Kotha, S., Bazzurro, P., Pagani, M. (2018): Effects of Epistemic Uncertainty in Seismic Hazard Estimates on Building Portfolio Losses. - *Earthquake Spectra*, *34*, 1, pp. 217—236.

DOI: http://doi.org/10.1193/020515EQS020M

Effects of Epistemic Uncertainty in Seismic Hazard Estimates on Building Portfolio Losses

Sreeram Reddy Kotha,[a),b)] M.EERI, **Paolo Bazzurro,**[c)] M.EERI, and **Marco Pagani**[d)]

In catastrophe risk modeling, a defensible estimation of impact severity and its likelihood of occurrence to a portfolio of assets can only be made through a rigorous treatment of uncertainty and the consideration of multiple alternative models. This approach, however, requires repeating lengthy calculations multiple times. To limit the demand on computational time and resources, a frequent practice in the industry is to estimate the distribution of earthquake-induced portfolio losses using a simulated catalog of events from a single representative *mean* ground motion hazard model for the region. This simplified approach is faster but may provide biased estimates of the likelihood of occurrence of the large and infrequent losses that drive many risk mitigation decisions. Investigation through case studies of different portfolios of assets located in the San Francisco Bay Region shows the potential for both a bias in the mean loss estimates and an underestimation of their central 70% interpercentile. We propose a simplified and computationally practical approach that reduces the bias in the mean portfolio loss estimates. This approach does not improve the estimate of the interpercentile range, however, a quantity of no direct practical use. [DOI: 10.1193/020515EQS020M]

INTRODUCTION

A robust, probabilistic quantification of the impact of future seismic events in terms of loss of life or economic property losses is crucial in mitigating earthquake risk. Risk assessment is of interest to various stakeholders, such as insurers, reinsurers, insurance brokers, capital market investors, corporations, engineers, and governmental institutions at both local and national levels. Most applications of interest outside of the academia are not limited to specific buildings but rather involve assessing the risk of portfolios of either private or public assets. Each of the hazard, exposure, and vulnerability components involved in any portfolio risk assessment study features modeling uncertainties that pervasively affect the resulting loss distributions. We only focus here on the effects of disregarding the modeling of seismic hazard uncertainty and on loss estimates for portfolio of buildings rather than for single structures.

a) GFZ German Research Centre for Geosciences, Telegrafenberg, 14473 Potsdam, Germany (fornerly at Understanding and Managing Extremes (UME) Graduate School, Pavia, Italy)
b) University of Potsdam, Institute of Earth- and Environmental Sciences, Potsdam, Germany
c) University School for Advanced Studies IUSS Pavia, Pavia, Italy
d) GEM Foundation, Pavia, Italy

Earthquake Spectra, Volume 34, No. 1, pages 217–236, February 2018; © 2018, Earthquake Engineering Research Institute

Although alternative approaches have been proposed (e.g., Ordaz and Arroyo 2016), in the more advanced Probabilistic Seismic Hazard Assessment (PSHA) studies, the epistemic uncertainty afffecting hazard input components is usually accounted for via logic trees (e.g., McGuire et al. 2005, Bommer and Scherbaum 2008). The hazard is computed for all the earthquake ruptures admitted by each branch of the logic tree and the mean hazard and its quantiles at any given site are extracted from the family of seismic hazard curves obtained. The weight assigned to each hazard curve is equal to the degree of belief assigned to each branch. At each site, the resulting mean hazard curve represents an average view of the hazard, which does not explicitly convey, by design, any specific information about lower and higher more extreme hazard results computed using some specific branches of the logic tree.

Knowing the mean hazard curve at all sites where the portfolio assets are located is sufficient to estimate the average annual loss (AAL) and the mean annual rate of exceedance of losses of different severity [whose ensemble of frequency-loss pair values is traditionally but improperly called the loss Exceedance Probability (EP) curve] for each single asset in the portfolio *considered as a separate entity*. However, given that, in general, multiple assets in the same portfolio can experience the same earthquake, knowing the mean hazard curves at all the sites where the portfolio assets are located is *not* sufficient to estimate AAL and loss EP curve for the entire portfolio. It is stressed here that in the large majority of risk assement studies carried out in practice, portfolio AAL and loss EP curve are indeed the critical end results of the analysis. The AAL and EP curve for any single asset are indeed used for insurance pricing, for example, but they are not valued for portfolio risk management purposes, a topic that is the core of this paper.

What is needed to estimate the portfolio AAL and loss EP curve is the distribution of the portfolio loss for any given earthquake rupture that is considered in any given branch of the logic tree-based hazard model. Hence for any such earthquake rupture, it is necessary to probabilistically model in the nearby affected region the random fields of the ground motion intensity measures (IMs; e.g., PGA or spectral acceleration [SA]) used for vulnerability computations. Again, the more complicated random field approach is necessary because the final objective is to evaluate, for any given earthquake, the portfolio loss that is caused by the ensemble of all the IMs at all the asset locations impacted by the earthquake. Of course, only the probabilistic distribution of each IM generated at each site by each earthquake is known. In other words, these IMs are random variables, which, as an added source of complication, are also statistically dependent. Hence their modeling is mathematically complex (e.g., Park et al. 2007). Although implicit in the previous discussion, it is worth emphasizing that the random field approach preserves the hazard at any single site that would be computed using the traditional PSHA approach. The two methods are fully consistent, but the former is made necessary by the portfolio nature of the problem at hand.

In catastrophe risk modeling, it is customary to assemble the totality of earthquake ruptures considered in the portfolio loss estimation into a so-called stochastic catalog of potential future events. As previously stated, to preserve the uncertainty in the frequency of occurrence and size of future earthquakes and in ground motion, the simulation of such catalog should be completed for every branch of the logic tree. However, routine practice in catastrophe risk modeling where speed of execution is of the essence does not allow for use of multiple stochastic catalogs of events. On the contrary, practicality dictates finding an expeditious procedure that preserves the robustness of the resulting mean loss estimates as much as possible. In a nutshell, this statement translates into using a "carefully crafted" single stochastic catalog of earthquake ruptures in place of a suite comprising as many catalogs as there are branches in the logic tree.

There is a quite large spectrum of procedures used in practice for the purpose of assembling a suitable single catalog of events but, in general, they can be summarized into two groups: those that attempt to obtain a catalog that is somewhat consistent with the mean hazard in the region, and those that attempt to obtain a catalog that preserves, again in an average sense, the mean losses caused to a subjectively chosen portfolio of assets that is representative of the building inventory located in the region of interest.

In the latter group of procedures, the target portfolio of assets utilized to fine-tune the procedure and identify the final stochastic catalog of simulated events is usually an aggregate representation of the entire regional asset exposure. Although this approach is intuitively appealing, the final stochastic catalog of events obtained this way is then used to evaluate losses to *any* portfolio of assets, including those that are not aggregated or those that are spatially different or who have different physical characteristics than the target portfolio. Since the accuracy of the loss estimates obtained by this latter group of methods is portfolio-dependent and cannot be investigated in a systematic manner, it will not be investigated here. We only offer a word of caution about using such a practice. The accuracy of final losses should be questioned, especially for portfolios that systematically differ from the target portfolio, either because they are concentrated in specific areas, or because made of assets that are rather homogeneous or with a distribution of vulnerability classes different than those in the target portfolio. The accuracy should also be scrutinized when considering the losses of subsets of large portfolios that may have similar characteristics to those of the target portfolio. In all these cases, the main hypotheses that led to the selection of the stochastic catalog are not valid and there is no guarantee that the loss estimates obtained are close to the desired mean estimates.

In this study, instead we concentrate our attention on the former group of procedures, which make use of a stochastic earthquake catalog that, in most practical applications, is representative of the "mean hazard" branch. Strictly speaking, the stochastic catalog for the mean hazard branch in the first aforementioned approach could only be constructed by collating all the ruptures of all the branches and by assigning its native annual frequency of occurrence times the weight assigned to that branch to each rupture in a given branch. In any practical application, this operation would result in an enormous and, therefore, unmanageable set of ruptures. The total number of ruptures would be identical to the sum of all the ruptures in all the branches—a route that would lead to no computational gain.

Therefore, in most practical applications, the consistency with the mean hazard everywhere in the region is only loosely enforced by considering the stochastic catalog of the specific branch that, at key locations, provides hazard estimates that are close to the mean estimates. Although not widely recognized, this simplified method suppresses the epistemic uncertainty in the seismic hazard representation by design, rather than accounting for it in the loss calculations.

We attempt here to quantify the impact on the accuracy of portfolio losses caused by the suppression of the hazard uncertainty when using a mean hazard-based stochastic catalog of simulated earthquakes selected according to a single branch of the logic tree. To do so, first we carry out a set of loss calculations based on multiple stochastic catalogs derived from all the branches of the hazard logic tree. This set of analyses, which is the most robust way of treating hazard uncertainty, results in loss estimates that constitute our benchmark against which all other loss estimates obtained using only a single-branch–based stochastic catalog are compared and contrasted. The measure of consistency between branch-specific hazard estimates and mean hazard estimates will be discussed in detail later. Finally, we will propose an alternative procedure that leads to more accurate mean loss estimates than the mentioned procedure usually adopted in practice.

MODELING METHODOLOGY AND CASE STUDIES

To assess whether the current procedures based on a single-branch stochastic catalog are indeed acceptable, we pragmatically examine the influence that the inclusion or exclusion of the hazard epistemic uncertainty (i.e., multiple alternative stochastic catalogs of simulated earthquakes) has on the resulting mean and quantile loss estimates for a variety of portfolios located in the San Francisco Bay Region (SFBR). To provide results that are representative of a range of cases of practical relevance, we consider portfolios with different numbers and asset density. However, to limit the computational effort to a manageable amount (recall that multiple stochastic catalogs need to be generated and utilized), we consider portfolios with a number of assets that are smaller than that of most applications and a simplified hazard model that includes only those faults that contribute more than 90% of the overall hazard (according to the United States Geological Survey [USGS] 2008 online Interactive Deaggregation tool) in the region, and only two ground motion prediction equations (GMPEs). Despite these simplifications, the final results certainly do not suffer from a loss of generality. Although significantly more complex models exist for the seismicity of SFBR (e.g., Petersen et al. 2014), the adopted hazard characterization is more complex than that of most PSHA studies in other regions of the world.

SFBR HAZARD MODEL

As alluded to previously, the seismic hazard model adopted here is a reduced version of the 2008 USGS model (see Petersen et al. 2008 and related appendices), which, in turn, is an adaptation of the Uniform California Earthquake Rupture Forecast, Version 2 (UCERF2) model by Field et al. (2009). For simplicity and manageability, the USGS model has been further trimmed to include only the seismicity produced by the San Andreas Fault, the Hayward–Rodgers Creek Fault, and the Calaveras Fault (Figure 1), which are the primary sources of events larger than M6.5 in the SFBR. By limiting the number of sources, we obtained a pruned version of the USGS 2008 source model logic tree that includes a total of eight branches for modeling the uncertainty in the earthquake occurrence, instead of the thousands of branches of the original logic tree. The epistemic uncertainty in the ground motion characterization was also reduced to considering only the BA08 (Boore and Atkinson 2008) and the CY08 (Chiou and Youngs 2008) equations with an equal weight of 0.5. The inclusion of these two GMPEs results in a logic tree with $8 \times 2 = 16$ hazard branches in all, each one with its own weight.

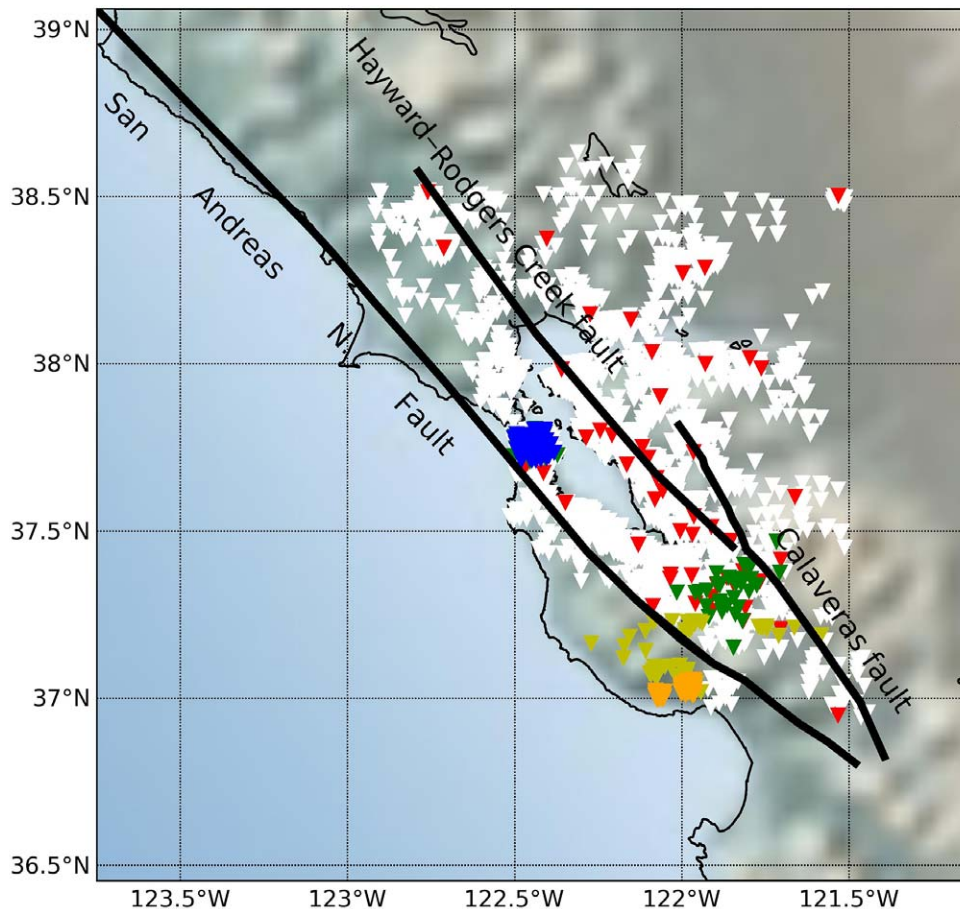

Figure 1. Fault system and the six portfolios of assets considered for seismic hazard and risk assessment in SFBR. Legend: #1 white, #2 red, #3 blue, #4 orange, #5 green, and #6 yellow.

Figure 2 compares the mean hazard curves obtained at a site in San Francisco (37.7750°N, 122.4183°W) with those extracted from the USGS online interactive Hazard Curve Application tool. The comparison of the two hazard curves for the 5%-damped S_a at 1.0 s oscillator period, $S_a(1.0\,\mathrm{s})$, is consistent with expectations, given the limitations of the simplified logic tree used. The two curves are very close for $S_a(1\,\mathrm{s}) \geq 0.25\,\mathrm{g}$, where most contributions to the hazard derive from the large magnitude events on the major faults that are included in the simplified model. At lower $S_a(1\,\mathrm{s})$ levels, on the other hand, our estimates of the annual frequencies of exceedance are lower due to the absence of less significant faults and by the exclusion of more frequent $\mathrm{M} < 6.5$ events from the background seismicity sources. In other words, the mean hazard is preserved for ground motion levels of engineering significance.

Therefore, this simplified, 16-branch hazard model is deemed adequate for the purpose of this study, and the benchmark loss calculations are performed using 16 stochastic catalogs of

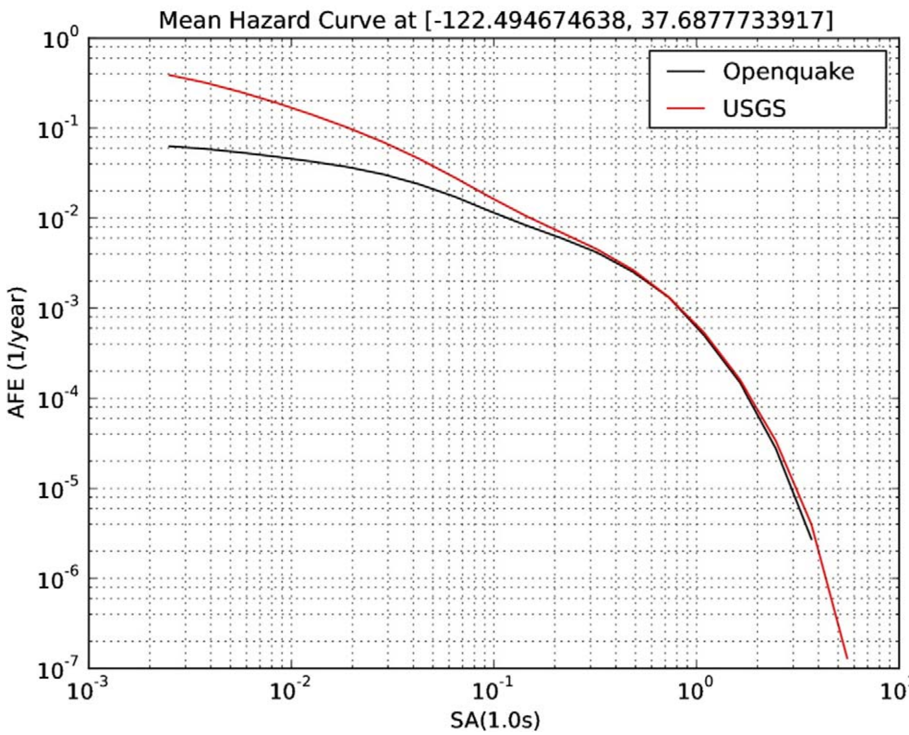

Figure 2. Comparison of the hazard curves for downtown San Francisco site obtained from the simplified model adopted for this study and from the USGS 2008 study. AFE stands for annual frequency of exceedance.

simulated events—one for each branch of this model. Each catalog is assembled assuming that each earthquake rupture occurs according to a Poissonian stochastic process with given annual rate of occurrence, and many realizations of one year of seismicity (here, 10,000 realizations) are simulated using a Monte-Carlo approach. The stochastic catalog and the loss calculations that followed were carried out using the software OpenQuake engine version 1.0.0, developed by the Global Earthquake Model (GEM) foundation (http://www.globalquakemodel.org/; see also Pagani et al. 2014, Silva et al. 2014).

EXPOSURE MODEL

HAZUS (Kircher et al. 2006) provides a detailed model of the built environment and demographics of the 19 counties in the SFBR. The building inventory is divided into several classes based on building type, height of structure, seismic design code, time of construction, occupancy level, functionality, retrofit level, etc. There are approximately 10 million inhabitants in the SFBR, but most live within 40 kilometers of the San Andreas Fault system. HAZUS demands such a complex exposure model, but that level of complexity is unnecessary for the scope of this study.

The exposure model considered here consists of a portfolio of 3,213 identical buildings and of five subsets of it (Figure 1) spatially distributed so as to reflect the population density in the SFBR. These six portfolios are composed of identical single-family, low-rise wood-framed structures constructed post-1950 ("W1−h−RES3AF−DF" in the HAZUS taxonomy). These buildings are the most common residential buildings in the SFBR. Each building has the same nominal replacement value of US $3 million and the same vulnerability function. Note that portfolios of buildings with identical vulnerability are common in the catastrophe risk modeling industry and they are routinely used in real applications. The details of these six portfolios follow:

Portfolio #1: 3,213 assets in the SFBR. It is the most numerous and widely spread portfolio in an area of 15,000 km^2. This is the portfolio whose results will be more carefully scrutinized because it is the most representative of those used in real applications.

Portfolio #2: 59 assets in the SFBR. It is the least dense portfolio in an area of 15,000 km^2. It is needed to investigate whether the conclusions made for Portfolio #1 are affected by the number and the density of the assets.

Portfolio #3: 66 assets in San Francisco. It is an exposure model of 66 identical buildings in 122 km^2 of area, concentrated only in San Francisco. It is needed to investigate whether the conclusions made for Portfolio #1 are affected by a specific hazard environment and by a more spatially concentrated portfolio with fewer assets.

Portfolio #4: 63 assets in Santa Cruz. It is the smallest and most dense (41 km^2) portfolio, far from San Francisco where Portfolio #3 is located. This portfolio was considered to further the investigation explained for Portfolio #3 but in a different hazard environment.

Portfolio #5: 59 assets in San Francisco and San Jose. This portfolio, whose area is 817 km^2, has intermediate characteristics between Portfolios #3 and #1 and has buildings in the two areas with the largest population in the SFBR.

Portfolio #6: 51 assets in Santa Cruz. This portfolio is less dense (600 km^2) than Portfolio #4. The results will help understand the effects of building concentration on the accuracy of loss estimates.

VULNERABILITY MODEL

The vulnerability function adopted here (Porter 2010) is shown in Figure 3. The ground motion IM used to link the mean damage factor (called MDF here and defined as the cost of repair normalized by the replacement cost of the building) to the level of shaking is $S_a(1.0\,\mathrm{s})$. The bands of increasing width with increasing $S_a(1.0\,\mathrm{s})$ represent the ± 1 standard deviation curves. The distribution of the MDF around the mean is assumed to be truncated lognormal.

LOSS ESTIMATION ANALYSES

We carried out two distinct sets of analyses using the OpenQuake-engine (Pagani et al. 2014): the fully probabilistic benchmark case with 16 stochastic catalogs and the simplified case based on the single branch of the 16 in the logic tree that provided the hazard estimates closest to the mean at key locations. To identify such a branch, the natural log of estimated annual frequency of exceedance of a range of $S_a(1.0\,\mathrm{s})$ values (0.2–2.0 g) from each branch

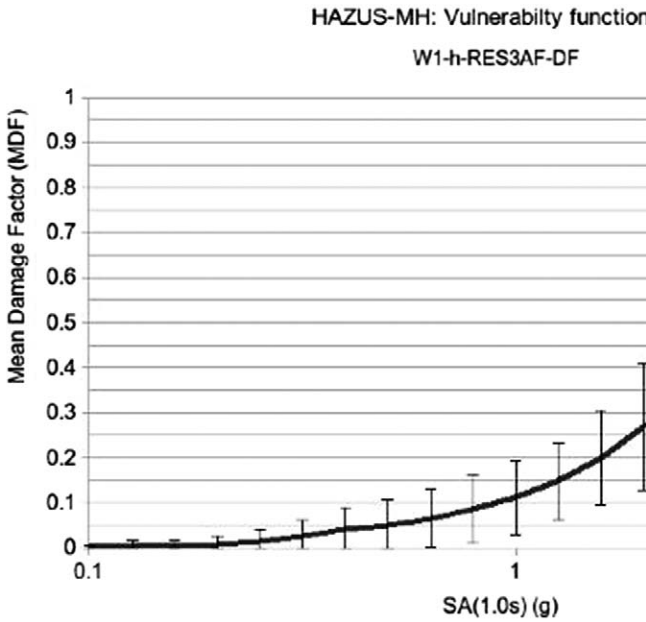

Figure 3. Vulnerability function used in the study showing the mean damage factor and the standard deviation (vertical error bars) at each ground motion level.

were compared to the natural log of mean annual frequency of exceedance values at 50 key sites (Figure 4) randomly chosen from the sites in Portfolio #1. Proportionally higher weight and, therefore, more importance was given to the locations with higher population density. The branch identified by the marker color in Figure 4 was found to be the closest to the mean at each site. The selected model (the dark green color in the circles displayed in Figure 4), namely, the "Segmented fault model – **H**anks and Bakun (2002) magnitude scaling relationship (MSR) – **A**-priori fault rates – **B**oore and Atkinson (2008) GMPE," abbreviated as AHB in this study, was the one that provided hazard estimates closer to the mean hazard at most sites, especially in San Francisco, San Jose, and Oakland, which are the three most populated cities in the region, but not everywhere. This is often the case in real applications. To test the sensitivity of the results, two other models corresponding to the second and third best branches are used in the simplified approach, as shown later.

In the fully probabilistic benchmark case, 10 stochastic catalogs simulating 10,000 realizations of yearly seismicity were developed for each one of the eight branches of the logic tree. The number of events in these 80 catalogs varies from 330 to 1,250, of which 200 to 440 ruptures are larger than M6.5. Another set of one hundred 10,000-year stochastic catalogs of the AHB branch were generated for computing the loss distributions corresponding to the mean hazard case. For each one of the events in any of these catalogs, OpenQuake was used to simulate one spatially correlated random field of $S_a(1.0\,\mathrm{s})$ modeled according to the work of Park et al. (2007) using the correlation model of Jayaram and Baker (2009). These random fields, which were generated using both the BA08 and the

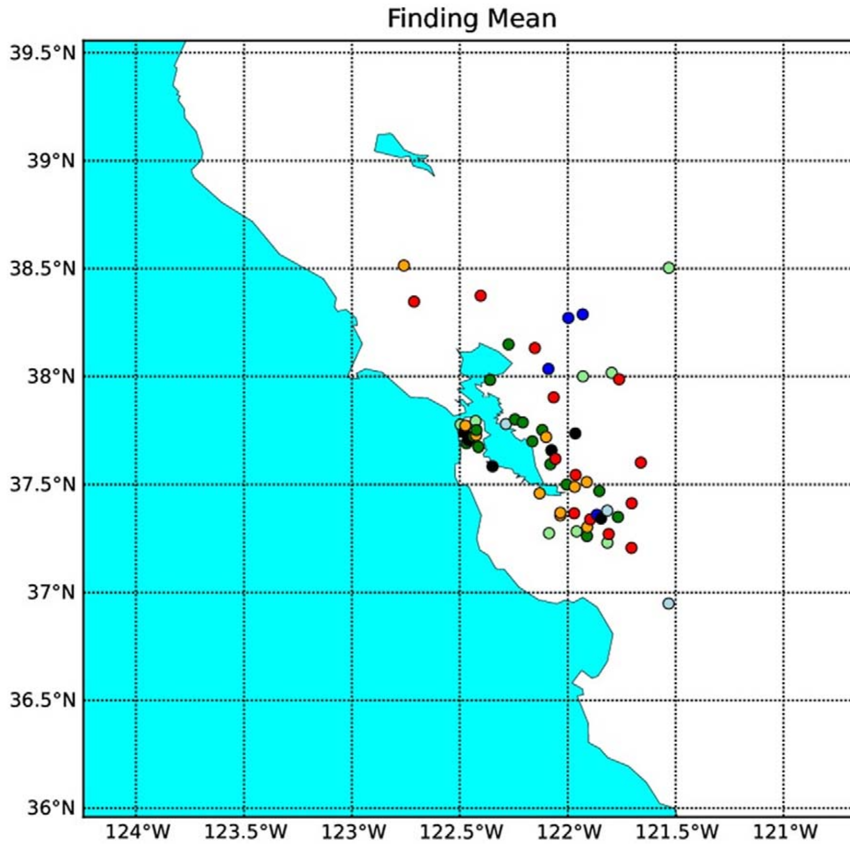

Figure 4. Location of the 50 sites in the SFBR used to identify the branch of the logic tree utilized in the simplified approach. The color within each circle represents the specific branch whose hazard estimate is closest to the mean hazard one at that location. The wide range of colors demonstrates the diversity of representative mean hazard models across the critical sites picked in the region.

CY08 GMPEs, were used as input for the loss estimation study as explained later. Therefore, in the fully probabilistic case, $80 \times 2 = 160$ stochastic catalogs of random fields were used.

Each random field was interpolated to extract the $S_a(1.0\,\mathrm{s})$ values at all the building sites, and these values were used to simulate the damage ratio from the damage function displayed in Figure 3. The damage factor for each building (say, 15% for a building experiencing $S_a[1.0\,\mathrm{s}]$ equal to 1.0 g) was then multiplied by the replacement cost of the building (here, US $3 million) to compute the loss in USD (i.e., US $450,000 in this example). The loss for a portfolio experiencing any of the simulated ground motion random field caused by any specific event is computed as the sum of the losses at all sites of that portfolio. This exercise is repeated for each one of the events in any of the 10,000-year catalogs. The losses generated by all events in a stochastic catalog are then ranked from the highest to the lowest (which may very well be zero). The highest loss is assigned an annual rate of exceedance of 1/10,000, the second highest a value of 2/10,000, the third highest a value of 3/10,000, and so on. This series of losses is then plotted versus the annual rate of exceedance (or versus its reciprocal, namely the mean return period) creating the loss EP curve introduced earlier.

In the fully probabilistic benchmark case, one loss simulation was performed for each one of the random fields generated for all the events contained in the 160 stochastic catalogs (80 catalogs of events × 2 GMPEs), leading to 160 loss EP curves for each one of the Portfolios #1–6. For the simplified, mean hazard-based case, one loss simulation was performed for each one of the 100 catalogs of ground motion random fields for the AHB branch, leading to 100 loss EP curves for each one of the Portfolios #1–6. Each of the 160 loss EP curves in the benchmark case is assigned 1/10th of the weight of its corresponding hazard branch. All the 100 loss EP curves in the mean hazard based case have all the same weight of 0.01. The mean and quantiles of these EP curves for each one of the portfolios are compared and contrasted in the next section.

COMPARISON OF LOSS ESTIMATES

The 160 EP curves obtained for the entire Portfolio #1 using the rigorous, fully probabilistic method discussed previously are shown in gray in Figure 5. This family of EP curves represents the uncertainty in the portfolio loss distribution stemming from the uncertainty in the seismicity model, the ground motion prediction model, and the severity of damage given the ground motion level at each site. To appreciate the effect on the final portfolio loss distribution of removing the uncertainty on both the earthquake occurrence model and on the ground motion model, these summary statistics curves are compared in

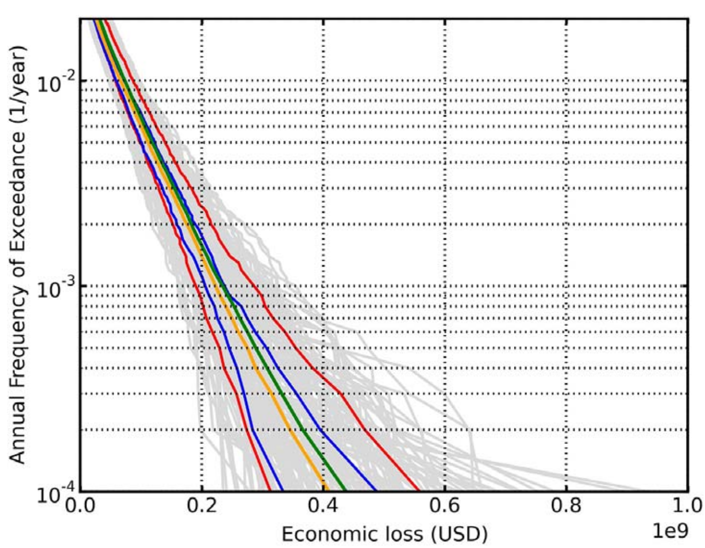

Figure 5. EP curves for Portfolio #1. The green (mean) and red (15th and 85th quantiles) curves correspond to benchmark calculation, and the yellow (mean) and blue (15th and 85th quantiles) curves correspond to the mean hazard-based simplified calculation carried on with the AHB branch.

Figure 5 with the curves corresponding to the simplified approach based on the AHB branch. It is clear that the two mean curves do not coincide and that central 70% interpercentile range delimited by the 15th and the 85th empirical quantile curves are much narrower in the simplified mean hazard-based case. The bias in the mean is even more evident in Figure 6, where the mean EP curves from the simplified case for Portfolios #1–6 are normalized by the corresponding mean EP curves from the benchmark case. For all portfolios, the simplified case provides mean EP curves that underestimate the benchmark cases by 10% or less for most annual rates, except for those of 10^{-2} or higher, which correspond to small and frequent losses where the bias reaches 20% or more, especially for Portfolio #4.

The bias in the mean EP loss curve is due to the fact that none of the branches, including the AHB branch used in Figure 6, represent the mean hazard perfectly well everywhere in the SFBR. The bias is also dependent on the branch selected to represent the mean hazard. The change in the bias trend is immediately apparent when inspecting Figures 7 and 8, where the AHB branch is replaced by the branches that provided the second best [MEC: **M**oment balanced rates – **E**llsworth (2003) MSR, **C**hiou and Young 2008] and the third best [MHB: **M**oment balanced rates – **H**anks and Bakun (2002) MSR, **B**oore and Atkinson 2008] match of the mean hazard at the 50 sites considered (Figure 4). The mean EP curve from the simplified mean hazard-based case does not always underestimate the mean EP curve from the benchmark case. The trend is significantly different than the one shown in Figure 6, implying that the branch chosen to produce the stochastic catalog used as input to the loss calculations controls shape and magnitude of the resulting bias in the mean loss estimates, which in turn depends on the choice of critical sites.

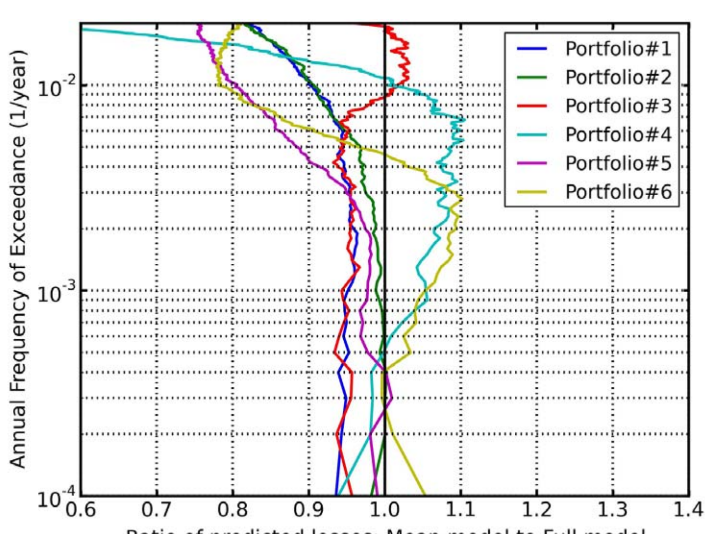

Figure 6. Ratio of the six mean EP loss curves for the benchmark case to the corresponding mean EP loss curves for the simplified mean hazard-based case that uses the AHB branch.

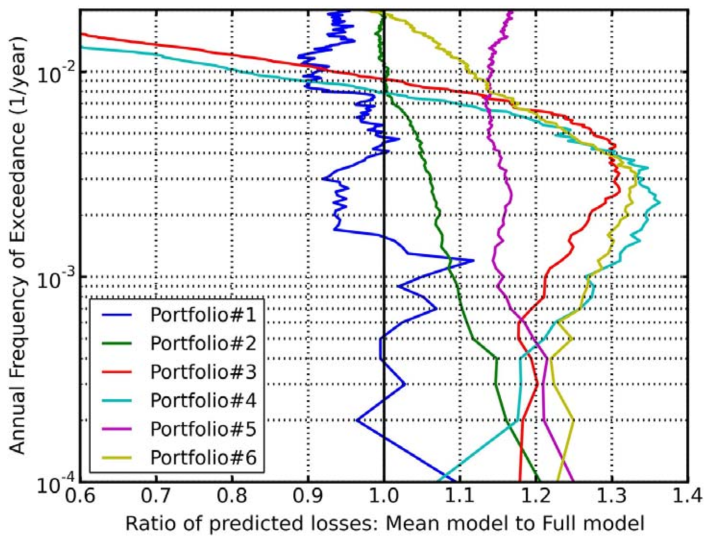

Figure 7. Ratio of the six mean EP loss curves for the benchmark case to the corresponding ones for the mean hazard-based simplified case that uses the MEC branch.

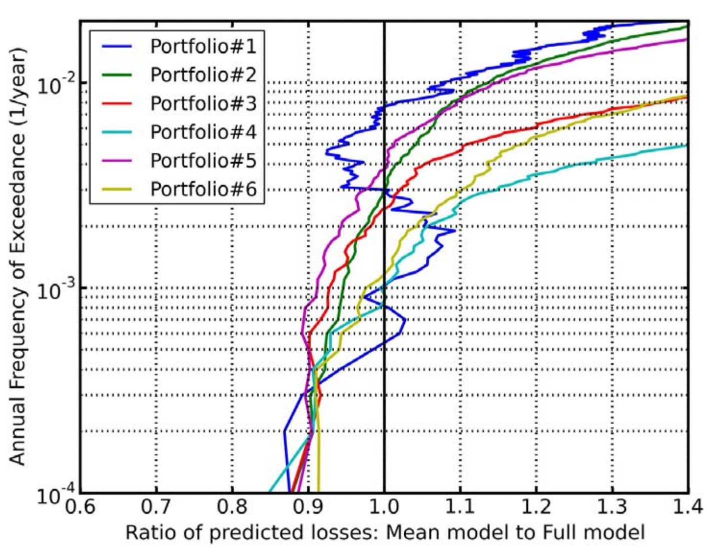

Figure 8. Ratio of the six mean EP loss curves for the benchmark case to the corresponding ones for the mean hazard-based simplified case that uses the MHB branch.

The width of the loss distribution generated by both the rigorous and the simplified approaches is evaluated by looking at the empirically derived central 70% interpercentile range (approximately bound by the ± 1 standard deviation curves in a Gaussian distribution) around the mean. To do so, we computed the ratio of the 15th and the 85th quantile curves from each distribution (i.e., the red curves for the benchmark case and the blue curves for the simplified case in Figure 5) to the mean loss curve from the same distribution (i.e., the green curve for the benchmark case and the yellow curve for the simplified case in Figure 5). These two interpercentile ranges for the true but unknown EP loss curve of Portfolio #1 are shown in Figure 9. The interpercentile range computed by neglecting the epistemic uncertainty in the hazard model is significantly narrower than the corresponding interpercentile range estimated in the benchmark case that accounts for that source of uncertainty as well. In this particular case, the 15th and 85th quantile curves predicted in the simplified case correspond approximately to the 30th and 70th quantile curves of the benchmark case. This means that the width of the central 70% interpercentile range of the simplified case is as narrow as the central 40% interpercentile range of the benchmark case. Moreover, recall that, besides being narrower, the interpercentile range in the simplified case bounds a mean loss EP curve that is biased by an amount that may be either positive or negative depending on which branch of the seismic model logic tree was selected to produce the underlying stochastic catalog of earthquake footprints. Although omitted here, the comparison of central 70% interpercentile ranges for Portfolios #2–6 shows a similar trend, but with a larger discrepancy for smaller portfolios, as shown in Figure 10 for Portfolio #3.

Another portfolio loss summary statistic that is widely used is the AAL. In general, there are many years where there are no losses and some years where there are large losses when earquakes occur. The AAL represents the expected loss per year when computed over a long period of time (or, better, across many realizations of next year earthquake activity). The AAL is computed here by summing up all the portfolio losses caused by all the earthquakes in any 10,000-year stochastic catalog and dividing them by 10,000.

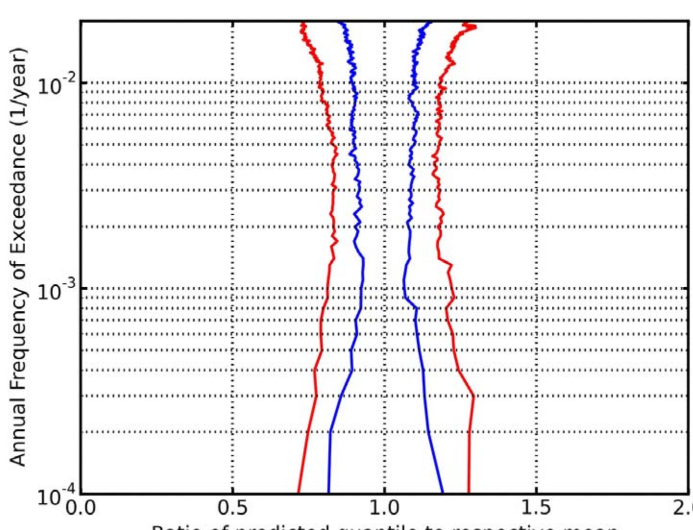

Figure 9. Empirical central 70% interpercentile range of loss estimates for Portfolio #1. Legend: red, benchmark case; blue, simplified case.

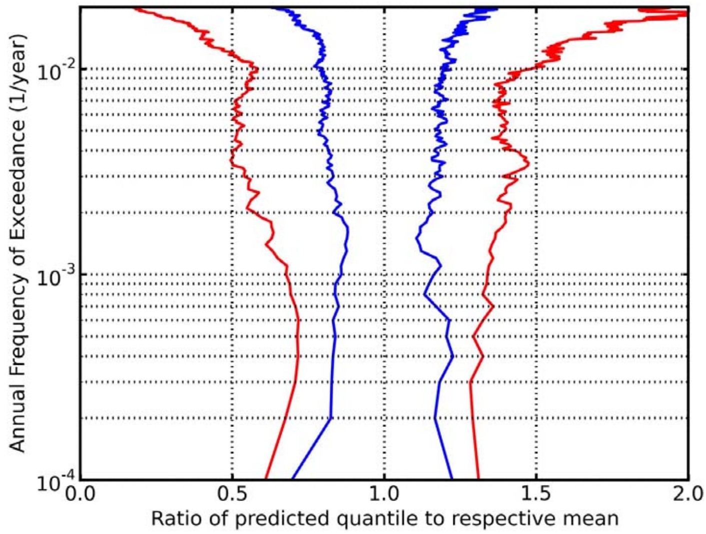

Figure 10. Empirical central 70% interpercentile range of loss estimates for Portfolio #3. Legend: red, benchmark case; blue, simplified case.

To be able to compare AAL results across the six portfolios, we normalized them by the total replacement value of the portfolios. The mean, the 15th, and the 85th quantiles of the normalized AALs (multiplied by 1,000 only for plotting purposes) from both the benchmark case and the primary mean hazard-based simplified case (AHB branch) are compared in Figure 11. The two adjacent subcolumns for each portfolio represent the loss estimates with (benchmark case, right column) and without (simplified case, left column) hazard

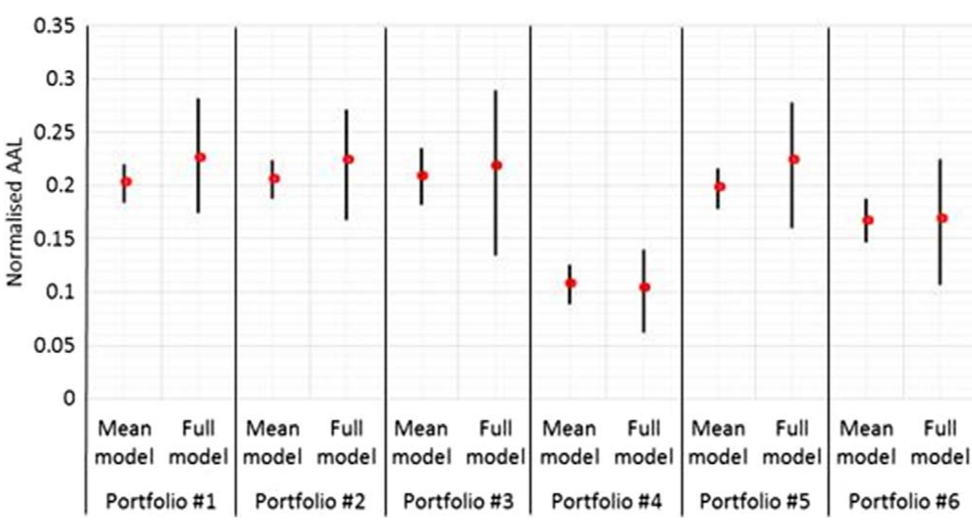

Figure 11. Normalized AAL values (scaled by 1,000 times) for the six portfolios.

uncertainties. Markers (red dots) referring to the mean AAL values for the same portfolio that are not horizontally aligned imply that the AAL estimate from the simplified approach is biased either positively or negatively. If the vertical black bar connecting the 15th and the 85th quantiles in the simplified case is shorter than that in the benchmark case, then the interpercentile range of the AAL estimate from the simplified approach is underestimated.

The AALs of Portfolios #3, #4, and #6 exhibit a much smaller, almost negligible, bias compared to those of Portfolios #1, #2, and #5. The AHB branch represents the mean hazard much better for the former set of three portfolios than it does for the latter set. For example, at most sites of the San Francisco Portfolio #3 (blue assets in Figure 1), the branch that estimates more closely the mean hazard is indeed branch AHB (green circles in Figure 4). The opposite holds true for the San Jose Portfolio #5, for which the AHB branch does not approximate the mean hazard well. Note that the overestimation or underestimation of the AAL for the AHB-based simplified cases could have been anticipated by inspecting Figure 6. The ratio of EP curves is always below one for Portfolios #1, #2, and #5 (this leads to AAL estimates that are too low), but it is above and below one for Portfolios #3, #4, and #6 for which errors compensate.

A different bias trend would be observed if the MEC (Figure 7) or MHB (Figure 8) branches were used to support the AAL calculations for the simplified case. Again, Figures 7 and 8 can provide indications about the sign of AAL bias for different portfolios. For example, the MEC-based simplified case overestimates the AAL for Portfolio #5 (purple line), but provides a rather unbiased AAL estimate for Portfolio #3 (red line) due to error compensation. However, in all cases, the central 70% interpercentile range of the AAL is also underestimated when the MEC and MHB branches are utilized to support the simplified approach computations (results not shown).

ALTERNATIVE SIMPLIFIED APPROACH

It is clear that the simplified method based on the identification of the single branch providing hazard estimates closer to the mean hazard at key locations is unreliable because, in general, it provides potentially inaccurate loss estimates. However, as stated earlier, the fully probabilistic approach is beyond the reach of most real-life applications with portfolios of hundreds of thousands of assets. We propose an improved simplified approach that assembles a single compendium catalog of simulated ruptures belonging to all the earthquake source models (here, 8) considered in the logic tree by exploiting the concept of a filtered Poissonian process (Parzen 1999). As mentioned earlier, before Monte-Carlo simulation is applied, the i^{th} rupture stemming from the j^{th} source model has an annual occurrence rate of $\lambda_{i,j}$ (say, 0.001). If the branch j is assigned a weight, w_j (say, 0.1), then the modified annual occurrence rate of that rupture would be simply $w_j\lambda_{i,j}$ (i.e., 0.0001). A single stochastic catalog of ruptures of any length of time, T_L (here, 10,000 years), can be generated from this compendium list of ruptures, by simulating, using a Monte-Carlo technique, the occurrence of each single rupture within T_L years as a filtered Poissonian process with modified annual rate of occurrence as previously specified. This means that catalogs that stem from a more credible branch will contribute, on average, more earthquakes to the compendium catalog than catalogs that have lower credibility. It also means that within each original catalog, the rarer events are less likely to be picked and included in the compendium catalog than the more frequent events. The resulting compendium catalog has approximately the same number of events as the weighted average number of events in the original single-branch catalogs, namely 570 events here.

This improved method is practical because (a) the simulation of the compendium stochastic catalog of ruptures is conceptually simple and computationally light and (b) it delivers a single catalog comprising a number of ruptures similar to that of any of the single-branch stochastic catalogs used in the simplified approach discussed in the previous section. The resulting compendium catalog of ruptures is then coupled with the ground motion models (two in this study) to generate a compendium catalog of spatially correlated random fields of IMs. It is worth noting that that some versions of this alternative simplified approach are currently in use in the catastrophe risk modeling industry but the details are not available in the public domain.

To test the robustness of this alternative simplified method, we again use the benchmark loss EP curves (Figure 5) for the six portfolios from the rigorous approach. To ascertain the potential bias and the uncertainty in the loss estimates computed from the compendium catalog, we sampled 100 realizations of it and used them to estimate portfolio losses. The resulting 100 loss EP curves for Portfolio #1 are shown in Figure 12, while the values of the AAL range from US \$2.25 billion to US \$2.92 billion. The potential bias (Figure 13) is assessed again by plotting the ratio of the *mean* EP curve from the 100 compendium catalogs to the *mean* curve obtained using the benchmark approach.

The bias at all levels of losses in the EP curve is within ±10% for all portfolios with the exception of small, frequent losses for the Santa Cruz Portfolio #4 that are controlled by few, localized earthquakes. The bias deriving from single-branch catalogs was significantly higher as apparent from Figures 6 to 8. Note that higher bias should be expected (gray lines in Figure 14) when only one realization of a 10,000-year catalog is used (as often done in

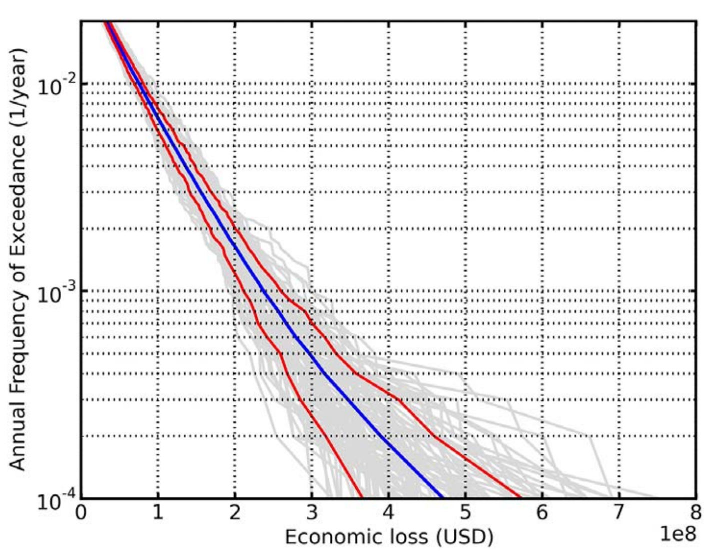

Figure 12. 100 EP curves for Portfolio #1 from 100 simulated compendium catalogs. Mean (blue), 15th and 85th quantiles (red) are shown.

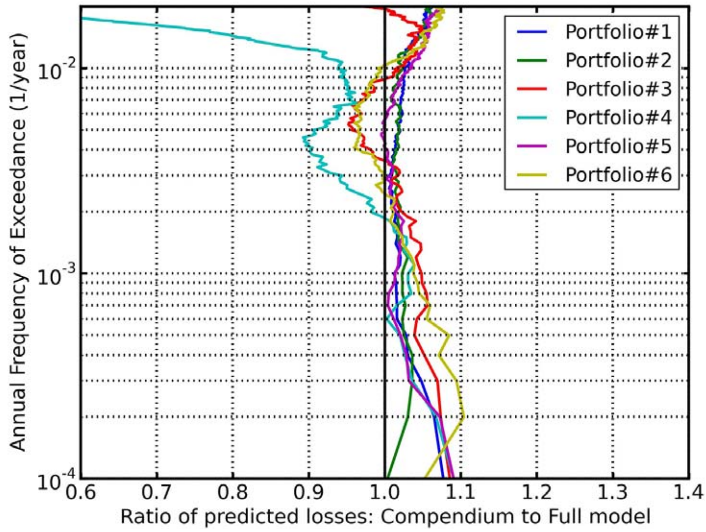

Figure 13. Ratio of the six mean EP loss curves for the benchmark case to the corresponding ones obtained for the 100 compendium catalogs.

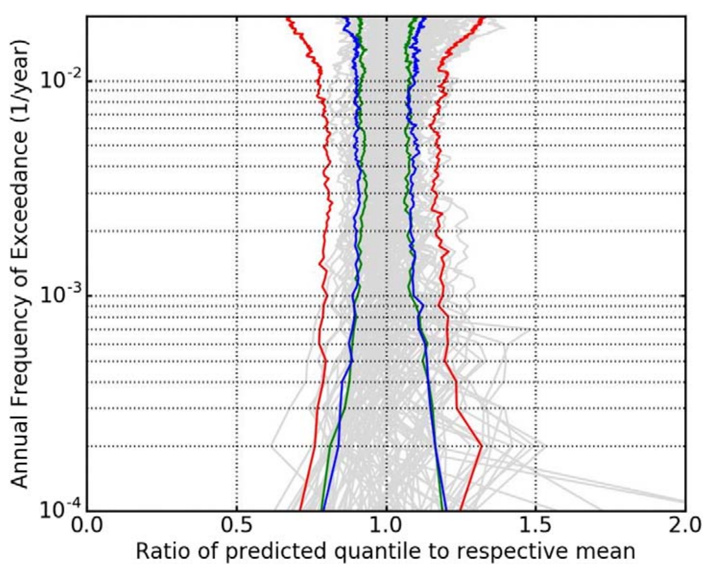

Figure 14. Three sets of central 70% interpercentile range on mean EP loss curve predicted by the fully-probabilistic model (red), the simplified branch-based approach that uses the AHB model (blue), the alternative simplified approach based on 100 compendium catalogs (green), and the ratio of each of the one hundred 10,000 year-long compendium loss catalogs to the mean of the fully-probabilistic model (gray).

practice) rather than the mean of many realizations, as done here. Furthermore, the central 70% interpercentile range around the mean (green lines in Figure 14) remains smaller than the interpercentile range exhibited by the benchmark method (red lines in Figure 14). The losses from the compendium catalog show lower variability than those stemming from the various hazard logic tree branches that fully capture the hazard epistemic uncertainty.

To summarize, the improved alternative simplified approach seems to remove, at least in the cases considered here, some of the bias and, more importantly, the vagaries of the results stemming from the selection of the branch providing hazard estimates closer to the mean. The full epistemic uncertainty in the loss distribution which is not important, however, for any real applications known to the authors, is not recovered.

CONCLUSIONS

Several applications in catastrophe risk modeling involve earthquake loss estimation for portfolios of assets, many of which deal with assessing insurance losses. For very compelling pragmatic reasons, all the earthquake risk assessment models that are used in this industry need to balance sophistication of calculations with speed of execution. In this trade-off, it is customary that rigorous approaches are, by design, set aside in lieu of approximate solutions that favor the expediency of calculations. The developers of these models usually investigate the consequences on the final loss estimates caused by the adoption of approximate solutions that are based on simplified approaches, but their findings often go unpublished. Hence the end users of catastrophe risk models that were developed using such simplified approaches are unaware of the potential consequences caused by the technicalities adopted during their development and, therefore, cannot take them into consideration during the decision-making process.

In this study we took advantage of the earthquake risk assessment OpenQuake engine tool developed by the GEM foundation to systematically assess the accuracy of portfolio loss estimates caused by adopting simplified, mean hazard-based approaches in the development of the stochastic catalog of future events used for loss estimation. Different variations of this simplified approach are present in all current catastrophe risk model used for practical applications. Note that we are not concerned here with the important but simpler problem of assessing the accuracy of single asset earthquake losses.

In the large majority of high-quality seismic hazard studies, the imperfect knowledge in modeling both earthquake occurrence and ground motion is explicitly accounted for by considering alternative models and weighting them in a logic tree framework. In portfolio loss estimation, this imperfect information should result in a plethora of different stochastic catalogs of future events—one per alternative seismicity/ground motion model—to be used for computing the final portfolio losses and their statistics. However, to ensure speed of execution, for portfolio loss calculations, it is customary to use only a single stochastic catalog of simulated earthquakes that contains a manageable number of events comparable to that of catalogs based on a single branch of the locic tree. Ideally, one would like to use a catalog of events that provide the mean hazard everywhere. This approach is possible but does not reduce the amount of computations needed in the rigorous approach consisting of as many catalogs as there are branches in the logic tree. In practice, this catalog is often generated using the single seismicity/ground motion model branch that provides the closest hazard estimates to the mean hazard at *key locations*. Using six portfolios of buildings located in the SFBR, we showed that this mean hazard-based, branch-specific simplified approach is bound to produce portfolio loss estimates that are biased. The sign of the bias (i.e., positive for overestimation and negative for underestimation) and the amount of the bias, which can be significant at times (e.g., 40% or more), depends on how well the seelcted hazard branch estimates the mean hazard in areas where the highest concentration of assets are located Because the selection of the branch on which the stochastic catalog is built depends on the key arbitrary locations chosen and the geographical distribution of the assets in the portfolio used to guide such a selection, the end user has no way of knowing the sign and amplitude of the bias of the final loss estimates obtained using this single-branch–based simplified approach on his/her portfolio, whose characteristics may be materially different than those hypothesized during the selection of the key locations.

In addition to the bias just discussed, the single-branch–based stochastic catalog also causes a consistent underestimation of the central 70% interpercentile range of key loss metrics (e.g., losses corresponding to given annual rates of exceedance or AALs), but the entity of such an underestimation varies from case to case. In general, we can say the central 70% interpercentile range around the mean loss estimates from the simplified approach in the cases considered here may be as narrow as the central 40% interpercentile range estimated using the fully probabilistic benchmark case. This underestimation is, however, not crucial because the uncertainty around the mean losses is of no use in any practical application.

To reduce or remove the bias, we investigated an alternative simplified but improved approach that makes use of a compendium catalog of events extracted from all seismic source and ground motion models (i.e., from all the branches of the logic tree). This alternative approach considerably improves the accuracy of the loss estimates. The EP curves and AAL estimates are, on average, less biased than those of the original branch-based approach. However, for the bias of such results to be contained within, say, 10% of the length in years of the compendium catalog needs to be much longer than that of the customary catalogs (e.g., 10,000 years) used in earthquake loss estimation models. Hence unlike the branch-based simplified approach whose accuracy of results can hardly be improved, the compendium catalog approach may lead to unbiased loss estimates, but at a price—namely, the use of a longer stochastic catalog of events. Note that the compendium catalog approach does not improve the estimates of the central quantile of the loss distribution, a caveat, however, that is not important in real-life applications.

ACKNOWLEDGMENTS

The original idea for this study was conceived several years ago during a discussion between one of the authors and Dr. Nesrin Basöz of GeoVera; her contribution is gratefully acknowledged. This work also benefited from thorough discussions with Dr. Nilesh Shome of Risk Management Solutions and Dr. Mario Ordaz of Evaluación de Riesgos Naturales.

REFERENCES

Bommer, J. J., and Scherbaum, F., 2008. The use and misuse of logic trees in probabilistic seismic hazard analysis, *Earthquake Spectra* **24**(4), 997–1009.

Boore, D. M., and Atkinson, G. M., 2008. Ground-motion prediction equations for the average horizontal component of PGA, PGV, and 5%-damped PSA at spectral periods between 0.01 s and 10.0 s, *Earthquake Spectra* **24**(1), 99–138.

Chiou, B. J., and Youngs, R. R., 2008. An NGA model for the average horizontal component of peak ground motion and response spectra, *Earthquake Spectra* **24**(1), 173–215.

Ellsworth, W. L., 2003. *Magnitude and Area Data for Strike Slip Earthquakes* US Geol. Surv. Open File Rep. 2003:03-214.

Field, E. H., Dawson, T. E., Felzer, K. R., Frankel, A. D., Gupta, V., Jordan, T. H., Parsons, T., Petersen, M. D., Stein, R. S., Weldon, R. J., and Wills, C. J., 2009. Uniform California earthquake rupture forecast, version 2 (UCERF 2), *Bulletin of the Seismological Society of America* **99**(4), 2053–2107.

Hanks, T. C., and Bakun, W. H., 2002. A bilinear source-scaling model for M-log A observations of continental earthquakes, *Bulletin of the Seismological Society of America* **92**(5), 1841–1846.

Jayaram, N., and Baker, J. W., 2009. Correlation model for spatially distributed ground-motion intensities, *Earthquake Engineering & Structural Dynamics* **38**(15), 1687–1708.

Kircher, C. A., Seligson, H. A., Bouabid, J., and Morrow, G. C., 2006. When the big one strikes again—estimated losses due to a repeat of the 1906 San Francisco earthquake, *Earthquake Spectra* **22**(S2), 297–339.

McGuire, R. K., Cornell, C. A., and Toro, G. R., 2005. The case for using mean seismic hazard, *Earthquake Spectra* **21**(3), 879–886.

Ordaz, M., and Arroyo, D., 2016. On uncertainties in PSHA, *Earthquake Spectra* **32**(3), 1405–1418.

Pagani, M., Monelli, D., Weatherill, G., Danciu, L., Crowley, H., Silva, V., Henshaw, P., Butler, L., Nastasi, M., Panzeri, L., and Simionato, M., 2014. OpenQuake engine: an open hazard (and risk) software for the global earthquake model, *Seismological Research Letters* **85**(3), 692–702.

Park, J., Bazzurro, P., and Baker, J. W., 2007. Modeling spatial correlation of ground motion intensity measures for regional seismic hazard and portfolio loss estimation, *Applications of Statistics and Probability in Civil Engineering,* Taylor & Francis Group, 1–8.

Parzen, E., 1999. *Stochastic Processes: Classics in Applied Mathematics*, volume **24**, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

Petersen, M. D., Frankel, A. D., Harmsen, S. C., Mueller, C. S., Haller, K. M., Wheeler, R. L., Wesson, R. L., Zeng, Y., Boyd, O. S., Perkins, D. M., Luco, N., Field, E. H., Wills, C. J., and Rukstales, K. S., 2008. Documentation for the 2008 Update of the United States National Seismic Hazard Maps: U.S. Geological Survey Open-File Report 2008–1128, 61 pp.

Petersen, M. D., Moschetti, M. P., Powers, P. M., Mueller, C. S., Haller, K. M., Frankel, A. D., Zeng, Y., Rezaeian, S., Harmsen, S. C., Boyd, O. S., Field, E., Chen, R., Rukstales, K. S., Luco, N., Wheeler, R. L., Williams, R. A., and Olsen, A. H., 2014. Documentation for the 2014 update of the United States national seismic hazard maps: U.S. Geological Survey Open-File Report 2014–1091, 243 p., https://dx.doi.org/10.3133/ofr20141091

Porter, K., 2010. Cracking an open safe: uncertainty in HAZUS-based seismic vulnerability functions, *Earthquake Spectra* **26**(3), 893–900.

Silva, V., Crowley, H., Pagani, M., Monelli, D., and Pinho, R., 2014. Development of the OpenQuake engine, the Global Earthquake Model's open-source software for seismic risk assessment, *Natural Hazards* **72**(3), 1409–1427.

(Received 5 February 2015; accepted 21 March 2017)