

# Losses to Single-Family Housing from Ground Motions in the 1994 Northridge, California, Earthquake

Robert L. Wesson,<sup>a)</sup> M.EERI, David M. Perkins,<sup>a)</sup> M.EERI,  
Edgar V. Leyendecker,<sup>a)</sup> M.EERI, Richard J. Roth, Jr.,<sup>b)</sup> M.EERI, and  
Mark D. Petersen,<sup>a)</sup> M.EERI

The distributions of insured losses to single-family housing following the 1994 Northridge, California, earthquake for 234 ZIP codes can be satisfactorily modeled with gamma distributions. Regressions of the parameters in the gamma distribution on estimates of ground motion, derived from ShakeMap estimates or from interpolated observations, provide a basis for developing curves of conditional probability of loss given a ground motion. Comparison of the resulting estimates of aggregate loss with the actual aggregate loss gives satisfactory agreement for several different ground-motion parameters. Estimates of loss based on a deterministic spatial model of the earthquake ground motion, using standard attenuation relationships and NEHRP soil factors, give satisfactory results for some ground-motion parameters if the input ground motions are increased about one and one-half standard deviations above the median, reflecting the fact that the ground motions for the Northridge earthquake tended to be higher than the median ground motion for other earthquakes with similar magnitude. The results give promise for making estimates of insured losses to a similar building stock under future earthquake loading. [DOI: 10.1193/1.1775238]

## INTRODUCTION

The objective of the present work is to determine a relationship between earthquake loss and strong ground motion. In other words, for a given ground motion, what distribution of loss should we expect? Using insurance loss data for single-family homes from the 1994 Northridge, California, earthquake, and both observed and estimated ground motions, we construct empirical probabilistic loss curves as a function of ground motion. These loss curves offer promise for estimating losses to similar housing from scenario earthquakes, and for making probabilistic estimates of annualized loss.

We briefly describe the Northridge loss data set, our technique for fitting the observed losses within a given ZIP code with a statistical distribution, the regression of the parameters in this distribution on ground motion parameters for many ZIP codes, and

---

<sup>a)</sup> U.S. Geological Survey, MS 966, Box 25046, Denver Federal Center, Denver, CO 80225-0046

<sup>b)</sup> California Insurance Actuaries, 8821 Baywood Dr., Huntington Beach, CA 92646

the application of the resulting loss curves to estimate losses from scenario earthquakes and to estimate annualized losses. Previous related work includes Steinbrugge and Algermissen (1990) and Steinbrugge and Roth (1994).

### THE DATA SET

The 17 January 1994 Northridge, California, earthquake (magnitude 6.7) caused widespread damage in the Los Angeles metropolitan area, especially in the San Fernando Valley and environs. Following the earthquake, the California Department of Insurance requested data from the insurers providing earthquake coverage to residents of Los Angeles and Orange counties. The data set used here is derived from the responses. The data were provided by ZIP code, and indicated for each claim, among other parameters, the value for which the structure was insured for fire, and the amount of the claim for damage to the structure, for damage to the contents and for loss of use. Typically, these structures were single-family homes of wood-frame construction. Most were of one or two stories. We analyzed only the data for claims on the Form 3 (Homeowner) policies and only for losses to the structure. It should be understood that these losses, which we refer to as “structural” to distinguish them from losses to contents and loss of use, are not the same as what a structural engineer would refer to as structural damage. While a structural engineer’s definition would distinguish between structural and non-structural damage to a building, our definition of “structural” loss would include certain types of damage that an engineer would refer to as “nonstructural.” Included in our definition of “structural” loss would be cracked chimneys, cracked plaster, painting, kitchen counters, damage to garden walls, etc., but not any loss for damage to contents or for loss of use.

Although some policies distinguished between values of structures for fire and earthquake insurance, we adopted the value for fire insurance as the preferred estimate of value because, first, it existed for all structures, and second, because it was judged more likely to be an up-to-date and accurate estimate. A counterintuitive characteristic of the data set is that there was a small fraction of claims paid which exceeded the fire-insured structural value. This arises in part because frequently the fire coverage understated the cost of replacement and the insurance company had to pay a larger amount. (In a typical homeowners policy today, the insurance company will pay up to 120% of the fire coverage to replace a home in the event of fire.) Another reason the fire coverage was used as the estimate of value is that at the time of the Northridge earthquake, some homeowners bought a limited earthquake coverage that was less than the fire coverage. In this case, the insurance company did not cover replacement, and the company did not pay more than the limit of the coverage.

The data set includes 784 ZIP codes with a total of 413,854 policies with earthquake coverage for structural damage, representing about 35.2% the total homeowner policies. The total fire-insured value of the structures with earthquake coverage was \$68,683,685,630, for an average insured value of \$165,961. (The market value of the house would be this amount plus the value of the land.) In this data set there were 80,727 claims. An insignificant 211 claims were excluded from the analysis because of incomplete data, or because the indicated ZIP codes included only post office boxes.

The resulting reduced data set includes 80,516 claims for structural losses above the deductible in 266 ZIP codes. The total amount of claims paid was \$3,406,762,035. Thus, for claims exceeding the deductible, the average loss was \$42,312.

A deductible near 10% was predominant at the time of the earthquake, with about 59% of the claims having a deductible estimated between 9% and 11%. That is, the insurance companies reimbursed the policyholders only for losses greater than the deductible (say 10%) of the insured value. About 39% of the claims had a deductible less than 9%, and only about 2% had a deductible greater than 11%. We estimate that the total loss to the insurers, had the deductible universally been 10%, would have been \$3.22 billion.

The fundamental measure of loss used in the following analysis was the loss before deductible. About 12% of the claims lacked data on the loss before deductible. For these claims, losses before deductible were estimated from the losses after deductible and the value of the dwelling assuming a deductible rate of 10%.<sup>1</sup>

### CHARACTERIZATION OF LOSS

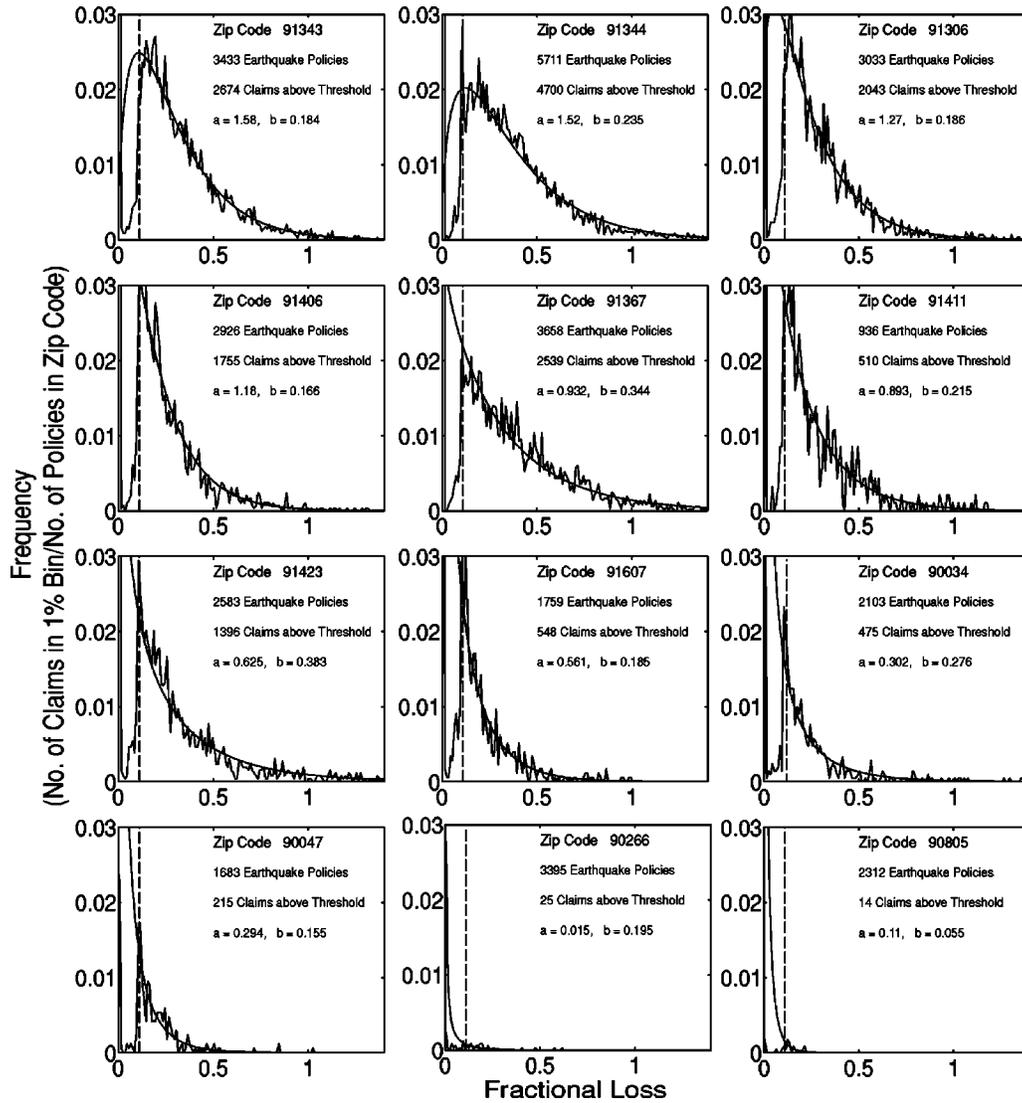
Because the geographic description of each loss is only known by ZIP code, we characterize the losses in each ZIP code by a statistical distribution of losses within that ZIP code. One simple measure of that distribution is the mean fractional loss for the entire ZIP code. For example, we could calculate the mean fractional loss after the deductible for the entire ZIP code by summing the paid claims and dividing by the total fire structural value for the ZIP code. However, since we have the claim data we are also able to consider the distribution of losses within each ZIP code. For each claim, we calculated the fractional loss, defined as

$$\text{Fractional Loss} = \frac{\text{Cost of repair for structural damage}}{\text{Fire-insured value of structure}}$$

where the cost of repair for structural damage is defined as the claim submitted before the deductible.

Then for each ZIP code, it is possible to construct a histogram of fractional loss (e.g., Figure 1). For the most heavily damaged ZIP codes, claims greater than the deductible were made against two-thirds or more of the policies in the ZIP codes. Although we do not know the value of the losses for policyholders who did not make claims, we can assume that these losses were less than the 10% deductible. Thus, although we do not know the distribution of losses less than the deductible, we can assume that the number of policies with claims less than the deductible is equal to the total number of earthquake policies minus the number of policies with claims greater than the deductible. Thus we can calculate the fraction of the total number of policies with losses less than the deductible.

<sup>1</sup> The estimated average deductible is 8.7%. Thus about 12% of the claims may have losses before deductible slightly overestimated. We judge the overall impact of this difference to be negligible.



**Figure 1.** Histograms of fractional losses for twelve representative ZIP codes. Fractional losses for claims in the ZIP code (jagged curves) are shown together with the best-fitting gamma distribution (smooth curves) obtained as described in the text. The vertical dashed line at a fractional loss of 0.11 indicates the threshold above which individual claims were considered. Since the losses for individual policies with losses less than the deductible are unknown, only the fraction of the total number of policies with losses less than this threshold is used in the fitting. Although the dominant deductible value was 0.10, we chose a threshold of 0.11 to avoid any possible aberrations at or near 0.10. About 98% of the claims had deductibles estimated at 11% or below.

Inspection of the histograms for loss in the various ZIP codes suggested that the distributions could be well fit by gamma distributions, although a Weibull or other distribution might also be considered. As will be shown below, it turns out that a desirable feature of the gamma distribution for this purpose is that relatively simple empirical relations can be established between the two parameters of the gamma distribution and ground motion.

We also carried out considerable analysis in terms of mean loss after the deductible, but we concluded that the analysis in terms of the gamma distribution captured significantly more information, and produced a more useful result.<sup>2</sup> Consideration of the whole distribution permits analysis of the high-loss tail of the distribution as well as the impact of changes in the deductible.

### FITTING GAMMA DISTRIBUTIONS TO THE LOSSES

The gamma distribution is a two-parameter distribution widely used in statistics. The probability density function for the distribution is given as (Evans et al. 2000)

$$f(x|a,b) = \frac{1}{b^a \Gamma(a)} x^{a-1} e^{-x/b} \quad (1)$$

The two parameters in the distribution, both greater than zero, are commonly referred to as  $a$ , the shape parameter, and  $b$ , the scale parameter. It should be noted that the mean of a gamma distribution is the product of the two parameters,  $ab$ . Thus, for a nearly constant  $b$ , the mean loss before deductible is proportional to  $a$ , and we can consider  $a$  as a surrogate for the mean loss. The coefficient of variation is  $a^{-1/2}$ .

For a scale parameter of  $b=0.2$ , typical of the distributions obtained below, the gamma distribution varies from a delta function at the origin as  $a$  approaches zero, to a decaying exponential shape for  $a$  less than about 1, to an asymmetrical hump-shaped distribution for  $a$  greater 1. Parameters in the gamma distribution are commonly estimated using the maximum likelihood approach. In our data set, information on losses less than 10% is incomplete, but we do know the fraction of losses less than that value. Consequently, we employed a maximum likelihood approach for estimating the parameters with the constraint that the fraction of losses less than a threshold must equal the observed fraction. (That is, the area under the gamma distribution for fractional losses less than the threshold must equal the fraction of insured houses with fractional losses less than the threshold.) To avoid any possible aberrations near the deductible value of 10%, we adopted a threshold value of 11%. Confidence intervals on  $a$  and  $b$  were determined using a bootstrap technique.<sup>3</sup>

<sup>2</sup> As will be shown below in Equation 3, a relationship between mean loss before deductible and mean loss after deductible can be written, assuming that the loss follows a gamma distribution and that the houses with losses follow the same value distribution as those without losses.

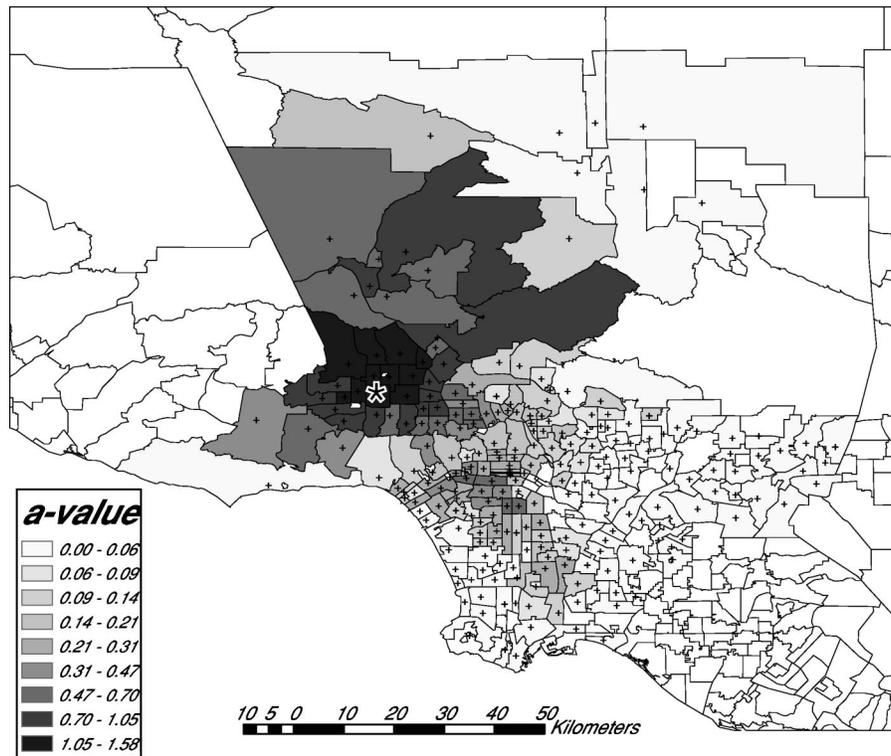
<sup>3</sup> In the bootstrap technique, the uncertainties in the calculated parameters are estimated by repeatedly resampling the data with replacement and considering the range of the parameters so calculated, c.f. Efron and Tibshirani, 1993.

For a satisfactory estimate of the parameters in the gamma distribution, we required that the ZIP code have more than one claim for earthquake structural damage and that the width of the 95% confidence interval for  $a$  be less than 1.0. By these criteria, we were able to obtain satisfactory estimates of the parameters for 234 residential ZIP codes out of the 266 ZIP codes for which losses were reported. Examples are shown in Figure 1. Most of the ZIP codes for which the fits were rejected had only one or a few claims. The two ZIP codes rejected, which had the largest number of claims, had 10 and 11 claims, respectively. The 234 ZIP codes for which parameters were calculated accounted for about 99.95% of the total insured loss.

Generally the scale parameter,  $b$ , in the gamma distributions obtained by this procedure is near 0.2, but the shape parameter,  $a$ , varies systematically from a high value near 1.6, for ZIP codes characterized by high damage, to low values approaching zero for ZIP codes experiencing very light damage. Our fitting procedure fails when there are no claims greater than the threshold, but it is not unreasonable to argue that  $a$  tends to zero as the losses tend to zero. The geographic pattern of the parameter  $a$  (Figure 2) shows that the high damage areas are most heavily concentrated in the areas closest to the rupture, and that the values generally diminish with distance away from the rupture. Some broad site and/or basin effects are evidenced by the distribution of  $a$ . In general, the values of  $a$  decrease with distance from the rupture zone, but a region of higher  $a$  values in the Los Angeles Basin lies to the south of a band of lower  $a$  values corresponding to the Santa Monica Mountains and Hollywood Hills.

### ESTIMATES OF GROUND MOTION

The Northridge earthquake produced one of the largest sets of strong-motion recordings obtained up to that time, but the recordings still constituted a sparse sampling of the actual wave field and were rather unevenly distributed throughout the region. As discussed above, the geographic locations of the sites of the claims are known only by ZIP code. Thus we chose to characterize the ground motion typical of each ZIP code by an estimate of the ground motion at the population centroid of the 1991 ZIP codes (Blodgett 2002). We considered estimates of the ground motions made in several ways, but in the end the most satisfactory were judged to be estimates made by ShakeMap (Wald et al. 1999a, b), and estimates made by directly interpolating the observed peak ground motions. ShakeMap provides estimates of the peak ground acceleration (PGA), the peak ground velocity (PGV), instrumental intensity (IIM), and spectral accelerations at 0.3 sec, 1 sec, and 3 sec. Readers interested in details should consult the publications of Wald et al. In brief, the ShakeMap interpolation is made by correcting the observed ground motions to a common site condition, using a correction appropriate for each recording site, then estimating the ground motions throughout the region by interpolation, corrected for the predicted site effects as determined from the near-surface site condition map of Wills et al. (2000). Although one may anticipate many improvements to the process of making ShakeMap estimates in the years to come, their undisputed virtue is that they provide a systematic estimate of the ground motion at each point in the region, taking into account known near-surface geology. Of course, perhaps the biggest improvement in ShakeMap estimates will come from an increased geographic density of record-

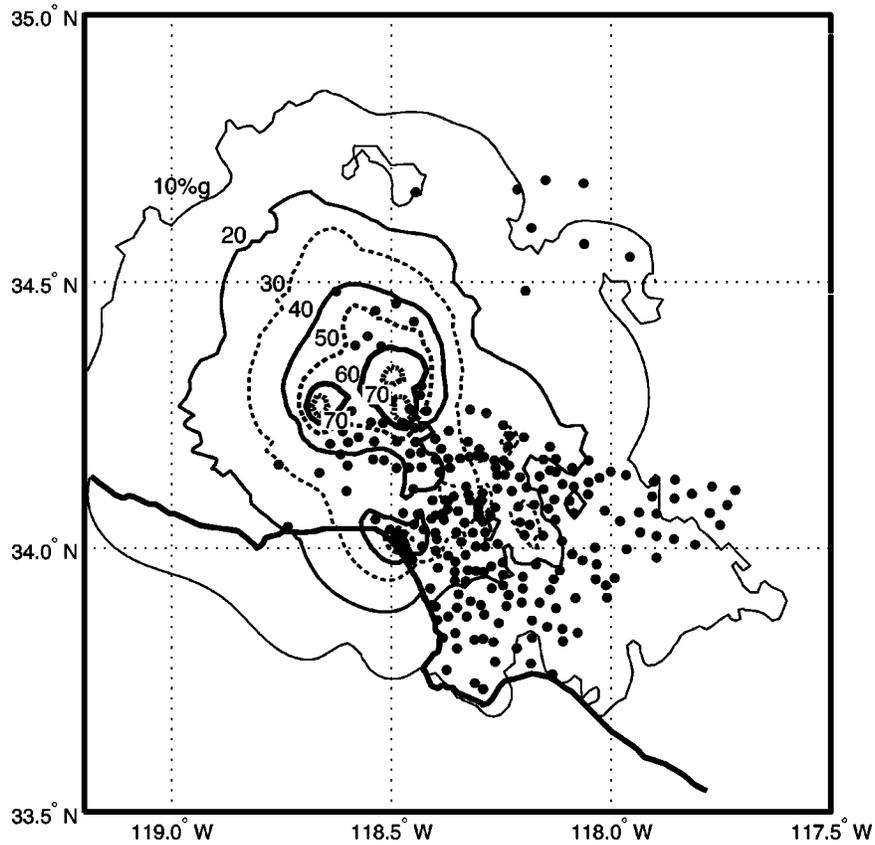


**Figure 2.** Geographic distribution of loss parameter,  $a$ . Each outlined ZIP code is colored according to the value of  $a$  determined from the losses in that ZIP code. Blank ZIP codes indicate regions where no data were available or where reliable  $a$  values could not be determined. The small crosses indicate the population centroids of the 1991 ZIP codes as determined by Blodgett (see text). For ZIP codes that are not geographically compact, the centroids suggest the geographic weighting of the loss data. Ground motion estimates were made for each ZIP code at these centroids. Star indicates epicenter of Northridge earthquake.

ing, and it should be noted that owing to the TriNet project, the number of recording sites has been increased significantly since the Northridge earthquake.

In addition, PGA and PGV values were interpolated at the centroid of each ZIP code directly from the observations using the Delaunay triangulation interpolation scheme implemented in MATLAB. In this scheme the interpolated surface fits each observed value exactly. All observed free field PGA and PGV values in the region were used except those derived from the recordings at Santa Monica City Hall and Tarzana.<sup>4</sup> These values appeared grossly inconsistent with both the ground motion and loss data sets and were discarded. (As will be shown below, the quite significant levels of loss in Santa Monica are consistent with ground motion values one-third to one-half those recorded at

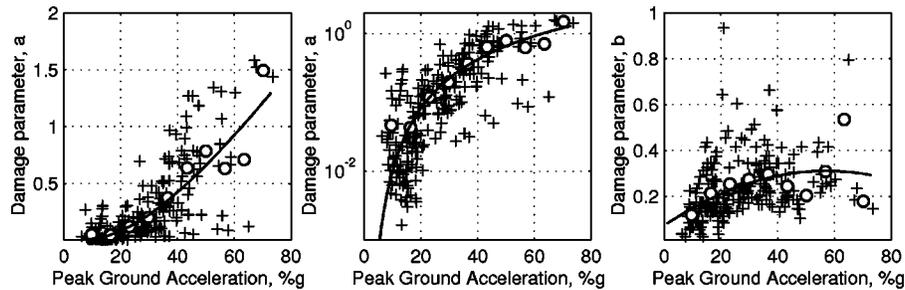
<sup>4</sup> The Tarzana recording is also ignored in ShakeMap.



**Figure 3.** An example of ground motion estimates for peak ground acceleration determined from ShakeMap. Black dots indicate population centroids of ZIP codes at which the ShakeMap values were interpolated. Contours slightly simplified for clarity. Peak estimates within the 70% g contours exceeded 80% g in small areas.

Santa Monica City Hall.) Inasmuch as the Delaunay interpolated values ignore what is known about site conditions, both at the observation sites and throughout the region, we are inclined to prefer the ShakeMap values, and include the interpolated values primarily as a check.

In total we consider eight estimates of ground motion parameters at each ZIP code centroid (PGA, PGV, Instrumental Intensity, 0.3 sec SA, 1.0 sec SA, and 3 sec SA from ShakeMap, and PGA and PGV from Delaunay interpolation). An example of these estimates for the region is shown in Figure 3.



**Figure 4.** Examples of loss parameters  $a$  and  $b$  plotted as a function of peak ground acceleration inferred from ShakeMap. Circles indicate means of data in each of 10 bins of equal width selected to cover the range of the ground motion parameter (here PGA). Note that the center plot uses a logarithmic ordinate.

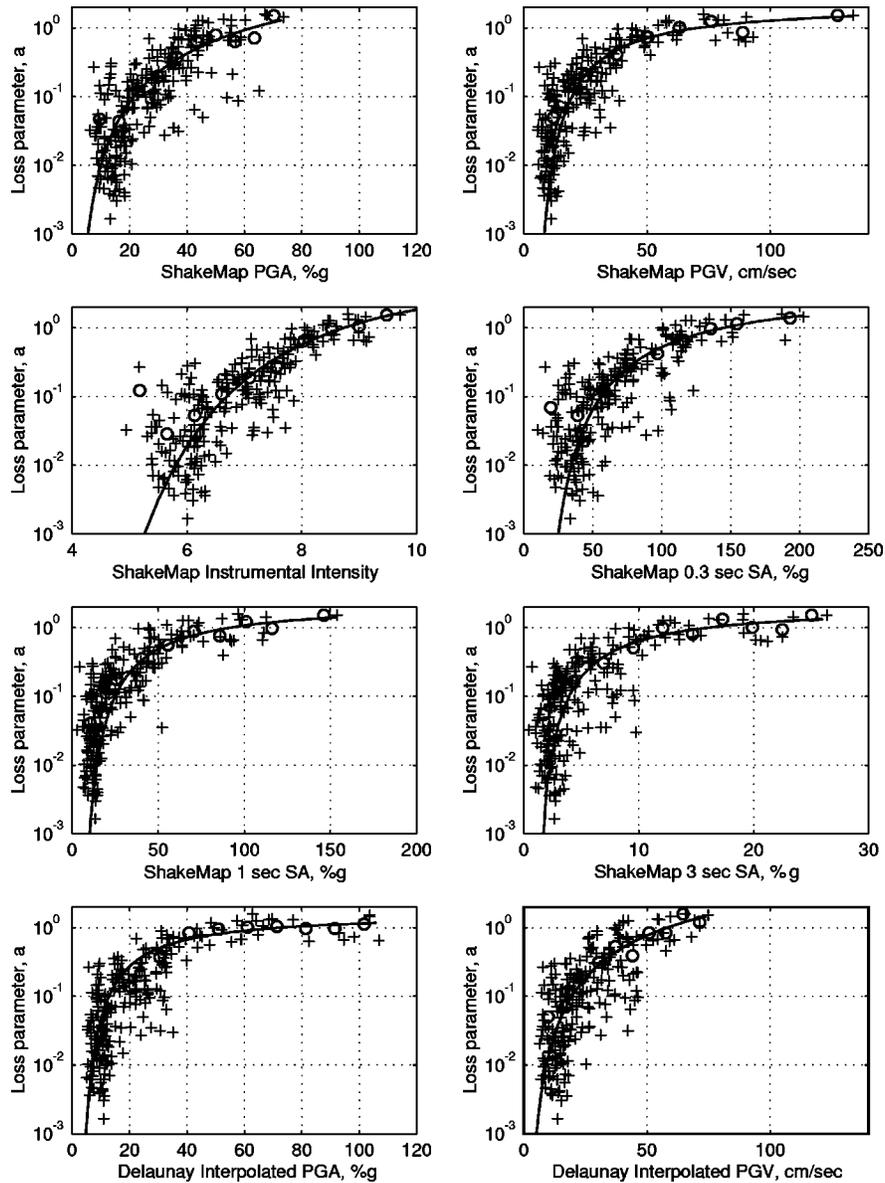
### REGRESSION OF LOSS PARAMETERS ON GROUND MOTION

Plots of the two parameters in the gamma distributions,  $a$  and  $b$ , for each ZIP code versus the eight ground-motion estimates listed above at the corresponding ZIP code centroid show considerable scatter, but nonetheless exhibit some strong tendencies.

Figure 4 shows examples of plots of  $a$  and  $b$  from each ZIP code versus the corresponding estimates of PGA from ShakeMap. The shape parameter  $a$  clearly increases rapidly with increasing PGA at small values of PGA, then less and less rapidly for higher values of PGA. This tendency is clearer in the semilog plot than in the linear plot. This behavior of  $a$  with increasing ground motion is also true for all the other measures of ground motion considered. In contrast, there do not seem to be any strong systematics in the variation of the scale parameter,  $b$ . In most cases,  $b$  scatters around a value near 0.2. The inquiring reader may ask, why, if there are no particular systematics to the variation of  $b$ , should its variation be considered. The answer is that carrying  $b$  through the regressions significantly improves the precision of the loss estimates.

Semilog plots of  $a$  versus the eight ground-motion parameters being considered are shown in Figure 5. In all cases  $a$  tends to zero at low ground motions, but at some ground motion level it begins to increase sharply with increasing ground motion. For about half of the ground motion parameters, e.g., instrumental intensity,  $a$  continues to increase across the range of estimated ground motions. For the other half, e.g., Delaunay interpolated PGA, there is a tendency for  $a$  to flatten out at large ground motions. Physically, this flattening out would mean that increasing ground motions above this range would not result in additional loss.

Because there are so many more observations at low ground-motion values (where the losses are small) as compared to the high ground-motion values (where losses are large), the observations for each ground motion parameter were divided into 10 bins with equal intervals and the bin means calculated. This approach has the characteristic of



**Figure 5.** Semilog plots of the loss parameter  $a$  versus the eight ground-motion parameters considered. The open circles show the bin means discussed in the text. The smooth curves are the best fitting equations of the form shown in Equation 2. In some cases, the parameter  $\delta$  in Equation 2 was constrained to equal one (see Table 1).

giving higher weight to the individual observations at high ground-motion levels. Regressions were then performed using the bin means for the  $a$  and  $b$  values at the mid-points of the ground motion bins.

Initially the regressions were attempted by fitting polynomials in the ground motion to the values of  $\log a$  and  $\log b$ . This approach was abandoned for  $a$  for two reasons. First, fitting in log space resulted in aggregate sums of losses (discussed in more detail below), which systematically underestimated the observed values. This is true because the parameters determined from regressions in log space systematically underestimate the values for large ground-motion values. Consequently, this approach underestimates the losses in ZIP codes with high ground motions where the losses are the largest components of the sum, thus leading to large errors in the sum. Therefore, regressions were carried out in linear space leading to no bias in the sum. Second, in linear space polynomials are poorly suited to the variation in  $a$  with ground motion, in particular because no negative values of  $a$  are permitted in the gamma distribution, and because  $a$  should tend asymptotically to zero at low ground motions and perhaps saturate at large ground motions. In place of polynomials, a three-parameter sigmoidal function was constructed for fitting  $a$  as function of a ground motion parameter,  $x$ , with the desired properties:

$$a(x) = \alpha \exp \left[ - \left( 1 + \frac{1}{\beta x^\delta} \right) \right] \quad (2)$$

This function has the desired asymptotic behavior as  $x$  approaches zero, can climb steeply approaching a linear growth with increasing  $x$ , and can flatten out if required. Examples of the regressions are shown in Figure 5. Note that Equation 2 has three free parameters,  $\alpha$ ,  $\beta$ , and  $\delta$ . Initially we used only two free parameters, that is with  $\delta=1$ . However, for some ground motion parameters, we required all three parameters to obtain satisfactory fits. Despite the scatter in the data, the fits are not bad, particularly when viewed in log space. The polynomial fitting procedure was retained for  $b$ . Regression parameters are shown in Table 1. We have not shown goodness-of-fit estimates for these parameters because the ground motion variability is so large as to render such estimates not particularly meaningful. Indeed, as mentioned above, the fits were carried out using bin means to stabilize the fitting process, and to increase the weight given to observations at higher ground-motion values.

### CONSTRUCTION OF CONDITIONAL LOSS CURVES

Using the regression parameters in Table 1, the dependence of  $a$  on the ground motion parameters from Equation 2 and of  $b$  on ground motion through a polynomial, and the gamma distribution in Equation 1, we can construct conditional probability-density functions (CPDF) for loss given a ground motion. An example is shown in Figure 6 for the ShakeMap 0.3 sec spectral acceleration, but similar relationships and sets of curves were constructed for all eight ground-motion parameters. In all cases, at very low ground motions the CPDF for loss approaches a delta function at zero loss. As the ground motions increase, the CPDF displays an exponential-like character. As the ground motion continues to increase, the CPDF shows an asymmetrical hump-like behavior.

**Table 1.** Parameters in regressions for  $a$  (Equation 2) and  $b$  ( $b(x)=p_0+p_1x+p_2x^2$ ), where  $x$  is the ground motion parameter. Entries denoted with asterisk indicate that  $\delta$  in Equation 2 was constrained to equal 1. These regressions are not valid beyond the observed range of the ground motion parameters, for example 0 to  $\sim 75$  cm/sec for Delaunay interpolated PGV.

Ground Motion Parameter	$\alpha$	$\beta$	$\delta$	$p_0$	$p_1$	$p_2$
ShakeMap PGA	346.49	.045618	.36456	.072583	.0083624	$-7.3239 \times 10^{-5}$
ShakeMap PGV	5.9099	.013531	1.0697	.22134	.00097165	$-7.3289 \times 10^{-6}$
ShakeMap Instrumental Intensity	15.723	.00062319	3.1458	-1.4874	.44895	-.028465
ShakeMap 0.3 sec SA	11.476	.0047879	1*	.028897	.0049338	$-2.2291 \times 10^{-5}$
ShakeMap 1 sec SA	6.4241	.012651	1*	.19822	.0014956	$-9.3265 \times 10^{-6}$
ShakeMap 3 sec SA	5.7134	.083894	1*	.18718	.0085717	$-2.8135 \times 10^{-4}$
Delaunay Interpolated PGA	4.4670	.027863	1*	.17387	.0021498	$-1.3083 \times 10^{-5}$
Delaunay Interpolated PGV	182.19	.046912	.39954	.098608	.0085343	$-9.8144 \times 10^{-5}$

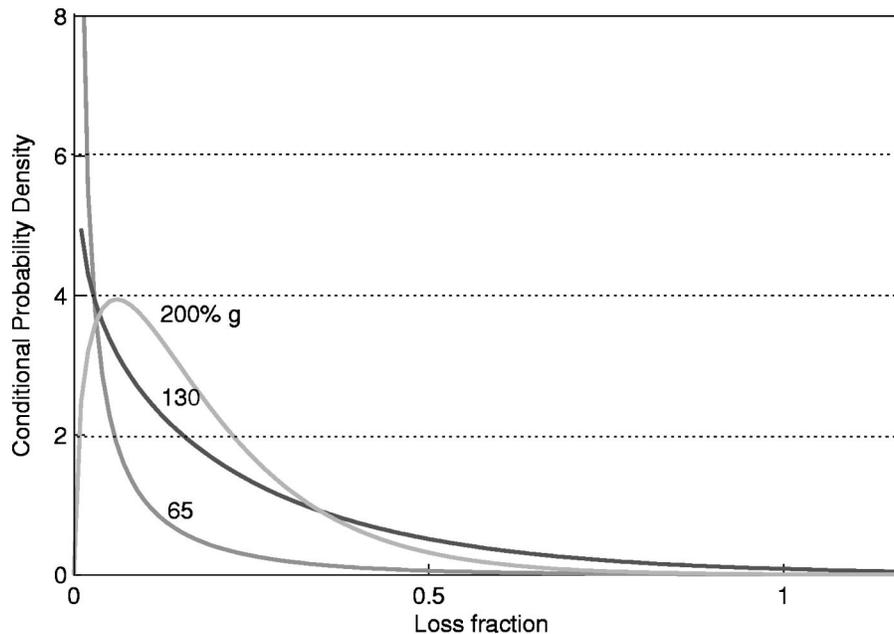
### COMPARISON OF OBSERVED AND PREDICTED LOSSES

As a check on the analysis described above, mean fractional loss to the insurer were calculated for each ZIP code for each of the ground motion parameters using the CPDF curves described above. It can be shown that if the fractional losses (ground up losses in insurance terminology) in a ZIP code follow a gamma distribution with parameters  $a$  and  $b$ , then the mean fractional loss to the insurer (that is, the losses to the insurer after the insured pays the deductible) is given by

$$\mu_{\text{loss\_to\_insurer}} = ab[1 - F(d|a+1, b)] - d[1 - F(d|a, b)],$$

where  $d$ , is the fractional deductible and  $F(x|a, b)$  is the cumulative distribution function for the gamma distribution. These predicted mean fractional losses to the insurer can then be compared with the actual for each ZIP code. Figure 7 shows an example of this comparison for the ShakeMap PGA. If a prediction were exact, the corresponding point would lie along the diagonal line. Knowing the predicted mean fractional loss to the insurer and the sum of the fire structural values for the properties insured against earthquake loss in each ZIP code, we can then predict the total dollar loss to the insurer in each ZIP code. Summing all ZIP codes we can predict the total dollar loss to the insurers. Figure 8 shows examples of the predicted dollar losses to the insurers for each ZIP code versus the actual losses. Again, if the predictions were perfect, all points would lie along the diagonal line.

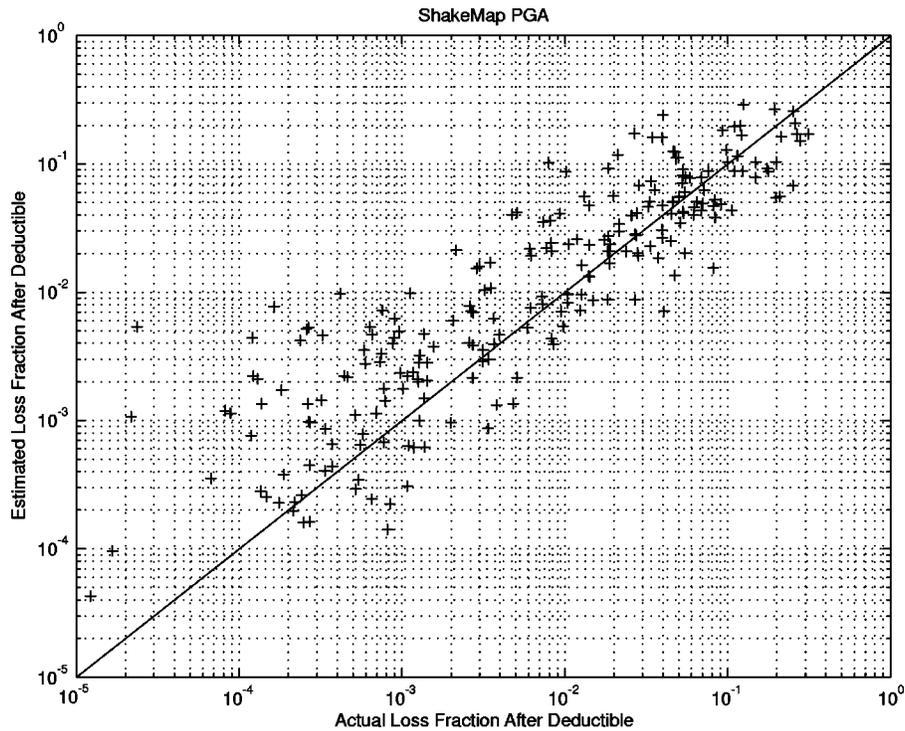
Table 2 shows the actual and predicted aggregate losses to the insurers for all ZIP



**Figure 6.** Examples of the conditional probability density functions (CPDF) for fractional loss given ground motion for values of the ShakeMap 0.3 sec spectral acceleration of 65, 130, and 200%  $g$ . At this scale, the curve for 10%  $g$  is indistinguishable from a delta function at the origin. Note the change of shape with increasing ground motion value, from a delta function at the origin for low ground motions, to a decaying exponential to a hump-shaped distribution for larger ground motions. This is characteristic of the CPDF's for all ground-motion parameters.

codes for all eight ground-motion parameters considered. Several points should be made about Figure 8 and this table. First, agreement among the actual and predicted total losses to the insurers is not an independent check of the approach described here, but it does provide some insight into the potential usefulness of the various ground motion parameters for loss prediction purposes. Second, the sums are dominated by the ZIP codes with the largest losses. Small changes in the predicted  $a$  values for these ZIP codes lead to significant changes in the predicted aggregate losses. Thus, given the dominance of a relatively small number of ZIP codes, one might not imagine that the sum is an extremely stable quantity.

Nonetheless, Table 2 shows that six of the eight estimates are within about 10% of the observed total, although the estimates are biased toward the low side. (The “actual” value in Table 2 has been reduced to adjust for the amounts paid for deductibles less than 10%. It therefore represents an estimate of what the insurers would have paid had a 10% deductible been universal.) Consequently, comparison of the sum, while tempting and useful, should not be the only metric for judging the efficacy of the predictions. Comparison of the fractional losses should also be considered. The width of the spread of the predictions should also be taken into account. In Figure 8 the spread of the values of

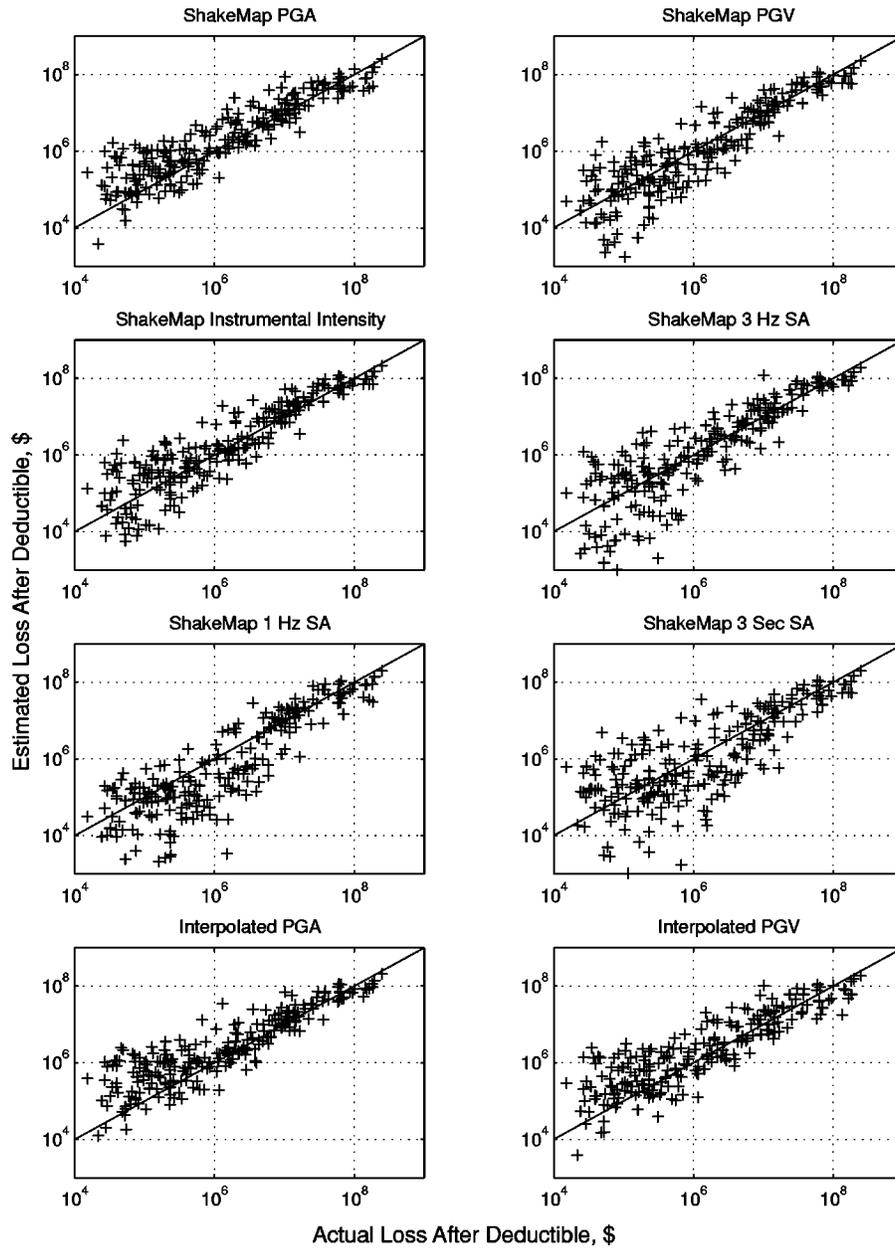


**Figure 7.** An example of the predicted and actual mean fractional loss using the ShakeMap PGA for 10% deductible.

ShakeMap 0.3 sec SA, for example, is much less than that for ShakeMap 3 sec SA, suggesting that the higher-frequency ground-motion parameter is a more consistent predictor of loss. Thus, considering Figure 8 and Table 2, it would appear that only ShakeMap 1 sec and 3 sec SA are indicated as inferior predictors of loss for this class of structures, that is, one- and two-story wood-frame, single-family dwellings. Since the fundamental period of these structures is much closer to 0.3 sec, than to either 1 sec or 3 sec, this result is not unexpected. The importance of the apparent differences among the results for the other ground motion parameters remains to be determined.

#### **IS THE FRACTIONAL LOSS INDEPENDENT OF THE VALUE AND AGE OF THE HOME?**

An implicit assumption in the foregoing fractional loss characterization is that the fractional loss to a particular residence does not depend on the value of the residence. To test that assumption, data from all claims were sorted by the fire structural value of the residence, irrespective of the geographic location of the property. The data were then divided into bins or cohorts of 1,000 properties according to this order, resulting in 80



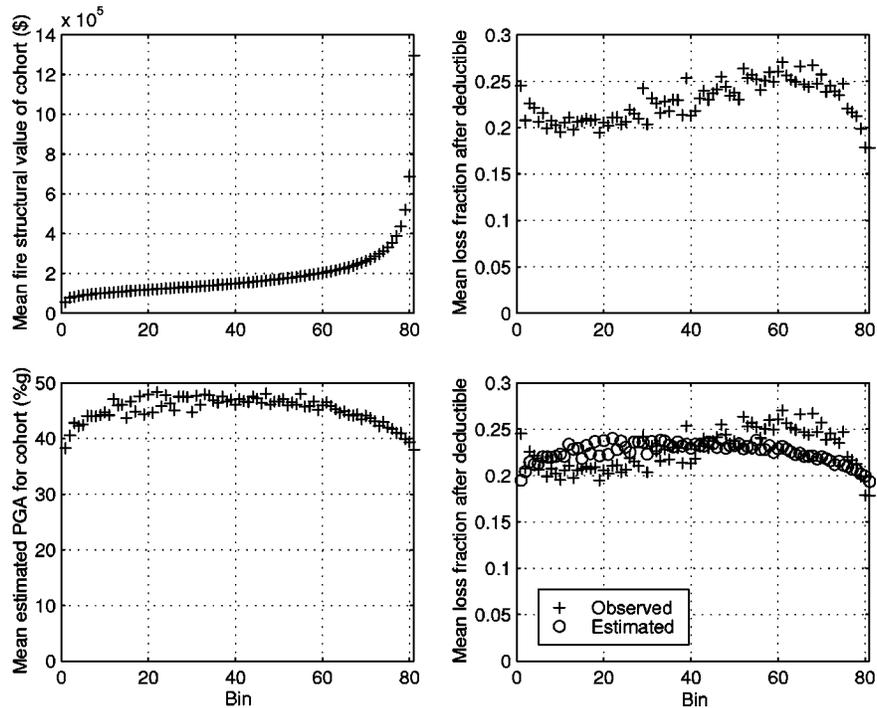
**Figure 8.** Predicted and actual losses by ZIP code for 10% deductible and for the eight ground-motion parameters considered.

**Table 2.** Actual and estimated aggregate losses for 10% deductible for 234 ZIP codes. Entries denoted with an asterisk indicate that  $\delta$  in Equation 2 was constrained to equal 1.

Losses	\$ Billions
Actual Loss	3.2
Estimated Aggregate Losses	
ShakeMap PGA	3.1
ShakeMap PGV	2.9
ShakeMap Instrumental Intensity	3.3
ShakeMap 0.3 sec SA	3.2*
ShakeMap 1 sec SA	2.5*
ShakeMap 3 sec SA	2.6*
Delaunay Interpolated PGA	3.2*
Delaunay Interpolated PGV	3.0

bins of 1,000 properties, and one, the highest value bin, of 615 claims. The mean fractional loss was calculated for each bin. Results are shown in Figure 9. Several observations deserve mentioning.

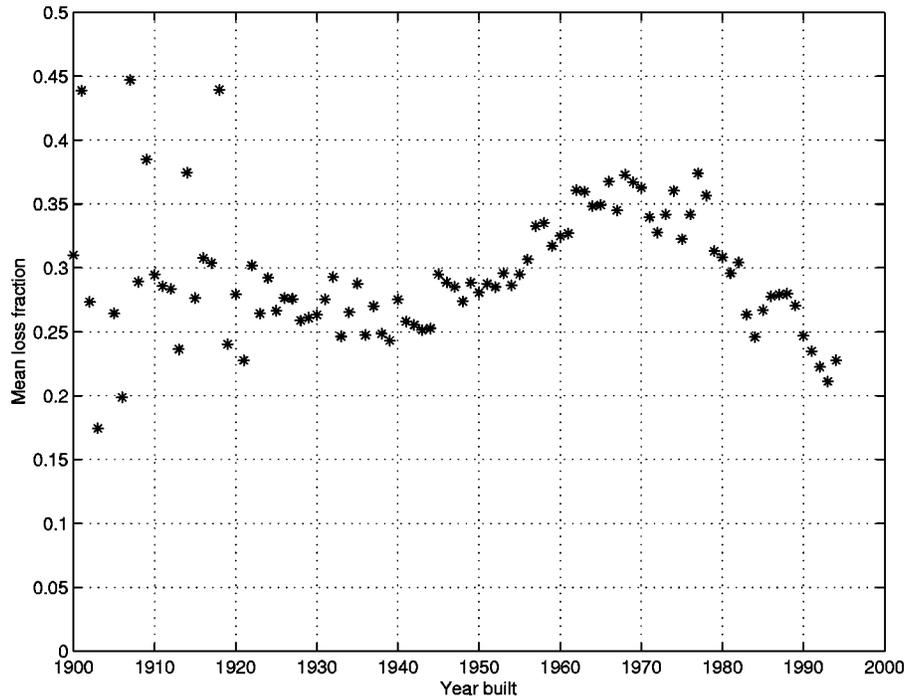
First, the fire structural value grows only very slowly from below \$100,000 to \$200,000 in about the 60<sup>th</sup> bin. For this entire range of value, the mean fractional loss after deductible in the bin varies between 0.2 and 0.25. Nonetheless, this variation appears to be systematic with slightly higher fractional losses to homes with the lowest fire structural values, a decrease to about 0.2 at about the 10<sup>th</sup> bin, then a gradual increase to about 0.25 by the 60<sup>th</sup> bin. Above the 60<sup>th</sup> bin, the fractional loss appears to decrease significantly. The mean estimated PGA is estimated for each bin as a surrogate for the mean ground motion (and indirectly, for distance from the fault rupture). Interestingly, on average, the homes with the lowest fire structural values also experienced lower ground motions on average. This observation would be consistent with lower fire structural values in the Los Angeles Basin at greater distances from the rupture. Similarly, the homes with the highest fire structural value experienced lower ground motions on average. This is consistent with the highest fire structural values, on average, being located in the band along, and just to the south of, the Santa Monica Mountains and Hollywood Hills. These properties on average are a little farther from the rupture than the properties in the San Fernando Valley, and may also have experienced a more favorable site condition on average. The comparison of the predicted mean loss fraction with the observed, however, shows that the explainable part of the difference in ground motion does not explain satisfactorily the variation in loss fraction with value. There remain some unexplained factors, perhaps systematic under- or overvaluation for the fire structural value.



**Figure 9.** Analysis of mean fractional loss for bins of 1,000 properties sorted by fire structural value. The values on the abscissa indicate the number of the bin from 1 to 81.

In conclusion, our assumption that the fractional loss does not depend on value is approximately true, but there is some interesting and unexplained systematic variation.

One factor that can be considered is the age of construction. Figure 10 shows the mean fractional loss before deductible as a function of the year of construction of the dwelling for dwellings with structural loss claims. Since the data set does not include information about houses that were not damaged, or were damaged less than the deductible, these fractional losses do not reflect losses to the entire housing stock. Nonetheless, they give a view of the conditional losses, in essence, the severity of loss, given a loss. It appears likely that the observed variation reflects different styles of construction and the evolution of building codes, rather than simply the age of the homes. The large scatter in the fractional loss for the older homes may result from smaller numbers of homes in the data set. The peak in the fractional loss in the 1960s and 1970s may reflect the popularity of irregular shaped homes (split levels), and large open spaces (cathedral ceilings) during those decades. The downturn in the fractional loss for homes built in the 1980s and 1990s may reflect improvements in the seismic provisions of the building codes. There also may be some geographic bias arising from the relative proximity to the earthquake of regions that were developed at the same time.



**Figure 10.** Mean fractional loss before deductible versus year of construction of dwelling.

### PREDICTED LOSSES FOR DETERMINISTIC NORTHRIDGE EARTHQUAKE SCENARIOS

In the foregoing, the ground motions inferred from ShakeMap have been taken as “observed.” In this section we attempt to estimate the losses for the Northridge earthquake without knowledge of the actual ground motions. That is, starting with a deterministic model of the rupture surface and magnitude of the Northridge earthquake, we estimate the losses.

Many attenuation relationships are available that predict ground motions as a function of distance and magnitude. A combination of three of these relationships, for example, was used to estimate peak ground acceleration from crustal earthquakes in California in the preparation of the National Probabilistic Seismic Hazard Map (Frankel et al. 1996). Such relationships can be used to estimate ground motion scenarios for postulated earthquakes. These ground motion scenarios can be adjusted for site conditions, and together with the loss functions described above, be used to estimate losses. We can develop an estimate of the average ground motion over a small region (e.g., a ZIP code) by summing estimates for the ground motions corrected for each site condition present, with each of these weighted by the fraction of the area covered by that site condition. In other words, if two site conditions are present in a ZIP code covering, say, one-third and two-thirds of the area, respectively, then the estimate of the average ground motion would be one-third the estimate of the ground motion for the first site condition, plus

two-thirds of the estimate for the second site condition. Two approaches were considered. In the first approach, ground motions were determined from attenuation curves specific to soil types. In the second approach, ground motions were estimated from the attenuation curve for a specific soil type (the so-called B-C boundary soil condition used in the National Map), and then site corrections were made for soil type using the NEHRP soil factors (BSSC 1995, 2001).

Calculations are shown here for a deterministic model based on the inferred rupture scenario for the Northridge earthquake. The geometry of the assumed fault plane was that used in the 1996 National Seismic Hazard Map (Petersen et al. 1996). The fractional area of each ZIP code covered by each of the standard soil types was determined using Geographic Information System analysis from the map of Wills et al. (2000). In the first approach, ground motions at the population centroid of each ZIP code were determined from the Boore-Joyner-Fumal (BJF) attenuation curves for the various soil types based on the measured shear wave velocities at the recording sites (Boore et al. 1997). Then the weighted average ground motion was determined according to the fractions of each soil type present. In the second approach, the standard set of attenuation curves used for the National Seismic Hazard Map were used to give an estimate of the median ground motion at the centroid of each ZIP code. Ground motions were calculated by application of the NEHRP soil factors for each of the soil types present in the ZIP code. These ground motions were then weighted in proportion to the fraction of area in the ZIP code covered by the corresponding soil type.

The weighted ground motions were then used to determine the  $a$  and  $b$  values from the regressions described above. Then the corresponding loss curves were used to calculate the mean fractional loss above a 10% deductible for each ZIP code. The results were multiplied by the total fire structural value for each ZIP code, and then summed to estimate the total loss. Results for the two approaches and various ground-motion parameters are summarized in Table 3.

As mentioned above, the aggregate sum of loss for the region is not an entirely satisfactory metric. Some ground motion relationships, for example, might overestimate ground motions near the source and underestimate ground motions at greater distances, but do better in estimating aggregate loss than a relationship that misestimates the ground motions by a constant factor. As can be seen from Figure 11, the ShakeMap ground motions are significantly higher than the median ground motions estimated by the National Map model with NEHRP soil corrections. Indeed, the median ground motions multiplied by 1.3 followed by the application of the NEHRP factors are a much more satisfactory fit to the ShakeMap data. It is well known that the Northridge ground motions were larger than the median values predicted by the strong-motion data collected to date. The factor of 1.3 corresponds to ground motions about one and a half inter-event standard deviations above the median. (Recall that the inter-event standard deviation of the BJF attenuation relationship is 0.08 in  $\log_{10}$  ground motion, or a factor of about 1.2.)

Thus the relatively poor agreement between the loss estimates based on the median deterministic model and the actual losses stems largely from the fact that the observed ground motions for the Northridge earthquake were high relative to the standard attenu-

**Table 3.** Loss estimates for 10% deductible for deterministic model of the Northridge earthquake calculated with various attenuation relationships and soil factors (see text). The estimates for the National Map ground motions with NEHRP soil factors, include in addition to the median values, estimates for 1.1, 1.2, 1.3, and 1.4 times the median before application of the NEHRP soil factors.

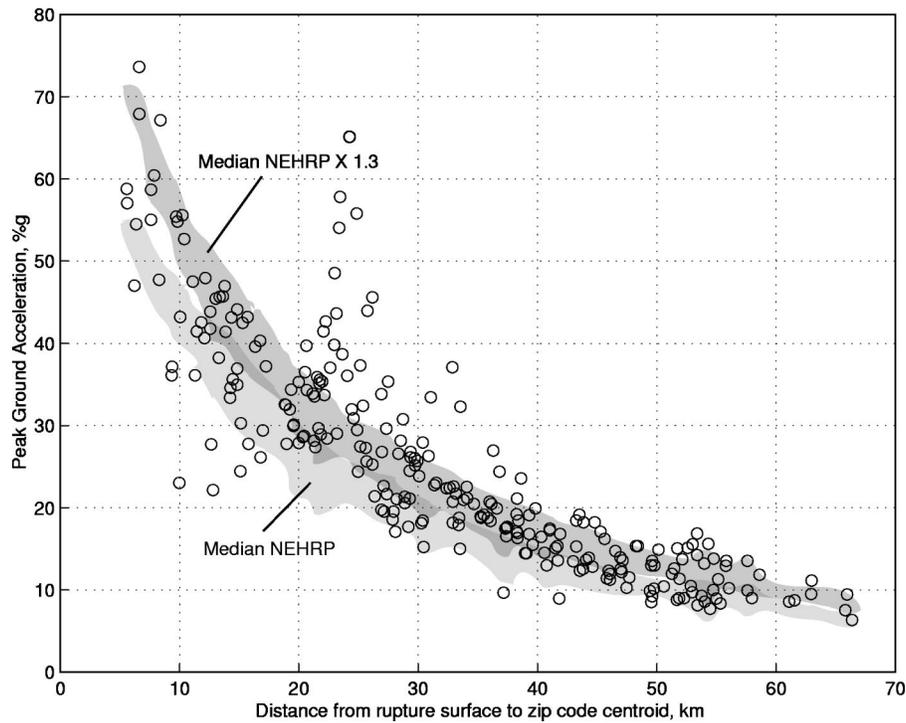
	Aggregate Losses to Insurers, \$ Billions				
	Median	Median× 1.1	Median× 1.2	Median× 1.3	Median× 1.4
Actual Loss	3.2				
Estimated Losses					
Boore-Joyner-Fumal PGA Weighted for Soil Types	2.9				
Boore-Joyner-Fumal 0.3 sec SA Weighted for Soil Types	3.2				
Boore-Joyner-Fumal 1 sec SA Weighted for Soil Types	2.3				
National Map PGA with NEHRP Soil Factors	1.8	2.3	2.8	3.3	3.9
National Map 0.3 sec SA with NEHRP Soil Factors	2.0	2.4	2.8	3.2	3.6
National Map 1 sec SA with NEHRP Soil Factors	1.9	1.9	2.3	2.5	2.8

ation relationships upon which the deterministic prediction is based. That is, the Northridge earthquake gave rise to higher ground motions and, consequently, higher losses than those predicted by the standard attenuation relationships. Notice from Table 3 that when the median ground motions are multiplied by a factor of 1.3, the estimated aggregate losses are much closer to the observed.

Table 3 makes an extremely important point. The loss estimates are critically dependent on the estimates of ground motion. Notice that a 30% increase in the estimated PGA values leads to a 43% increase in the estimated losses.

### PROBABILITY DENSITY FUNCTION FOR ANNUAL LOSS: AN EXAMPLE

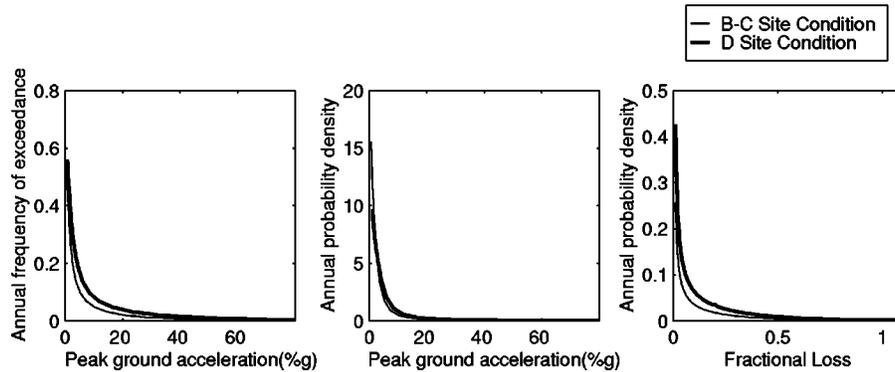
An analogous procedure can be used to estimate the probability density function for annual loss from the hazard curves developed as a part of the 1996 National Seismic Hazard Map for Southern California (Frankel et al. 1996, Petersen et al. 1996). Following a method similar to that of Cao et al. (1999), the hazard curve, originally calculated for a B-C boundary site condition, can be corrected to other site conditions using the NEHRP factors, then converted to a probability density function (PDF) for annual



**Figure 11.** Comparison of ShakeMap and deterministic predictions of peak ground accelerations (PGA). Circles indicated ShakeMap values of PGA at ZIP code centroids (that is, “observed” as interpolated by ShakeMap). The two gray clouds indicate the fields of values from deterministic ground-motion predictions at the centroid of each ZIP code. “Median NEHRP” denotes the median PGA adjusted by the NEHRP soil factors and weighted by the fractional area of each soil type found within the ZIP code. “Median NEHRP×1.3” denotes median PGA multiplied by 1.3 then adjusted by the NEHRP soil factors as above. See text for discussion.

ground motion (Algermissen and Perkins 1976, Wesson and Perkins 2001). Then this curve can be integrated with the conditional PDF for loss to give the PDF for annual loss. Figure 12 shows an example of this procedure for one of the ZIP codes in Northridge, 91325. In this case, the ZIP code is 96% type D soil, so type D soil is assumed.

Numerical calculations based on integration of the annual PDF for loss can answer questions such as What is the annual probability of loss greater than a particular value? and What is the expected annual loss given a particular deductible? The annual probability for losses greater than loss fractions of  $10^{-6}$ , 0.1, and 0.2 are shown in Table 4. For example, the annual probabilities of suffering losses greater than these three loss fractions are about 4.5%, 1.2%, and 0.8%, respectively. The recurrence times are roughly the reciprocal of these values. Thus a loss of 10% or larger is expected about every 83



**Figure 12.** Example from ZIP code 93125, Northridge, CA, shows the hazard curve, the annual probability density function for ground motion corrected for site condition, and annual probability density function for loss.

years, and a loss greater than 20% is expected about every 133 years. Similarly, Table 4 shows the expected annual losses above a range of deductibles.

It should be remembered that these estimates pertain to a loss distribution representing the entire ZIP code and would not necessarily apply to any one structure. They could be thought of as pertaining to a kind of “average” structure in the ZIP code that represents a weighted average of all the single-family homes in the ZIP code with average site conditions.

### DISCUSSION AND FUTURE WORK

The results presented above raise several questions that require further investigation. First, what are the relationships between the CPDF’s for loss found here and other pro-

**Table 4.** Examples of annual probabilities for loss and expected annual loss assuming various deductibles for ZIP code 91325 assuming type D soil condition. The integrals for the PDF for loss were truncated at 2.13 g, the maximum value given in the hazard curves calculated for the National Map. Sensitivity tests indicate that the PDF is not sensitive to the details of the hazard curve below a few % g.

Annual probability of loss greater than a particular loss fraction		Expected annual loss fraction above specified deductible	
Loss fraction	Annual probability of loss greater than loss fraction	Deductible	Loss fraction
$10^{-6}$	0.0449 (4.49%)	0	0.00464 (0.46%)
0.1 (10%)	0.0119 (1.19%)	0.1 (10%)	0.00413 (0.41%)
0.2 (20%)	0.0075 (0.75%)	0.2 (20%)	0.00350 (0.35%)

posed methods for estimating damage and loss, e.g., HAZUS (e.g., Kircher et al. 1997a, b), *ATC-13* (ATC 1985), and red-tagging (Boatwright et al. 2001)? Second, the CPDF's for loss found here are only strictly applicable to single-family housing in the Los Angeles-San Fernando Valley region. How should they be modified for the building types and practices in other regions exposed to earthquake hazard (e.g., northern California, Pacific Northwest, Intermountain states, central and eastern United States)? One would expect, for example, that for regions in which masonry is more common than in southern California, losses might display a different behavior. Third, is it possible to develop a better method of estimating ground motion throughout the region, perhaps including better estimates of site conditions that would reduce the scatter in the results? Fourth, if a more precise estimate of the location of each property was available, could scatter be reduced, or is the inherent variability in ground motion such that more precise locations would not improve the results significantly? Fifth, there seem to be some subtle relations between the value and age of the homes, and fractional loss. Can significant relationships along these lines be extracted from the data set? Can some of these variations be confidently ascribed to differences in construction design and practice? Sixth, clearly the inter-event variability in ground motions is extremely important in estimating losses. How can our understanding of inter-event variability be improved?

### CONCLUSIONS

Empirical conditional probability density functions for loss given ground motion have been obtained from data on insurance losses from the 1994 Northridge, California, earthquake. Predictions from the eight ground-motion parameters, except for the 1 sec and 3 sec spectral accelerations, predict losses within about 10% of the actual. The best predictions are for PGA, instrumental intensity, and 0.3 sec spectral acceleration. Comparison of losses predicted from a deterministic model of the earthquake with the actual is not as satisfactory, however, the agreement is good if source ground-motion values about 1.3 times the median values (that is, about one and one-half inter-event standard deviations above the median) are assumed. Nonetheless, the results presented here provide support for optimism that conditional loss curves based on ground motion are possible and practical, and that such curves can provide the basis for estimating losses from scenario earthquakes and annualized losses from hazard curves as shown here by example.

### ACKNOWLEDGMENTS

The authors are indebted to Ken Rukstales for continuing help with geographic information systems (GIS) analysis; to Erdal Safak for helpful discussions during the early phases of the work; to Dave Wald, Bill Graf, and three anonymous reviewers for very thoughtful and constructive reviews; and to Art Frankel for continuing discussion, support, and encouragement. The authors also express their appreciation to the California Geological Survey for access to the GIS version of the map of Wills et al. (2000).

### REFERENCES

- Algermissen, S. T., and Perkins, D. M., 1976. A probabilistic estimate of maximum acceleration in rock in the contiguous United States, *U.S. Geol. Surv. Open-File Rep.* 76-416.

- Applied Technology Council, 1985. *ATC-13, Earthquake Damage Evaluation Data for California*, Redwood City, CA.
- Blodgett, J. G., 2002. The ZIP code resources Web page, University of Missouri Outreach and Extension, Office of Social and Economic Data Analysis, <http://www.oseda.missouri.edu/jgb/ZIP.resources.html>.
- Boatwright, J. L., Thywissen, K., and Seekins, L. C., 2001. Correlation of ground motion and intensity for the 17 January 1994 Northridge, California, earthquake, *Bull. Seismol. Soc. Am.* **91**, 739–752.
- Boore, D. M., Joyner, W. B., and Fumal, T. E., 1997. Equations for estimating horizontal response spectra and peak accelerations from western North American earthquakes: A summary of recent work, *Seismol. Res. Lett.* **68**, 128–153.
- Building Seismic Safety Council (BSSC), 1995. *NEHRP Recommended Provisions for Seismic Regulations for New Buildings*, 1994 edition, Part 1—Provisions, developed for the Federal Emergency Management Agency, *FEMA-222A*, Washington, D.C., p. 8.
- Building Seismic Safety Council (BSSC), 2001. *NEHRP Recommended Provisions for Seismic Regulations for New Buildings and Other Structures*, 2000 edition, Part 1—Provisions, developed for Federal Emergency Management Agency, *FEMA-368*, Washington, D.C., p. 43.
- Cao, T., Petersen, M. D., Cramer, C. H., Topozada, T. R., Reichle, M. S., and Davis, J., 1999. The calculation of expected loss using probabilistic seismic hazard, *Bull. Seismol. Soc. Am.* **89**, 867–876.
- Efron, B., and Tibshirani, R. J., 1993. *An Introduction to the Bootstrap*, Chapman and Hall, New York.
- Evans, M., Hastings, N., and Peacock, B., 2000. *Statistical Distributions*, 3rd Edition, John Wiley & Sons, Inc., New York.
- Frankel, A., Mueller, C., Barnhard, T., Perkins, D., Leyendecker, E. V., Dickman, N., Hansen, S., and Hopper, M., 1996. National seismic-hazard maps: Documentation, *U.S. Geol. Surv. Open-File Rep. 96-352*.
- Kircher, C. A., Nasser, A. A., Kustu, O., and Holmes, W. T., 1997a. Development of building damage functions for earthquake loss estimation, *Earthquake Spectra* **13**, 663–682.
- Kircher, C. A., Reitherman, R. K., Whitman, R. V., and Arnold, C., 1997b. Estimation of earthquake losses to buildings, *Earthquake Spectra* **13**, 703–720.
- Petersen, M. D., Bryant, W. A., Cramer, C. H., Cao, T., Reichle, M., Frankel, A. D., Lienkaemper, J. J., McCrory, P. A., and Schwartz, D. P., 1996. Probabilistic seismic hazard assessment for the state of California, *California Division of Mines and Geology Open-File Report 96-08*.
- Steinbrugge, K. V., and Algermissen, S. T., 1990. Earthquake losses to single-family dwellings: California experience, *U.S. Geol. Surv. Bulletin 1939-A*.
- Steinbrugge, K. V., and Roth, R. J., Jr., 1994. Dwelling and mobile home monetary losses due to the 1989 Loma Prieta, California, earthquake with an emphasis on loss estimation, *U.S. Geol. Surv. Bulletin 1939-B*.
- Wald, D. J., Quitoriano, J. V., Heaton, T. H., and Kanamori, H., 1999a. Relationships between peak ground acceleration, peak ground velocity and modified Mercalli intensity in California, *Earthquake Spectra* **15**, 557–564.
- Wald, D. J., Vincent, Q., Heaton, T. H., Kanamori, H., Scrivner, C. W., and Worden, C. B., 1999b. TriNet “ShakeMaps”: Rapid generation of instrumental ground motion and intensity maps for earthquakes in Southern California, *Earthquake Spectra* **15**, 537–556.

- Wesson, R. L., and Perkins, D. M., 2001. Spatial correlation of probabilistic ground motion and loss, *Bull. Seismol. Soc. Am.* **91**, 1498–1515.
- Wills, C. J., Petersen, M. D., Bryant, W. A., Reichle, M. S., Saucedo, G. J., Tan, S. S., Taylor, G. C., and Treiman, J. A., 2000. A site-conditions map for California based on geology and shear wave velocity, *Bull. Seismol. Soc. Am.* **90**, S187–S208.

(Received 28 January 2003; accepted 2 February 2004)