones. To evaluate the ground-motion uncertainties from these floating events, the magnitude uncertainty is set to ± 0.1. The variation of a-value, which is the seismicity rate in the G-R magnitude–frequency relation, is obtained using the aperiodicity value of 0.58 obtained by Petersen *et al.* (2002) based on paleoseismic recurrence data. The deep seismicity in northern California is modeled differently from the background seismicity in the rest of California. There are two alternative attenuation models weighted equally for in-slab earthquakes. One of them uses two attenuation relations given by Atkinson and Boore (2003); the other uses only one attenuation relation by Youngs *et al.* (1997). To avoid a sudden drop of uncertainty that would be caused by including only one attenuation relation in the second model, an artificial variation of 15% is added to the attenuation relations to account for this difference. The fault slip rate uncertainties for Nevada and Oregon faults are not available yet, so the uncertainty contributions from these faults are calculated without slip rate uncertainties. However, these uncertainties should have a small contribution to the hazard uncertainty in California because of their low slip rates.

Figure 16 is the California COV map for PGA at 10% of exceedance in 50 years. The uncertainty from background seismicity is also included so the mean hazard is similar to the 2002 update of the California mean hazard to within a few percent. This map is calculated with a grid spacing of 0.1°. Because of the added uncertainty due to background seismicity, the uncertainties for the San Francisco Bay and

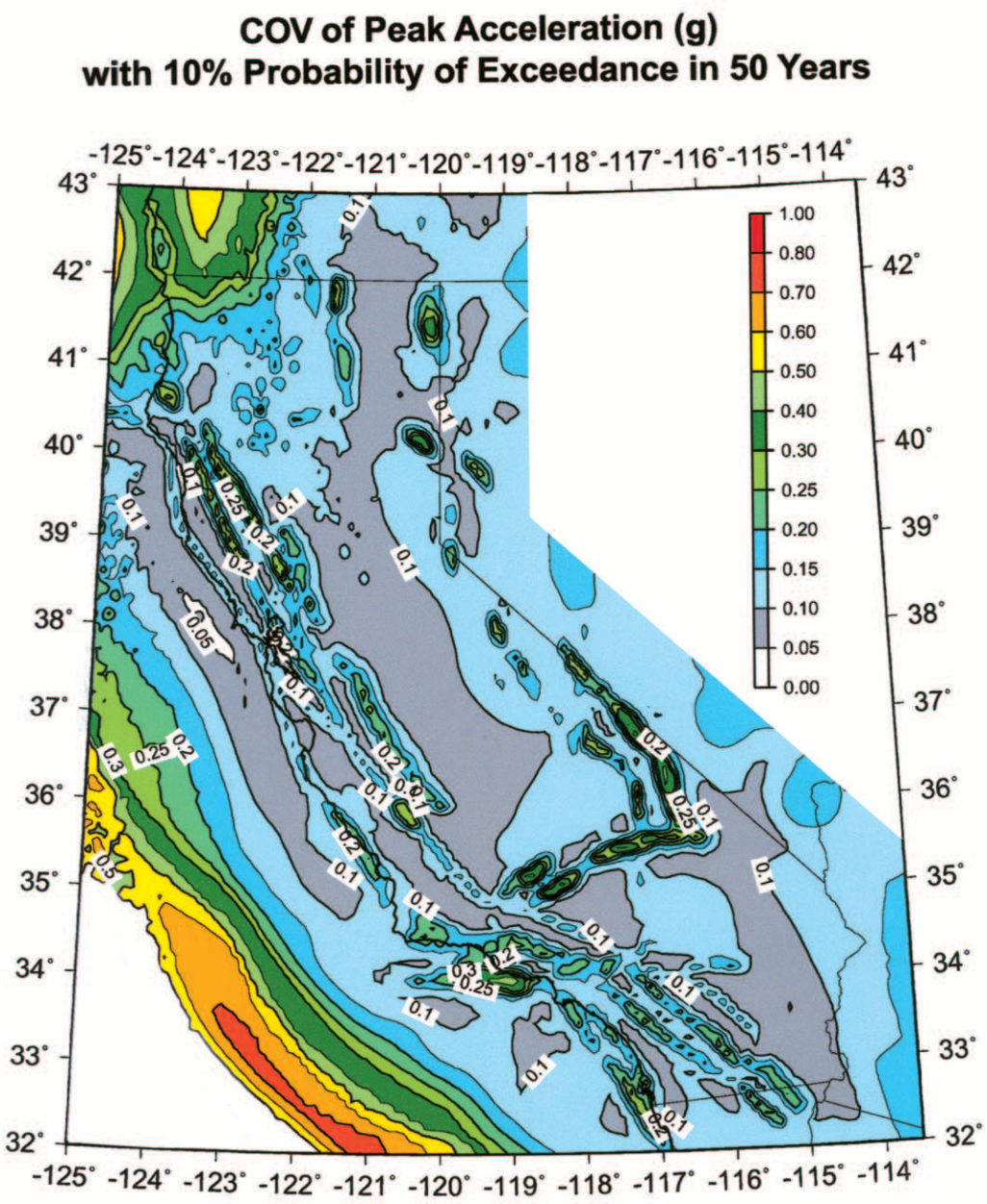

Figure 16. California COV map for PGA (10% of exceedance in 50 years). COV values along the San Andreas fault system from southern to northern California range from 0.1 to 0.15. It decreases with distance from faults in both directions east and west and then increases. COV values along most of the class B faults range from 0.2 to 0.3. The pattern that COV increases at large distances from faults is not overwhelmed by local sources to the west into the ocean but is overwhelmed in central and east California by the local and Nevada sources and the background seismicity. High COV values are observed around Anacapa-Dume, Palos Verdes, and Garlock faults in southern California and around Maacama fault and Cascadia subduction zone in northern California. These are poorly defined faults with large magnitudes.

tricounty regions are not exactly the same as shown in Figures 11 and 13. However, we find that all the observed patterns of spatial-uncertainty distributions and their relations to faults in the San Francisco Bay and tricounty regions are even more clearly seen on the state uncertainty map. The COV values range from 0.1 to 0.15 along the San Andreas fault system and other class A faults from southern to northern California. The COV values around class B faults, which are located adjacent to the northwest-trending class A faults, are generally higher and range from 0.2 to 0.3. The distance-dependent pattern of COV that decreases with distance from a fault and then increases with distance beyond 50 km is also

shown along every fault with magnitude less than M 7.5. The principal San Andreas fault system has a northwest–southeast strike, and this contour pattern continues out into the Pacific Ocean. The pattern revealed in test case 1 is observed in the contours beyond the coastline where the 2002 hazard model does not include seismic sources. To the east of the San Andreas fault system, however, such a pattern is not observed because of the influence of nearby fault sources in eastern California and Nevada. The decrease of COV values because of independent sampling of multiple nearby faults is also observed at many places across California.

Besides those faults highlighted in the preceding section, high COV values are also observed at the Cascadia subduction zone in northern California and near the Nevada border in eastern California. The high COV values around the Garlock and Maacama faults, partially shown in Figures 11 and 15, are now shown along their entire lengths (Fig. 16). All these faults that have high-hazard uncertainties share some common features: they produce large-magnitude earthquakes with long recurrence times; they are poorly defined; and they have no nearby faults to reduce uncertainties through independent sampling.

Summary and Discussions

In this study we proposed to use the independent Monte Carlo sampling method to determine the epistemic uncertainty in the 2002 California hazard model. The COV values for PGA obtained by using independent sampling are lower than the values obtained by using dependent sampling because of the basic statistical feature of probabilistic hazards.

We evaluated the ground-motion uncertainties (COV) for four designed test cases. From these test cases we showed the basic features or patterns of uncertainties for systems containing one fault and two faults. In case 1 we found that for faults with $M \sim 7.0$, the COV of ground motion decreases at a short distance from a fault and then increases with a distance beyond 50 km. The lowest uncertainties are located at distances at about 30–50 km from the fault, where most of the strong-motion data are grouped in the development of attenuation relationships. We also observed that when the return-period increases (i.e., AFE value is lowered) but magnitude and fault slip rate are unchanged, COV decreases significantly (Fig. 5). Case 2 showed that when fault magnitude is increased but the return period and fault slip rate are unchanged, the COV increases significantly (Fig. 6). At high magnitudes ($M \sim 7.5$) the COV increase is mostly due to the characteristic versus G-R modeling. In cases 3 and 4 we compared the uncertainty differences between a one-fault and a two-fault systems. In the two-fault system the annual frequency of exceedance is doubled compared with the one-fault system, but the COV decreases because of the independent sampling of multiple nearby faults.

The California COV map for PGA at 10% exceedance in 50 years shows all the features or patterns discussed previously. In the future, the hazard maps will evolve but these general patterns of uncertainty (COV) will remain the same because they are generic patterns obtained from simple one-fault and multiple-fault systems. The COV values along the class A San Andreas fault system range from 0.1 to 0.15, which are lower than the values of 0.2 to 0.3 along the class B faults. The COV decreases from class A faults to 0.1 or less and then increases to 0.3 or more in the direction of the Pacific Ocean where there are no seismic sources that will dominate the uncertainty. To the east, this pattern is interrupted by the contribution from local seismic sources in the Great Valley, eastern California, Nevada, and by background seismicity. The COV values in the Great Valley and eastern California are in the range of 0.05 to 0.1. It is low because of the low fault earthquake magnitudes (<7.0) in this area. In this case the characteristic versus G-R modeling is no longer a major contributor to the uncertainty. Much higher COV values (0.4–0.6) are observed around the Garlock, Anacapa-Dume, and Palos Verdes faults in southern California and the Maacama fault and Cascadia subduction zone in northern California, where there are greater uncertainties about the slip rates and magnitudes of the potential ruptures.

These uncertainty results may be useful to the engineers for better understanding the safety factor in building and structural designs. The uncertainty results may also help in determination of insurance rates, which largely depends on using the hazard maps and their uncertainties to estimate the potential losses to a portfolio of properties. One of the fast-developing applications of hazard maps is in loss estimation, which will benefit greatly from the uncertainty analysis of this study (Cao *et al.*, 2000; Wesson and Perkins, 2001). Most importantly, this study provides a reference point for comparison with future studies to determine how these uncertainties may be reduce through better data and research.

Acknowledgments

We are grateful to William Bryant and Jerome Treiman for providing us the fault data. We thank Ivan Wong and two anonymous reviewers for their critical reviews. We thank Chris Cramer, Robert Wesson, David Perkins, Stephen Harmsen, Chris Wills, and Cathy Slater for helpful discussions. This study is supported by USGS/NEHRP grant 04HQGR0027.

References

Abrahamson, N. A., and W. J. Silva (1997). Empirical response spectral attenuation relations for shallow crustal earthquakes, *Seism. Res. Lett.* **68,** 94–127.

Atkinson, G. M., and D. M. Boore (2003). Empirical ground-motion relations for subduction zone earthquakes and their application to Cascadia and other regions, *Bull. Seism. Soc. Am.* **93,** 1703–1729.

Boore, D. M., W. B. Joyner, and T. E. Fumal (1993). Estimation of response spectra and peak acceleration from western North American earthquakes: an interim report, *U.S. Geol. Surv. Open-File Rept. 93-509.*

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, *Seism. Res. Lett.* **68,** 128–153.

Campbell, K. W., and Y. Bozorgnia (1994). Near-source attenuation of

peak horizontal acceleration from worldwide accelerograms recorded from 1957 to 1993, in *Proc. of the 5th U.S. National Conference on Earthquake Engineering*. Chicago, Illinois, 10–14 July 1994, Vol. 3, Earthquake Engineering Research Institute, Oakland, California, 283–292.

Campbell, K. W., and Y. Bozorgnia (2003). Updated near-source ground motion attenuation relations for the horizontal and vertical components of peak ground acceleration and acceleration response spectra, *Bull. Seism. Soc. Am.* **93,** 314–331.

Cao, T., M. D. Petersen, C. H. Cramer, T. R. Toppozada, M. S. Reichle, and J. F. Davis (2000). The calculation of expected loss using probabilistic seismic hazard, *Bull. Seism. Soc. Am.* **89,** 867–876.

Cao, T., M. D. Petersen, and M. S. Reichle (1996). Seismic hazard estimation from background seismicity in southern California, *Bull. Seism. Soc. Am.* **86,** 1372–1381.

Coppersmith, K. J., and R. R. Youngs (1986). Capturing uncertainty in probabilistic seismic hazard assessment within intraplate tectonic environments, in *Proceedings of the 3rd U.S. National Conference on Earthquake Engineering*, August 24–28, 1986, Charleston, South Carolina, 301–312.

Cramer, C. H. (2001a). The New Madrid seismic zone: capturing variability in seismic hazard analysis, *Seism. Res. Lett.* **72,** 664–672.

Cramer, C. H. (2001b). Aseismic hazard uncertainty analysis for the New Madrid seismic zone, *Eng. Geol.* **62,** 251–266.

Cramer, C. H., M. D. Petersen, and A. D. Frankel (2001). Monte Carlo analysis for assessing uncertainty in national hazard maps (abstract), Presented at Workshop by Applied Technology Council (ATC-35) and the USGS, May 2001 San Francisco, CA.

Cramer, C. H., M. D. Petersen, and M. S Reichle (1996). A Monte Carlo approach in estimating uncertainty for a seismic hazard assessment of Los Angeles, Ventura, and Orange counties, California, *Bull. Seism. Soc. Am.* **86,** 1681–1691.

Cramer, C. H., R. L. Wheeler, and C. S. Mueller (2002). Uncertainty analysis for seismic hazard in the southern Illinois Basin, *Seism. Res. Lett.* **73,** 792–805.

Flück, P., R. D. Hyndman, and K. Wang (1997). Three-dimensional dislocation model for great earthquakes of the Cascadia subduction zone, *J. Geophys. Res.* **102,** 20,539–20,550.

Frankel, A. (1995). Mapping seismic hazard in the Central and Eastern United States, *Seism. Res. Lett.* **66,** 8–21.

Frankel, A., C. Mueller, T. Barnhard, D. Perkins, E. Leyendecker, N. Dickman, S. Hanson, and M. Hopper (1996). National seismic-hazard maps: documentation June 1996, *U.S. Geol. Surv. Open-File Rept. 96-532*, 110 pp.

Frankel, A. D., S. Harmsen, C. Mueller, T. Barnhard, E. V. Leyendecker, D. Perkins, S. Hanson, N. Dickman, and M. Hopper (1997). USGS national hazard maps: uniform hazard spectra, de-aggregation, and uncertainty, in Proceedings of FHWA/NCEER Workshop on the National Representation of Seismic Ground Motion for New and Existing Highway Facilities, *NCEER Technical Report 97-0010,* 39–73.

Frankel, A. D., M. D. Petersen, C. S. Muller, K. M. Haller, R. L. Wheeler, E. V. Leyendecker, R. L. Wesson, S. C. Harmsen, C. H. Cramer, D. M. Perkins, and K. S. Rukstales (2002). Documentation for the 2002 Update of the National Seismic Hazard Maps, *U.S. Geol. Surv. Open-File Rept. 02-420*, 33 pp.

Geomatrix Consultants (1995). Seismic design mapping state of Oregon, Final Report, prepared for the Oregon Department of Transportation, Salem, Oregon.

Hanks, T. C., and W. H. Bakun (2002). A bilinear source-scaling model for M-log A observations of continental earthquakes, *Bull. Seism. Soc. Am.* **92,** 1841–1846.

McGuire, R. K., and K. M. Shedlock (1981). Statistical uncertainties in seismic hazard evaluation in the United States, *Bull. Seism. Soc. Am.* **71,** 1287–1308.

Petersen, M. D., W. A. Bryant, C. H. Cramer, T. Cao, N. S. Reichle, A. D. Frankel, J. J. Lienkaemper, P. A. McCrory, and D. P. Schwartz (1996). Probabilistic seismic-hazard assessment for the state of California, *California Division of Mines and Geology Open-File Rept. 96-08, U.S. Geol. Surv. Open-File Rept. 96-706.*

Petersen, M. D., C. H. Cramer, and A. D. Frankel (2002). Simulation of seismic hazard for Pacific Northwest of the United States from earthquakes associated with the Cascadia subduction zone, *Pure Appl. Geophys.* **159,** 2147–2168.

Petersen, M. D., C. H. Cramer, M. S. Reichle, A. D. Frankel, and T. C. Hanks (2000). Discrepancy between earthquake rates implied by historic earthquakes and a consensus geologic source model for California, *Bull. Seism. Soc. Am.* **90,** 1117–1132.

Reiter, L. (1990). *Earthquake Hazard Analysis: Issues and Insights*, Columbia University PressNew York.

Sadigh, K., C. Y. Chang, J. Egan, F. Makdisi, and R. Youngs (1997). Attenuation relationships for shallow crustal earthquakes based on California strong motion data, *Seism. Res. Lett.* **68,** 180–189.

Spudich, P., W. B. Joyner, A. G. Lindh, D. M. Boore, B. M. Margaris, and J. B. Fletcher (1999). SEA99: a revised ground motion prediction relation for use in extensional tectonic regimes, *Bull. Seism. Soc. Am.* **89,** 1156–1170.

Wells, D. L., and K. J. Coppersmith (1994). New empirical relationships among magnitude, rupture length, rupture width, and surface displacements, *Bull. Seism. Soc. Am.* **84,** 974–1002.

Wesson, R. L., and D. M. Perkins (2001). Spatial correlation of probabilistic earthquake ground motion and loss, *Bull. Seism. Soc. Am.* **91,** 1498–1515.

Working Group on California Earthquake Probabilities (WG02) (2002). Earthquake probabilities in the San Francisco Bay region: 2002–2031, *U.S. Geol. Surv. Circular 1189.*

Youngs, R. R., S. J. Silva, and J. R. Humphrey (1997). Strong ground motion attenuation relationships for subduction zone earthquakes, *Seism. Res. Lett.* **68,** 58–73.

Appendix

COV Decreasing with Increasing Return Period

In the main text we have stated that COV decreases with increasing return period. We use a single fault source and its hazard curves at a site from different branches of the logic tree to show this in three steps. In the first step we show that the branch differences of all nodes in the logic tree can be summarized in three types: magnitude, occurrence rate, and attenuation relation, if we discretize the continuous G-R magnitude–frequency distribution. In the second step we show that these three types of differences cause a hazard curve, which is defined by a set of magnitude, occurrence rate, and attenuation relation and plotted in a log–log scale, to translate to another location without rotation. The translated hazard curve is for another set of magnitude, occurrence rate, and attenuation relation in the logic tree if we temporarily assume all the attenuation relations having the same standard deviation σ value for the lognormal distributions of ground motion. We will show that the S.D. of ground motion for the horizontally translated curves increases with increasing ground motion (return period) at the same rate. Therefore, the COV does not change with increasing ground motion (return period). But the S.D. for the vertically translated hazard curves decreases with increasing ground motion exponentially and, therefore, the COV decreases with increasing ground motion (return period). In the third step we show that the σ differences among the atten-

uation relations cause the hazard curves to diverge from each other and COV values to increase. But it is too small to alter the overall decrease of COV with increasing ground motion (return period).

Step 1. Magnitude, Occurrence Rate, and Attenuation Relation Defining a Hazard Model (Hazard Curve)

In the California logic tree (Fig. 1), to change from one branch to another is to change the magnitude, occurrence rate, and attenuation relation under each node except under the node for characteristic versus G-R modeling. The earthquake occurrence rate here is converted from the fault slip rate in the logic tree. For the G-R modeling, the magnitude integration in the hazard calculation can be discretized (see next step). For each discretized magnitude bin, the differences from a characteristic modeling are still in magnitude, occurrence rate, and attenuation relation.

Step 2. Hazard Curve Translations According to the Differences in Magnitude, Occurrence Rate, and Attenuation Relation

The AFE at a site from a single-fault source modeled with a characteristic recurrence process of magnitude m can be written as (Reiter, 1990; Cao *et al.*, 1996):

$$\lambda(u \geq u_0) = \alpha P(u \geq u_0 \mid m, r) \tag{A1}$$

where $\lambda(u \geq u_0)$ is the AFE for ground motion u exceeding given value u_0, α is the mean occurrence rate for characteristic earthquakes of magnitude m on the fault, $P(u \geq u_0|m, r)$ is the conditional probability that u exceeds u_0 when an earthquake of magnitude m occurs on the fault with a distance r to the site. Because here we are interested in one fault source and one magnitude only, all the integrations and summations in the general formula are not needed in (A1). The conditional probability term in (A1) determines the shape of the calculated hazard curve.

If the fault is modeled with a G-R recurrence process, there is an integration to magnitude to be added to (A1), which can be discretized and becomes a summation to many magnitude bins. In this case the preceding discussion on the characteristic modeling is still valid for each magnitude bin. The hazard curve for the G-R modeling is the sum of hazard curves from all magnitude bins and is dominated by the curve from the bin with the lowest magnitude because the occurrence rate decreases exponentially with increasing magnitude.

Let us first show that a different attenuation relation or a different magnitude causes a hazard curve to have a horizontal shift. For all the attenuation relations used in the 2002 California hazard model, the conditional probability is calculated from a lognormal distribution of the ground motion, u, or a normal distribution of parameter $x = \ln(u)$ with $x_0 = \ln(u_0)$. We have,

$$P(x \geq x_0 \mid m, r) = \int_{x_0}^{\infty} (\sqrt{2\pi}\sigma)^{-1} \cdot \exp(-((x - x_m)/\sigma)^2/2)dx, \tag{A2}$$

where σ is the standard deviation of the normal distribution and x_m is the log of the mean ground motion predicted by an attenuation relation, or $x_m = \ln(u_m)$. Both σ and x_m depend on the attenuation relation used and x_m also depends on m, r, and other parameters such as fault mechanism. Let us do a substitution:

$$y = (x - x_m)/(\sqrt{2}\sigma).$$

Then we get:

$$P(y \geq y_0 \mid m, r) = (1/\sqrt{\pi}) \int_{y_0}^{\infty} \exp(-y^2)dy, \tag{A3}$$

where $y_0 = (x_0 - x_m)/\sqrt{2}\sigma)$. This expression can be written in the following form using the error function, *erf*:

$$P(y \geq y_0 \mid m, r) = (erf(-y_0) + 1.0)/2. \tag{A4}$$

This result shows that a change from one attenuation relation to another, which causes x_m to change, is equivalent to a ground-motion change or a hazard-curve shift in the ground-motion (horizontal) direction; however, the shape of the hazard curve is not changed, if assuming σ is the same.

In Figure A1, the difference between curves A and B is the mean ground motion, u_m. Curves A and B have the same shape and one can be obtained from the other by a horizontal shift. So we have $\ln(u_A) - \ln(u_B) = $ const. or $u_A/u_B = $ const. for any AFE value, where u_A and u_B are the ground-motion values from curves A and B at a given AFE value. Then we have $(u_A - u_B)/u_B = u_A/u_B - 1 = $ const. or the COV does not change with increasing ground motion (return period) for a group of hazard curves that are different only by horizontal shifts. If only the magnitude is different we will have a different x_m and, again, it is equivalent to a ground-motion change or a hazard curve shift in the horizontal direction. So far, we have assumed the same σ-value for all the attenuation relations. The case with different σ-values is discussed in step 3.

An occurrence rate change will cause the hazard curve to have a vertical shift in a log–log plot (Fig. A1). This becomes clear if we take logarithm on both sides of (A1). In Figure A1, the difference between curves A and C is the occurrence rate. We see that curves A and C have the same shape, and one can be obtained from the other by a vertical shift. There is an important difference between a vertical shift and a horizontal shift in ground-motion difference between two curves for a given AFE value. It is obvious in Figure A1 that the ground-motion difference between curves

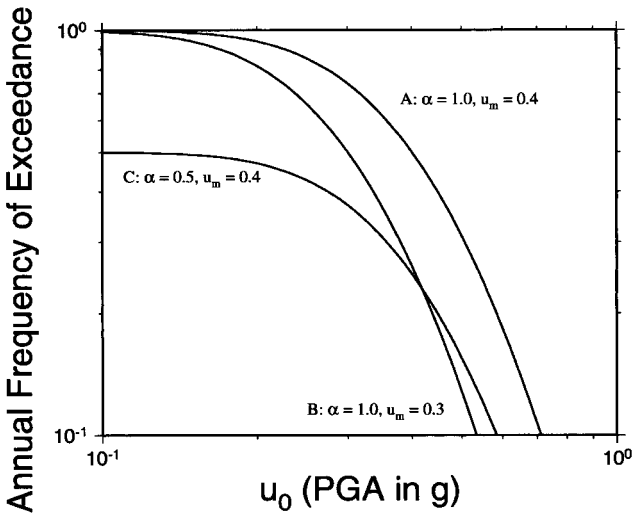

Figure A1. Three hazard curves with different occurrence rates (α) or different mean ground motions (u_m), but the same standard deviation value of σ = 0.45. Curves A and C are different only in their occurrence rates. They have the same shape and one can be obtained from the other by a vertical shift. Curves A and B are different only in their mean ground motions. They also have the same shape and one can be obtained from the other by a horizontal shift.

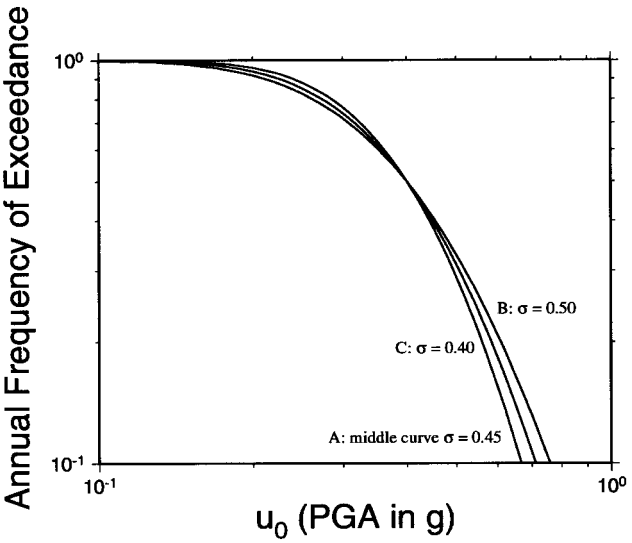

Figure A2. Three hazard curves with three different σ-values but the same occurrence rate (α = 1.0) and mean ground motion (u_m = 0.4). They have different shapes. One can be obtained from the other by a stretch or compression in the horizontal direction. The amount stretch or compression needed depends on the ratio of σ-values. These curves diverge slightly from each other.

A and C for the same AFE is no longer increasing with increasing ground motion. The numerical calculation shows that the ground-motion difference between curves A and C for same AFE decreases exponentially with increasing ground motion (return period). This is equivalent to saying that the predicted ground motion is more sensitive to the occurrence rate change at lower ground-motion levels (shorter return periods) than at higher ground-motion levels (longer return periods). This decrease of ground-motion difference causes the S.D. and COV of ground motion for a group of hazard curves that are different only by vertical shifts to decrease. It is similar to the COV decrease from P1 to P4 in Figure 8. This explains why the S.D. decreases by changing from a one-fault system to a two-fault system in Figure 9.

Step 3. Different σ-Values Causing Hazard Curves to Diverge from Each Other

In Figure A2 three different hazard curves are from three different σ-values. These values span from the upper to the lower standard deviations of the attenuation relations used in the 2002 California hazard model. Now we cannot get one hazard curve from the other by shifts. From the substitution $y = (x - x_m)/(\sqrt{2}\sigma)$, we know it needs a stretch or compression in the horizontal direction. The amount stretch or compression needed depends on the ratio of σ-values for the two curves. For example, to get curve B from curve A, curve A has to be stretched by a factor of 0.5/0.45 = 1.11. Figure A2 shows that because of the relatively narrow range of σ-values for the attenuation relations used in the 2002 California model, the hazard curves from the logic tree diverge from each other slightly if only σ differences are considered. The results in Figure 5 show that this divergence due to the σ-value differences is overwhelmed by the convergence due to occurrence rate differences.

In summary, the hazard curves of the logic tree with different occurrence rates only are convergent from each other and cause S.D. and COV to decrease with increasing return period (ground motion) along a descending mean hazard curve; the hazard curves with different magnitudes only are neither divergent nor convergent from each other and cause COV unchanged with increasing ground motion; the hazard curves with different σ-values only are slightly divergent from each other and cause COV to increase with increasing return period (ground motion). When the differences in magnitude, occurrence rate, and attenuation relation are mixed, the COV of ground motion decreases with increasing return period (ground motion) or decreasing AFE. In real cases, these three differences always act inseparably even if we wish to study the uncertainty contribution from one node only. For example, if we want to study the uncertainty due to attenuation relations only, we average the models sharing the same attenuation relation, that will suppress the uncertainty contributions from other nodes but we never can eliminate them totally. The COV may still decrease with increasing return period (ground motion) despite the previous conclusion for attenuation relation difference only. The 2002 model is a moment-balanced model. A magnitude change is always tied with an occurrence rate change. The

total effect of the previous three differences causes a net decrease of COV with increasing return period (ground motion). This reflects that the ground motion of the current hazard model is more sensitive to the input parameter uncertainties at lower ground-motion levels (shorter return periods) and less sensitive at higher ground-motion levels (longer return periods).

California Geological Survey
801 K Street, MS 12-32
Sacramento, California 95814
(T.C.)

U.S. Geological Survey
Box 25046 Federal Center, MS 966
Denver, Colorado 80225
(M.D.P., A.D.F.)

Manuscript received 28 January 2005.