Model Order Reduction for UCERF3-TD for Loss Exceedance

Keith Porter, University of Colorado Boulder
Kevin Milner, University of Southern California
Ned Field, United States Geological Survey

The Uniform California Earthquake Rupture Forecast version 3-Time Dependent is a complex mathematical model of where California's seismic faults are and how frequently they produce earthquakes. The model is represented by a logic tree with 5,760 leaves, each representing one combination of modeling choices to represent epistemic uncertainties. To use the model in risk analysis, one must often add epistemic uncertainties, such as which of several ground-motion-prediction equations to use. Doing so can increase the model size to 172,800 leaves. Each leaf still has explicit uncertainties, often called aleatory uncertainties, such as whether each of 6 million possible earthquakes will occur and the resulting map of spatially correlated shaking. To use the model in practice with all these uncertainties can be computationally demanding. It is desirable to find a subset of epistemic uncertainties that preserve the distribution of important dependent variables. We previously showed how to trim the logic tree to preserve the distribution of expected annualized repair cost to a large portfolio of California buildings. Here we show how to trim the logic tree to preserve the distributions of expected annualized repair cost and of loss with various rare exceedance frequencies: 1 time in 100 years, 250 years, 400 years, 550 years, and 2,500 years. It appears that one can reduce the logic tree to as few as 15 logic tree leaves (from 172,800), varying only ground motion model and ground motion model added epistemic uncertainty, at least for the 400-year and 550-year losses. A hypothetical risk calculation that takes 24 hours for the full model (evaluating all 172,800 leaves) can be reduced to a calculation that takes 8 seconds (for a reduced-order model with 15 leaves). But in all cases examined here, one can trim the logic tree by at least three variables. Because of the exponential relationship between the number of logic tree branches and the size of the model, even trimming three variables can produce a huge savings in computational effort, reducing the number of logic tree leaves to 4% of the size of the full model. Model order reduction can have important financial benefits. With more study, it may be possible to reduce uncertainty around the choice of ground motion model and of added epistemic uncertainty in the ground motion model. Lower uncertainty means thinner tails to the distribution of loss, that is, lower loss associated with rare exceedance rates. If an insurer can reduce its estimate of rare loss, it can buy less reinsurance and save money.

1. Introduction

1.1 Model order reduction for functions that include nominal random variables

The Uniform California Earthquake Rupture Forecast version 3-Time Dependent (UCERF3-TD, Field et al. 2015) mathematically models seismic activity in California. UCERF3-TD can be represented using a logic tree with eight modeling choices—branches in the logic tree—often called epistemic uncertainties. Some of the uncertainties have to do with an uncertain scalar quantity such as the maximum magnitude of earthquake that can occur off the fault. Some are presented by nominal variables, meaning variables that can take on different values, but the values have neither scale nor order, such as the choice between two depictions of the physical geometry of the larger and more active faults. Each added epistemic uncertainty multiples the size if the model—the number of possible logic-tree leaves—by the number of possible values the uncertainty can take on. The model grows exponentially with the number of variables. Using the full model can be computationally expensive if one wants to quantify the tails of a distribution of a function of those variables.

How expensive? Consider the insurance loss to a portfolio of 1 million policies subject to 6,000,000 possible ruptures on 15,700 known faults, with a model that comprises 172,800 combinations of model elements. (UCERF3-TD has 5,760 possible combinations, but one needs three more variables to estimate uncertain ground motion, increasing the model by a factor of 30.) Despite Moore's Law, quantifying the tails of a distribution of a function of such a complex model can cost so much that the analyst must resort to simplifications, while trying to maintain the shape of the distribution of the function. How can the analyst best reduce the model without introducing bias into the tails of the distribution?

Most model order reduction techniques work when the model is a function of scalar random variables. In a prior work, we offered a path-search technique that can reduce a model that has nominal random variables too. One starts with the full model and one at a time trims off variables that do not strongly affect the distribution of the output function. In that prior work we illustrated the technique by trimming the UCERF3-TD logic tree to maintain the distribution of expected annualized loss, which matters to the premium rates that insurers charge their policyholders. This time we care about a value at a high tail of the distribution. Insurers care about the tails of the distribution of loss in a coming year because they must buy reinsurance to ensure they can financially survive the coming year with high confidence.

1.2 Objectives

This document adapts our previous path-search model-order reduction technique to maintain the distribution of a value on the tail of the model's dependent variable. Let L denote the dependent variable, for example the largest loss that an insurer will suffer in the coming year. Let L_p denote a point on the high tail of the distribution of L, for example the value of portfolio loss L that is expected to be exceeded only 1-p times in the coming year. If p = 1/250, for example, L_p reflects the 250-year loss.

L_p is a function of many independent variables, many of which are nominal, each of which has a probability distribution with a weight or probability applied to each possible value. Variables are arranged as branching points in a logic tree. The problem addressed here is, how can we remove logic-tree branches by fixing them at one of their possible values, reducing the number of leaves to minimize the model size without introducing unacceptable error into L_p?

This is a problem of model order reduction for an extreme value of a high-order model that includes nominal uncertain variables (see e.g., Schilders et al. 2008 for general treatment). The model order reduction technique proposed here is illustrated using an earthquake insurance problem, but generally applies to extreme values in high-order models that include nominal uncertain variables. We seek to show how to quantify the tradeoff between degree of model order reduction and error in L_p. We also repeat the demonstration for expected annualized loss, *EAL*.

2. Literature review

2.1 Logic tree model of mixed independent variables in an earthquake rupture forecast

Field et al. (2014) offer a new earthquake rupture forecast for California (the Uniform California Earthquake Rupture Forecast version 3, Time-Independent, or UCERF3-TI) with seven uncertain model components. Field et al. (2015) add an eighth element that models aperiodicity in earthquake recurrence to make the model time-dependent (and hence dubbed the Uniform California Earthquake Rupture Forecast version 3, Time-Dependent, or UCERF3-TD). Model elements are arranged in a logic tree with 5,760 possible combinations (leaves) of their eight branches. See Figure 1. The eight branches represent uncertain variables, of which three are scalar random variables and five are nominal, meaning that their possible values lack scale or order. Such a large model can make calculating losses to a large earthquake portfolio computationally demanding, if one wants to treat all possible combinations of the variables and their associated weights (or Bayesian probabilities). Reducing the size of such a model can be challenging because most model order reduction techniques work on functions of scalar random variables, not on models with nominal random variables. The computation problem of estimating loss to an insurance portfolio is compounded by the need to treat additional uncertainties: site characteristics, ground motion model, and added epistemic uncertainty in ground-motion models. (There are more as well that are not treated here.) A model that includes all these variables has 172,800 leaves.

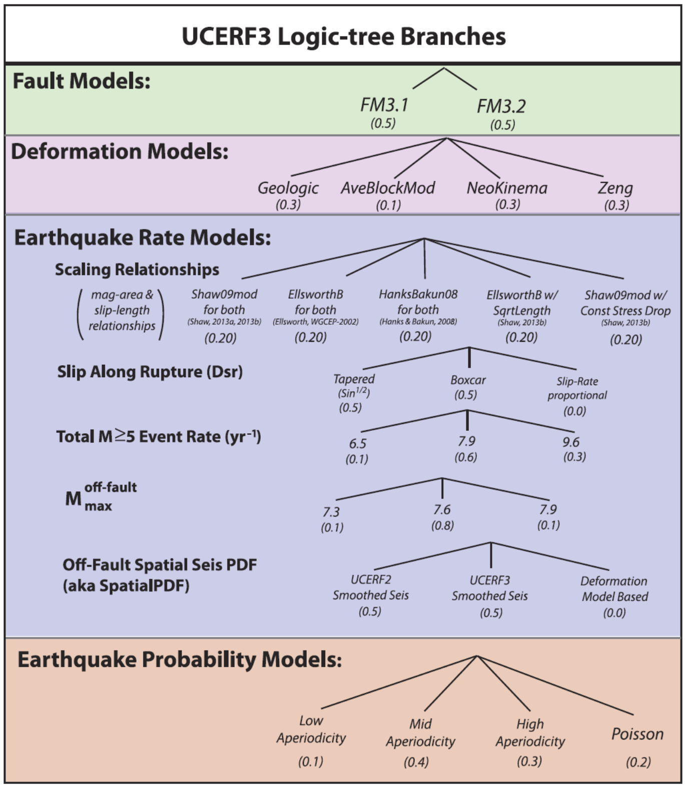


Figure 1. UCERF3-TD logic tree. Each branching point represents an uncertain variable; each branch a possible value.

2.2 Model order reduction techniques

To reduce the computational expense of UCERF3-TD, some authors have replaced the earthquake rupture forecast with a Monte Carlo time series called an event set. That is, one creates a sequence of scenario earthquakes spread over thousands of years or more, consistent with the earthquake rupture forecast. For example, Perkins and Taylor (2003) use a 50,000-year event set to estimate risk to a roadway system. They find the effort highly computationally demanding and attempted a variety of model order reduction techniques, including bootstrap sampling, the use of antithetic variates, the use of Latin Squares (or permutation) sampling, the use of control functions, a compound Poisson approach, and importance sampling. They find that large reductions in the number of simulations needed could be achieved for the mean and confidence limits of the conditional loss distribution (the loss distribution given some loss in a specific year). However, for the unconditional, annual-loss distribution, the reduction of the number of simulations achieved through post-sampling techniques was only a multiplicative reduction factor of slightly above 3.

In prior work (Porter et al. 2012) we applied a deterministic sensitivity analysis technique called tornado-diagram analysis meant to identify the likely important variables in a complex earthquake rupture forecast. Later, we discussed how existing model-order reduction techniques seem to be intended for functions of scalar random variables (Porter et al. 2017). In that work, we offer a new model order reduction technique that works on models with nominal random variables. We applied it to the UCERF3-TD, using the expected annualized loss, *EAL*, to a proxy for the California Earthquake Authority's statewide insurance portfolio. (*EAL* measured ground-up repair cost rather than insured loss after deductibles and limits.) We applied what one might call a path search from the full model to a reduced-order model that was 3 orders of magnitude smaller (involving 60 leaves rather than 57,600) without significantly biasing the distribution of *EAL* relative to the full model. In that work, the model did not include an added epistemic uncertainty to the ground motion models, which with three possible values increases the model size to 172,800 leaves.

3. Methodology

3.1 Overview

The objective of the present model-order-reduction technique is to reduce the model without greatly disturbing the probability distribution of a point on the tail of the dependent variable. To do so, we apply the path-search model-order reduction technique developed in Porter et al. (2017) to the present problem. To make the analysis easier to follow, let us consider it in stages:

1. Evaluate the model output (here, the loss exceedance curve at a given non-exceedance probability) for one logic tree leaf.
2. Evaluate the cumulative distribution function (CDF) of the output for the original (full) model.
3. Evaluate the CDF of output for a reduced-order model, that is, the full model minus one variable, which is fixed at one of its possible values. Compare the CDF it with that of the full model. Do so for each possible reduction of the model by one variable. Identify the one reduced model whose output CDF is most like that of the full model.
4. Apply the path-search approach to find the best reduced-order models, successively trimming variables from the full model. That is, repeat step 3, successively reducing the reduced order model until all variables are fixed, tracking the goodness of fit of the CDF of each successive reduced order model's output to that of the full model.

3.2 Evaluate the model output for one logic-tree leaf

This model-order-reduction technique addresses a model that outputs a scalar value that can depend on the value of each model element, each independent variable. Each logic tree leaf can produce a different output value. In this step, we calculate the deterministic model output conditioned on the values of all the independent variables.

In the present application, we are concerned with an extreme value on the cumulative distribution function of insured loss for one year. Insurers tend to view the distribution of loss in a year with the complement of the CDF, that is, the probability that the value of loss exceeds each of many dollar values. The function is referred to as a loss exceedance curve. Often the insurer is more concerned with the rate at which losses are exceeded (in events per year) rather than the loss with a certain exceedance probability in a year (which is unitless), but at values smaller than about 0.1, rate and probability take on almost the same numerical value and differ only by units. Let us ignore the distinction here between low rates and low probabilities. The extreme value we are about is the loss with a prescribed low exceedance probability in the coming year, for example p = 0.002, which one can think of as the 500-

year loss (p = 1/500 = 0.002). Because the 500-year loss can vary depending on the value of the model variables, and because those values have prescribed weights or Bayesian probabilities, the 500-year loss has an associated cumulative distribution function. Let is first find its value conditioned on fixing all the variables at one of their possible values.

In the case of an insurance portfolio, here is one way to evaluate the loss exceedance curve for one logic-tree leaf, as applied to a portfolio of earthquake insurance policies in California. In the following, important aspects of the leaf are implicit, such as which ground motion model is used, which Vs30 model to use, etc.

N_k = number of possible ruptures among full UCERF3-TD model

k = an index to scenario ruptures (“ruptures”), $k \in \{0, 1, \ldots N_k - 1\}$

N_a = number of assets in the portfolio

a = an index to assets in the portfolio, $a \in \{0, 1, \ldots N_a - 1\}$

V_a = replacement cost of asset a

V = replacement cost of the portfolio

$$V = \sum_a V_a \tag{1}$$

x = a measure of ground motion, e.g., 5% damaged elastic spectral acceleration response at 1.0 sec period

$y_a(x)$ = mean repair cost as a fraction of replacement cost new for asset a, given ground motion x

$X_{a|k}$ = uncertain ground motion at asset a given rupture k

$f_{Xa|k}(x)$ = probability density function of $X_{a|k}$, evaluated at x

$\mu_{L,|k}$ = expected value of portfolio loss L given rupture k

$$\mu_{L|k} = \sum_{a=0}^{N_a-1} V_a \int_{x=0}^{\infty} y_a(x) f_{X_a|k}(x)\,dx \tag{2}$$

$\delta_{L|k}$ = coefficient of variation of portfolio loss in rupture k. See Appendix 1 for a method to estimate $\delta_{L|k}$ as a function of $\mu_{L|k}$. As others have found for individual assets (e.g., Porter 2010), portfolio loss uncertainty decreases with increasing portfolio loss, as in the exponential relationship

$$\delta_L \approx c_1 \cdot \left(\frac{1000}{V} \mu_L \right)^{c_2} \tag{3}$$

In the equation, the coefficient 1000/*V* expresses the mean loss in terms of loss per $1000 of replacement cost, a normalized loss measure sometimes used in the catastrophe-risk modeling industry. Appendix 1 presents a regression analysis that suggests coefficients c_1 and c_2 as follows. The resulting curve gradually drops from 2 (at low portfolio losses) to 0.5 (at high portfolio loss).

$c_1 = 0.9832$

$c_2 = -0.117$

$\theta_{L|k}$ = median value of portfolio loss *L* given rupture *k*, assuming that *L* is approximately lognormally distributed. See Appendix 1 for evidence that this assumption is reasonable.

$$\theta_{L|k} = \frac{\mu_{L|k}}{\sqrt{1+\left(\delta_{L|k}\right)^2}} \tag{4}$$

$\beta_{L|k}$ = standard deviation of the natural logarithm of portfolio loss *L* given rupture *k*, assuming that *L* is approximately lognormally distributed

$$\beta_{L|k} = \sqrt{\ln\left(1+\left(\delta_{L|k}\right)^2\right)} \tag{5}$$

r_k = rate at which rupture *k* occurs (here and elsewhere: given logic tree leaf)

L = uncertain portfolio loss

l = a value of *L*

$G(l)$ = number of earthquakes per year producing $L \geq l$. The relationship between $G(l)$ and *l* is often called the loss exceedance curve. By the theorem of total probability, the rate is the sum of event rates r_k times probability that the loss in rupture *k* is greater than or equal to *l*:

$$G(l) = \sum_{k=0}^{N_k-1} r_k \cdot \left(1 - \Phi\left(\frac{\ln\left(l/\theta_{L|k}\right)}{\beta_{L|k}}\right)\right) \tag{6}$$

Later we will be interested in the loss with a specified exceedance rate, rather than the exceedance rate of some value of loss, so recalling that L_p denotes the loss with exceedance rate *p*, it is the inverse of the loss exceedance curve evaluated at *p*:

$$L_p = G^{-1}(p) \tag{7}$$

3.3 Evaluate the CDF of the model output for the original model

Equations (1) through (7) show how to calculate the loss exceedance curve for one leaf. Next, we combine the exceedance curves for the model leaves and evaluate the exceedance curve for the original, full, model.

Z = number of leaves in the original model

j = an index to leaves, $j \in \{0, 1, \ldots Z-1\}$

w_j = weight of leaf j in the full model

$L_{p,j}$ = loss associated with exceedance frequency p in logic-tree leaf j, from equation (7). Note that each leaf j can have a different loss exceedance curve and therefore a difference value of loss associated with exceedance frequency p, and therefore a probability distribution of L_p, as illustrated in Figure 2. The figure shows a suite of loss exceedance curves for many logic-tree leaves. It also shows a horizontal line at some exceedance rate p of interest (0.004 per year), and a probability density function of L_p. The probability density function has some mean value that we could denote by μ_{Lp} and a coefficient of variation denoted by δ_{Lp}. Note that it will not be necessary to assume a parametric form of the distribution of L_p such as normal or lognormal.

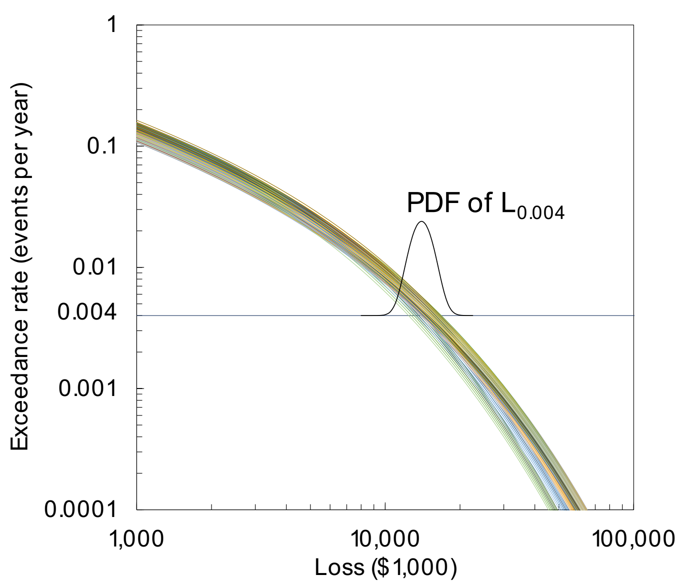


Figure 2. Illustration of the probability density function of L_p. The colored curves represent loss-exceedance curves for different logic-tree branches. The present model-order-reduction effort aims to reduce the number of possible loss exceedance curves (thereby simplifying the model and reducing computational effort) without strongly affecting the PDF of large rare loss.

$F_{Lp}(l)$ = cumulative distribution function for L_p

$$F_{L_p}\left(l\right) = \sum_{j=0}^{Z-1} w_j \cdot H\left(l - L_{p,j}\right) \tag{8}$$

where H is the Heaviside function, that is,

$$\begin{aligned} H(x) &= 0 \qquad x < 0 \\ &= 0.5 \qquad x = 0 \\ &= 1 \qquad x > 0 \end{aligned} \tag{9}$$

μ_{Lp} = expected value of L_p in the original model,

$$\mu_{L_p} = \sum_{j=0}^{Z-1} L_{p,j} \cdot w_j \tag{10}$$

σ^2_{Lp} = variance of L_p in the original model,

$$\sigma^2_{L_p} = \left(\sum_{j=0}^{Z-1} \left(L_{p,j} \right)^2 \cdot w_j \right) - \left(\mu_{L_p} \right)^2 \tag{11}$$

δ_{Lp} = coefficient of variation of L_p in the original model:

$$\delta_{L_p} = \frac{\sigma_{L_p}}{\mu_{L_p}} \tag{12}$$

3.4 Loss exceedance curve for a reduced model

Next let us evaluate the exceedance curve for one reduced model and measure the error in L_p.

I_j = a binary indicator (1,0) whether a reduced model includes (I_j = 1) or excludes (I_j = 0) logic-tree leaf j

z = model size of reduced-order model, meaning the number of leaves in the reduced model

$$z = \sum_{j=0}^{Z-1} I_j \tag{13}$$

c_0 = normalizing constant for weights in the reduced-order model

$$c_0 = \sum_{j=0}^{Z-1} w_j \cdot I_j \tag{14}$$

Now find the cumulative distribution function of L_p in the reduced model:

$\hat{F}_{L_p}(l)$ = cumulative distribution function for L_p in reduced model

$$\hat{F}_{L_p}(l) = \frac{1}{c_0} \cdot \sum_{j=0}^{Z-1} w_j \cdot I_j \cdot H\left(l - L_{p,j} \right) \tag{15}$$

where *H* is still the Heaviside function.

$\hat{\mu}_{L_p}$ = expected value of L_p in the reduced model

$$\hat{\mu}_{L_p} = \frac{1}{c_0} \sum_{j=0}^{N_j - 1} L_{p,j} \cdot w_j \cdot I_j \tag{16}$$

$\hat{\sigma}^2_{L_p}$ = variance L_p in the reduced model

$$\hat{\sigma}^2_{L_p} = \left(\frac{1}{c_0} \sum_{j=0}^{Z-1} \left(L_{p,j} \right)^2 \cdot w_j \cdot I_j \right) - \left(\hat{\mu}_{L_p} \right)^2 \tag{17}$$

$\hat{\delta}_{L_p}$ = coefficient of variation of L_p in the reduced model

$$\hat{\delta}_{L_p} = \frac{\hat{\sigma}_{L_p}}{\hat{\mu}_{L_p}} \tag{18}$$

Now we check the goodness of fit for the reduced-order model, that is, how well $\hat{F}_{L_p}$ matches that of the full model, F_{Lp}. We apply the two-sample Kolmogorov-Smirnov goodness-of-fit test at the 1% significance level. One calculates the maximum difference in the cumulative distribution functions, D_n, as in equation (19), and checks that it is less than the limit shown in equation (20).

$$D_n = \max_l \left(\left| F_{L_p}(l) - \hat{F}_{L_p}(l) \right| \right) \tag{19}$$

$$D_n \le 1.63 \sqrt{\frac{z + Z}{z \cdot Z}} \tag{20}$$

It is also desirable to ensure that errors in the mean and coefficient of variation of L_p are both less than some reasonable limit, say 5%:

$$\left| \varepsilon_\mu \right| \le 0.05 \tag{21}$$

$$\left| \varepsilon_\delta \right| \le 0.05 \tag{22}$$

Where

$$\varepsilon_\mu = \frac{\hat{\mu}_{L_p} - \mu_{L_p}}{\mu_{L_p}} \tag{23}$$

$$\varepsilon_\delta = \frac{\hat{\delta}_{L_p} - \delta_{L_p}}{\delta_{L_p}} \tag{24}$$

If the reduced model passes the test specified in equation (20), we can reject at the 1% significance level that the two distributions differ. If it fails equation (21), the reduced model is drifting too far in the mean, even if the Kolmogorov-Smirnov test says that it and the full model are still drawn from the same distribution. If it fails equation (22), the reduced model is (probably) getting too certain, again even if the Kolmogorov-Smirnov test says it is drawn from the same distribution.

3.5 Model order reduction technique: path search

Finally, let us apply the path-search technique from Porter et al. (2017) to model order reduction for L_p. Here are its steps:

1. Evaluate $F_{Lp}(l)$, μ_{Lp}, and δ_{Lp} for the original model.
2. Let a denote an index to independent variables and b denote an index to possible values (e.g., Table 1). For each (a, b) pair, fix variable a at value b. For each leaf j, calculate D_n, ε_μ, and ε_δ from equations (19), (23), and (24), where I_j = 1 if the leaf has variable a equal to value b, or I_j = 0 if otherwise.
3. Trim the first branch (c = 0) by selecting the (a, b) pair with the smallest value of D_n that satisfies the goodness of fit test in equation (20) and the additional inequalities (21) and (22). Fix variable a at value b. Variable a is no longer a free variable. One can say the model has been reduced by variable a. Record the model size z of the model with one trimmed branch.
4. Trim the second branch (c = 1) by repeating steps 2 and 3 starting with the reduced-order model from step 3, but allowing every remaining (a,b) pair where a has not already been fixed.
5. Repeat until all branches are fixed (c = 2, 3, ... N_c – 1) where N_c is the number of branches in the logic tree.

4. Application to UCERF3-TD tree trimming problem

4.1 Variables

4.1.1 Independent variables: branches of UCERF3-TD plus three intensity-related branches

Table 1 summarizes the independent variables considered here: their type (scalars, denoted by S, ordinals, denoted by O, and nominal, denoted by N), their possible values, weights (that is, their conditional probabilities in a Bayesian sense), and a brief description. The description explains to the reader who is unfamiliar with UCERF3 what the variable models. The description includes notes about how influential one might expect the variable to be on overall uncertainty in rare portfolio loss. These notes are largely drawn from observations by Field et al. (2013) on the influence each variable has on peak ground acceleration with 2% exceedance probability in 50 years.

Variables 0 through 7 are elements of UCERF3-TD. They represent 2 × 4 × 5 × 2 × 3 × 3 × 2 × 4 = 5,760 possible combinations. To calculate the repair cost to a portfolio of buildings requires additional variables 8 through 10, that is, variables that are exogenous to UCERF3-TD but endogenous to the (broader) loss model used here to trim the UCERF3-TD logic tree using losses. Variables 8, 9, and 10 have 2 × 5 × 3 = 30 possible combinations, for a total of 172,800 model leaves when combined with the UCERF3-TD leaves. Of the 11 variables, four (numbers 4, 5, 7, and 10) involve scalar quantities and the others are nominal, that is, a choice among values with no order or scale.

To calculate repair cost for a single scenario and a loss exceedance curve for the suite of earthquakes in the rupture forecast also requires some more input information that one could consider to be independent variables:

1. ***Portfolio.*** This is an estimate of the assets exposed to risk, each asset parameterized with its geographic location, site conditions (Vs30), replacement cost new (the cost to build a new facility approximately functionally and aesthetically equivalent to the existing one), and a building type. "Building type" is often parameterized (as it is here) by structural material (e.g., wood), lateral force resisting system (e.g., shearwall), height category (e.g., 1-3 stories), and era of construction (e.g., pre-1940). We estimated the inventory of woodframe single-family dwellings in California using a 2002-era database in Hazus-MH, factored up on a statewide basis to account for population growth and construction costs, and then factored down on a county-by-county basis to account for the CEA's market penetration rate—that is, the fraction of homes that are insured by the CEA. We use a fixed value of the portfolio, rather than varying it. In the present case, the portfolio has an estimated replacement cost new of $483 billion (2019 USD).
2. ***Vulnerability functions.*** These relate ground motion to mean (and sometimes variability) of repair cost as a fraction of replacement cost new. We used the Hazus-based vulnerability functions described in Porter (2009a, b, 2010). Vulnerability functions can be considered a variable that we fixed. Other models are available, but to vary the vulnerability functions seems relatively unimportant for the present objective of model order reduction of UCERF3-TD.

Table 1. Independent variables, variable types, possible values, weights (conditional probabilities), and brief description.

a	Variable (branch) name	Type	*b*	Possible value[1]	Weight	Description. See Field et al. (2013) Table 15 for maps of size and extent of effects.
0	Fault model	N	0	3.1	0.5	Geometry of larger, more active faults. FM3.1 has 2,606 subsection and 253,706 multi-subsection ruptures; FM3.2, 2,665 and 305,709. Modest (±12% PGA) differences in a few (~5) local (≤100 km) areas. Neither choice is closer to UCERF3.3 average.
			1	3.2	0.5	
1	Deformation model	N	0	Geol	0.3	Slip rates and related factors for each fault section; strain accumulation before fault rupture; energy released. Reflects approach to handling earthquake dynamics. Significant effects on 2%/50-year PGA (±25%) over many large regions (≥200 km). GEOL and ZENGBB are closer to UCERF3.3 average than others.
			1	ABM	0.1	
			2	NeoK	0.3	
			3	ZengBB	0.3	
2	Scaling relationship	N	0	SHAW 09m	0.2	Relates earthquake magnitude to rupture surface area or to area and rupture aspect ratio (length divided by width). Also relates slip length to rupture length and width. Effects are modest (±12%) but affects many large regions (≥200 km). ELL B SQL and SHAW09m closer to UCERF3.2 average 2%/50-year PGA than others.
			1	ELL B	0.2	
			2	H&B 08	0.2	
			3	ELL B SQL	0.2	
			4	SHAW CSD	0.2	
3	Slip along rupture	N	0	Tapered	0.5	Relates fault slip to location along rupture. Very little influence: modest effect (±12%) in a few (~5) local (≤100 km) areas.
			1	Boxcar	0.5	
4	Total M>5 event rate yr^{-1}	S	0	6.5	0.1	Small (±5%) effect throughout much of California, but mostly away from metro areas. 7.9 closest to UCERF3.3 average 2%/50-year PGA.
			1	7.9	0.6	
			2	9.6	0.3	
5	Maximum off-fault magnitude	S	0	7.3	0.1	Maximum magnitude of earthquakes away from mapped faults. Almost no noticeable influence on 2%/50-year PGA from any of the three models.
			1	7.6	0.8	
			2	7.9	0.1	
6	Off-fault spatial seismicity PDF	N	0	UCERF2	0.5	Depicts the spatial distribution of off-fault gridded seismicity. Significant (±25%) influence throughout much of California, but mostly away from metro areas. Neither choice is closer to UCERF3.3 average.
			1	UCERF3	0.5	
7	Earthquake probability model	N	0	Low COV	0.1	Estimates how ready each fault segment is to rupture given stress accumulation since last rupture. Probabilities are lower on faults with recent large earthquakes, higher the last rupture occurred at least half the average recurrence interval. Mid to high COV (aperiodicity) likely closer to average than the other, more extreme, options.
			1	Mid COV	0.4	
			2	High COV	0.3	
			3	Poisson	0.2	
8	Vs30 model	N	0	Wills (2015)	0.5	Average shear-wave velocity in upper 30m of soil based either on correlation between observed Vs30 and geologic unit (Wills et al. 2015) or topographic slope (Wald and Allen 2007).
			1	Wald & Allen (2007)	0.5	
9	Ground-motion-prediction equation	N	0	ASK2014	0.22	Relates ground motion (e.g., 5% damped spectral acceleration response) to magnitude, distance, fault attributes, and site conditions. BSSA 2014 and CY2014 tend to be closer to the average of the four for common conditions in the middle distance (10-30 km) for a large (M7.8) earthquake on common site conditions (Vs30 = 300 m/sec, D1.0 = 100 m, D2.5 = 1 km). Significant (±25%) influence statewide.
			1	BSSA2014	0.22	
			2	CB2014	0.22	
			3	CY2014	0.22	
			4	IDR2014	0.12	
10	Added epistemic uncertainty	S	0	Low	0.185	Adds ground motion uncertainty to account for collaboration among the NGAW-2 developers and their use of common sets of statistical analyses and simulations to constrain parts of the models. Likely to have significant statewide effect.
			1	Med	0.630	
			2	High	0.185	

1. Abbreviations consistent with Field et al. (2013)

4.1.2 Intermediate variables: within- and between-event variability

The dependent variables involve earthquake shaking-induced loss to a portfolio of assets. To calculate them will require propagating uncertainty in ground motion conditioned on each of many fault ruptures and attendant

1. ***Within-event ground motion variability.*** Within-event positive spatial correlation of ground motion tends to result in within-event positive correlation of repair costs between assets. That makes buildings located within a few kilometers of each other experience higher or lower ground motion together. As a result, repair costs are not independent and identically distributed (IID) conditioned on median ground motion, and we cannot simply sum the variances of the

asset repair costs to get the variance of the total portfolio repair cost. The variance of portfolio repair cost will tend to be larger, potentially much larger, than this desirable simplification would imply. To deal with this problem we simulated within-event ground motion using 100 realizations of spatially correlated standard normal fields at grid spacing of 1-km each way (north-south and east west), each field being 800 km on a side (i.e., 801 by 801 gridpoints), using the model of spatial correlation proposed by Jayaram and Baker (2009) for 1-second period 5% damped elastic spectral acceleration response. See Figure 3 for four of the 100 realizations. Why 1 second? This is an artifact of the Hazus-based vulnerability model (Porter 2009a, b), which measures ground motion with a vector of 0.3-sec and 1.0-sec 5% damped elastic spectral acceleration response. At high excitations, which cause most of the damage, the 1-second component matters much more than the 0.3-sec component. Let $e_{\phi o}$ denote the value of the within-event field in one realization at a location denoted by o.

2. ***Between-event ground motion variability.*** Between-event variability affects all buildings in the portfolio simultaneously. A positive difference between the actual earthquake ground motion across the entire field and the median will tend to increase damage. The between-event variability is modeled with lognormal distribution where the natural logarithm of the residual has zero mean and standard deviation denoted by τ. One can normalize the residual by dividing by τ, and denote the normalized residual by e_τ, which has standard normal distribution.

Ground motion at a location *o* (denoted here by x_o) can then be estimated as

$$x_o = \hat{x}_o \exp\left(e_\tau \cdot \tau + e_{\phi o} \cdot \phi\right) \tag{25}$$

where

$\hat{x}_o$ = median ground motion conditioned on magnitude, distance, Vs30, ground-motion-prediction equation, and other parameters of the ground-motion-prediction equation

e_τ = standardized residual of between-event uncertainty in ground motion

τ = standard deviation of the natural logarithm of ground motion associated with between-event variability, the part that varies uniformly between events—the ground motion field is uniformly higher or lower in a single event than predicted by the ground-motion-prediction equation, with uncertainty quantified by τ.

$e_{\phi o}$ = standardized residual of within-event uncertainty in ground motion at location *o*.

ϕ = standard deviation of the natural logarithm of ground motion associated with within-event variability, the part that varies spatially within a single earthquake.

We propagate uncertainty in the Gaussian e_τ by 5-point moment matching. That is, we substitute 5 weighted sample values for the continuous Gaussian distribution of e_τ as shown in Table 2. Why these values and weights? With 5-point moment matching, we have nine free variables (the positions and weights of five samples, minus one degree of freedom because the weights must sum to unity). Which means one can set their values to exactly match the first nine moments of the continuous distribution they replace (mean, variance, skewness, etc.). If we only care about matching the first few moments, say the first three, the problem is underdefined. We can set some values so that the positions are symmetrical about zero and the weights are easy to remember. The samples shown here are selected for these reasons of convenience. When applied to a lognormal variable, these choices reproduce the first three moments in the real domain well, within a few percent.

We propagate uncertainty in e_{ϕ} by Monte Carlo simulation, using 100 realizations of the map of e_{ϕ}, like those in Figure 3, centering the map at each epicenter. Given a fixed portfolio and set of vulnerability functions, repair costs of individual assets can be considered independent, conditioned on ground motion as calculated by equation (25). See Appendix 1 for evidence that this simulation approach reproduces the proper mean.

Sample	e_{τ}	weight
0	-2	0.1
1	-1	0.1
2	0	0.6
3	1	0.1
4	2	0.1

Table 2. Moment matching points for e_{τ}

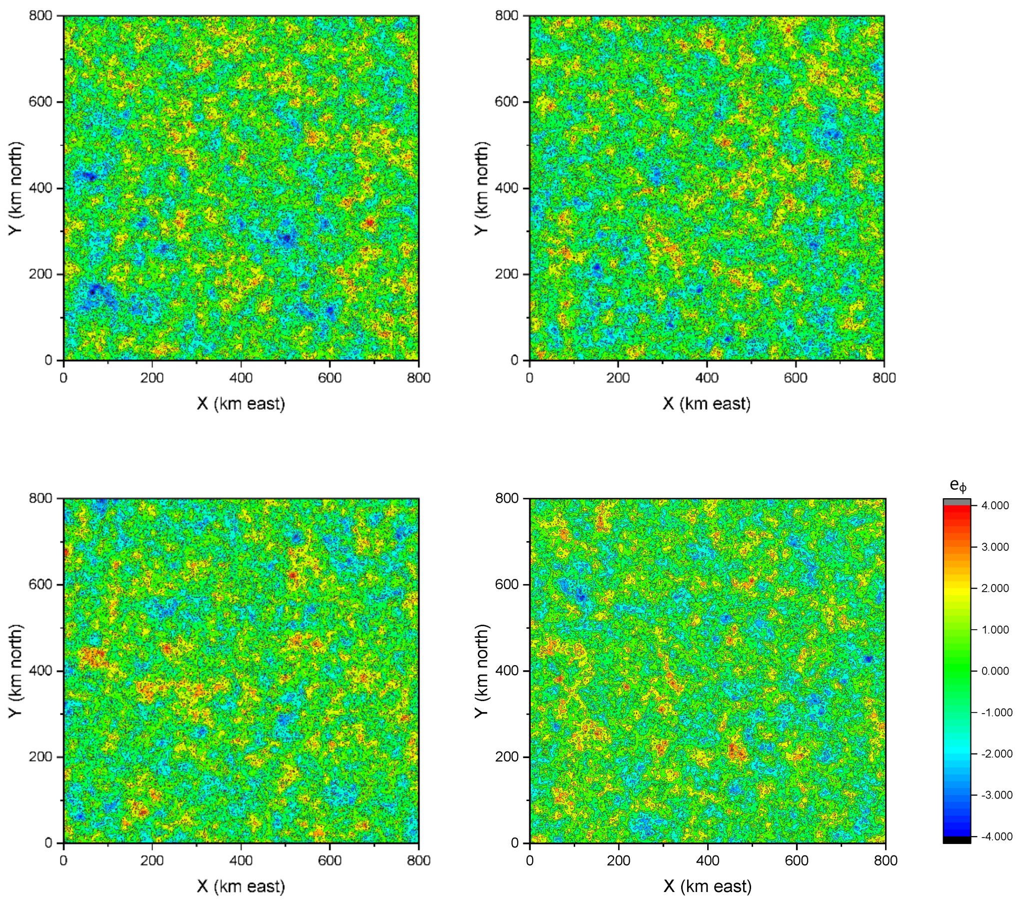


Figure 3. Four realizations of a field of spatially correlated standard normal variates, with spatial correlation as suggested by Jayaram and Baker (2008) for 1.0-second 5% damped elastic spectral acceleration response.

4.1.3 Dependent variable 1: loss *L* with mean exceedance probability *p* in one year

Insurers commonly evaluate liquidity at the 1-in-250-year mark (p = 0.004 per year) primarily because of rating agencies' target and stress-test levels since the 2004/2005 hurricane seasons. That target assumes a multi-line, multi-state insurer with diversification benefits. The California Earthquake Authority is different for exactly those reasons – one line, one state, all catastrophe risk. The California Earthquake Authority's current risk transfer strategy approved by its board and in the public domain is to maintain a minimum of 1-in-400 and a maximum of 1-in-550-year claim-paying capacity (here, p = 0.0025 to 0.0018). Therefore, we evaluate $p \in \{0.01, 0.004, 0.0025, 0.0018, 0.0004\}$.

4.1.4 Dependent variable 2: expected annualized loss *EAL*

Insurers commonly evaluate profitability by comparing the total amount of premiums charged to policyholders (called the gross premium) to the present value of future claims (called the net premium, and which catastrophe risk modelers sometimes call the expected annualized loss, *EAL*). The insurer's net premium accounts for the fact that when policyholders who incur a loss pay a deductible, that is, they pay for losses up to a prescribed amount and the insurer pays for the claim in excess of that amount. Deductibles for earthquake insurance commonly exceed 5% of the replacement cost of the property, so the insurer's *EAL* is typically much lower than the expected present value of the total loss—the policyholder's payment plus the insurer's payment—but for present purposes we ignore the deductible.

4.2 Results with loss L_p as the dependent variable

Table 3 shows the successive steps in trimming the logic tree to match the cumulative distribution function of the full model's 100-year loss. Rows labeled 0 through 9 indicate the order in which the path search fixed the variables. Row 0 provides information about the full logic tree: the number of leaves, z = 172,800. The expected value of the 100-year loss, L_p, = \$9.239 billion. The coefficient of variation of 100-year loss, δ_{Lp}, = 0.34. The following rows, labeled 1 through 11, show the trimmed branches in the order in which they are trimmed, and the value to which each is set as it is trimmed from the model. For example, the first variable to be trimmed is the maximum off-fault magnitude, 7.6. It is the variable to which L_p is least sensitive, meaning that fixing its value has the least effect on the cumulative distribution function of L_p. Fixing it reduces z to 57,600 leaves, changes L_p slightly to \$9.238 billion, has no effect on δ_{Lp} to two significant figures. The column labeled D_n shows the maximum difference between the cumulative distribution functions of the original model and the model with the first variable trimmed. There is no observable difference to three significance figures. A difference up to 0.008 would be allowable. The error in the mean and in the coefficient of variation, ε_μ and ε_δ are both less than 0.5%. Considering the goodness of fit test and the two error terms, the reduced model passes the tests laid out earlier.

The table shows that three variables can be trimmed, and the reduced-order model still reasonably approximates the full model. The smallest reduced-order model has 7,200 leaves, passes the goodness-of-fit test, and differs with the full model by less than 5% in either the mean or coefficient of variation of 100-year loss. Shaded rows below the third trimmed variable indicate that the model order reduction fails one or more of the three tests in those steps. The table shows that continuing to trim the tree introduces unacceptable error in the coefficient of variation, making it too small. If one wanted to relax the constraint on coefficient of variation to ±10%, one could trim 7 variables, leaving only 180 leaves, and still pass the goodness-of-fit and error tests.

Figure 4A shows the mean loss exceedance curves for the full model and all the reduced-order models for the cumulative distribution function of the 100-year loss. Figure 4A shows that the loss exceedance curves for the full model and reduced-order models are virtually indistinguishable. It may seem surprising that the curves do not pinch to a point at the loss with 100-year mean exceedance frequency. Remember however that the point of the technique is not simply to match that scalar value but to match the cumulative distribution function of 100-year loss, which Figure 4A does not show.

Figure 4B shows that cumulative distribution function. It shows how ever-greater model order reduction makes the cumulative distribution functions of the reduced-order models more and more loosely approximate that if the full model, but the differences tend to be small until the last three branches are trimmed (the lightest curves labeled 9, 10, and 11).

Figure 4C shows the error in the mean value of the 100-year loss, ε_μ as a function of model size *z*, along with ± 5% bounds. Figure 4D shows the error in the coefficient of variation of the 100-year loss, ε_δ, as a function of *z*. In Figure 4C and Figure 4D, only the markers (the circles) have meaning; the lines connecting them just make it easier to see the pattern of smaller error with a larger model.

Table 4 and Figure 5 present similar results for the 250-year loss. Table 5 and Figure 6 depict the tree trimming for the 400-year loss. Table 6 and Figure 7 depict the process for the 550-year loss, and Table 7 and Figure 8 do so for the 2,500-year loss.

	Trimmed branch	z	μ_{Lp}	δ_{Lp}	D_n	D_{nmax}	ε_μ	ε_δ	Pass
0	Full Tree = N/A	172800	$9,239	0.34			0%	0%	
1	MMax Off Fault = 7.6	57600	$9,238	0.34	0.000	0.008	0%	0%	TRUE
2	Fault Model = Fault Model 3.1	28800	$9,194	0.34	0.006	0.010	0%	0%	TRUE
3	ERF Probability Model = Mid COV Values	7200	$9,256	0.34	0.008	0.020	0%	1%	TRUE
4	Scaling Relationship = Shaw (2009) Modified	1440	$9,088	0.32	0.020	0.043	-2%	-6%	FALSE
5	Vs30 Model = Wald & Allen (2007)	720	$9,194	0.32	0.016	0.061	0%	-7%	FALSE
6	Slip Along Rupture Model (Dsr) = Tapered Ends	360	$9,010	0.31	0.038	0.086	-2%	-7%	FALSE
7	Spatial Seismicity PDF = UCERF3	180	$9,435	0.31	0.036	0.122	2%	-8%	FALSE
8	Total Mag 5 Rate = 7.9	60	$9,191	0.30	0.045	0.210	-1%	-12%	FALSE
9	Deformation Model = Average Block Model	15	$8,581	0.27	0.151	0.421	-7%	-19%	FALSE
10	Ground Motion Model = Abrahamson, Silva & Kamai (2014)	3	$9,080	0.28	0.202	0.941	-2%	-18%	FALSE
11	GMM Additional Epistemic Uncertainty = None	1	$8,763	0.00	0.387	1.630	-5%	-100%	FALSE

Table 3. Trimming path for 100-year loss (p = 0.01)

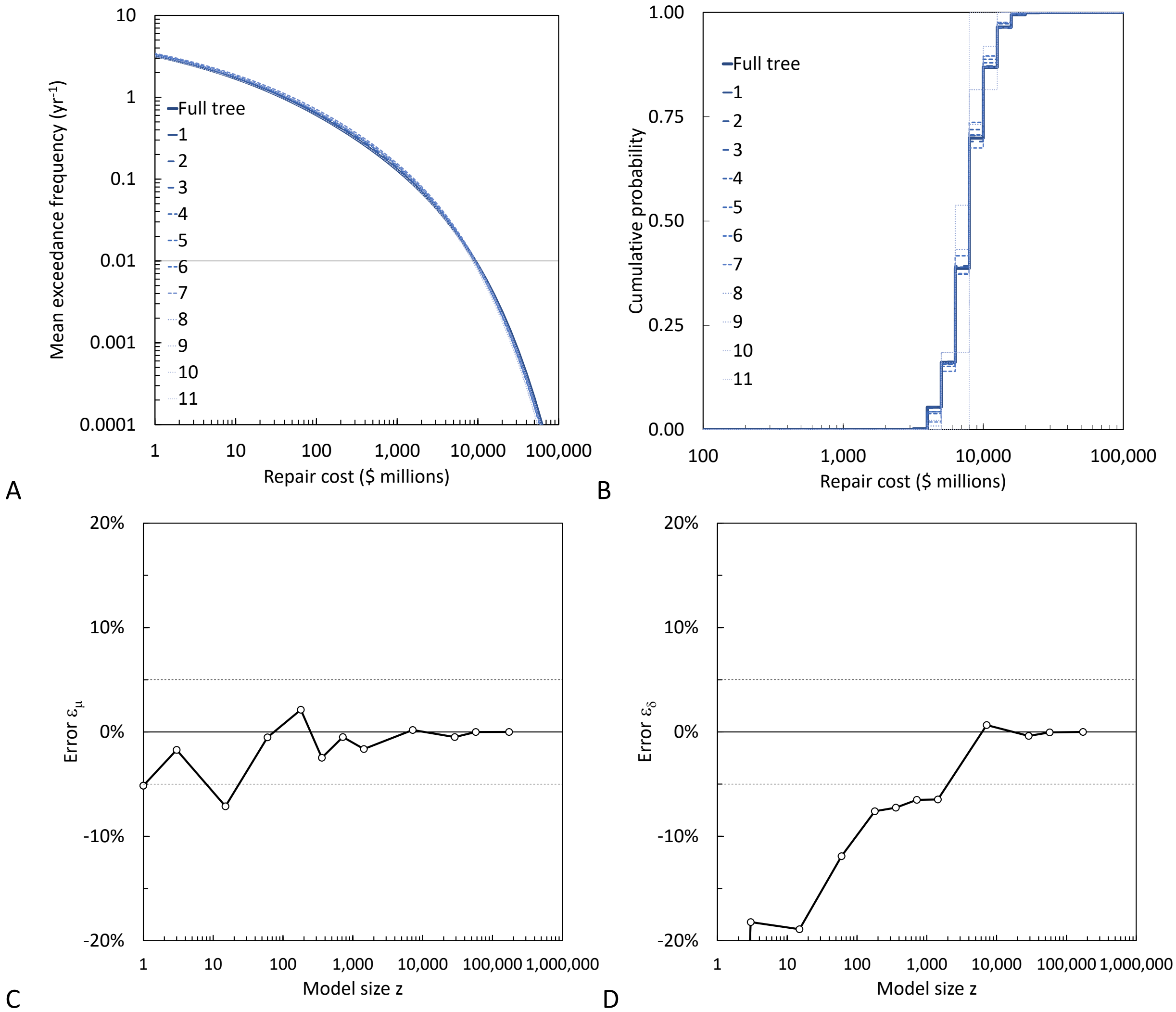


Figure 4. UCERF3-TD model trimmed to 100-year loss for CEA-proxy portfolio (A) loss exceedance curve; (B) cumulative distribution function of 100-year loss; (c) mean error ε_μ versus model size z; (D) coefficient of variation error ε_δ versus model size z

	Trimmed branch	z	μ_{Lp}	δ_{Lp}	D_n	D_{nmax}	ε_μ	ε_δ	Pass
0	Full Tree = N/A	172,800	$15,144	0.34					
1	MMax Off Fault = 7.6	57,600	$15,145	0.34	0.000	0.008	0%	0%	TRUE
2	Fault Model = Fault Model 3.2	28,800	$15,255	0.34	0.011	0.010	1%	0%	FALSE
3	Vs30 Model = Wills et al. (2015)	14,400	$15,099	0.34	0.004	0.014	0%	0%	TRUE
4	ERF Probability Model = Mid COV Values	3,600	$15,015	0.34	0.019	0.027	-1%	0%	TRUE
5	Slip Along Rupture Model (Dsr) = Uniform	1,800	$15,313	0.34	0.008	0.039	1%	1%	TRUE
6	Total Mag 5 Rate = 7.9	600	$15,022	0.33	0.016	0.067	-1%	-1%	TRUE
7	Scaling Relationship = Hanks & Bakun (2008)	120	$14,937	0.30	0.022	0.149	-1%	-10%	FALSE
8	Deformation Model = Neokinema	30	$14,726	0.30	0.055	0.298	-3%	-11%	FALSE
9	Spatial Seismicity PDF = UCERF3	15	$15,295	0.30	0.110	0.421	1%	-12%	FALSE
10	Ground Motion Model = Boore, Stewart, Seyhan & Atkinson (2014)	3	$14,033	0.27	0.167	0.941	-7%	-21%	FALSE
11	GMM Additional Epistemic Uncertainty = None	1	$13,575	0.00	0.352	1.630	-10%	-100%	FALSE

Table 4. Trimming path for 250-year loss (p = 0.004)

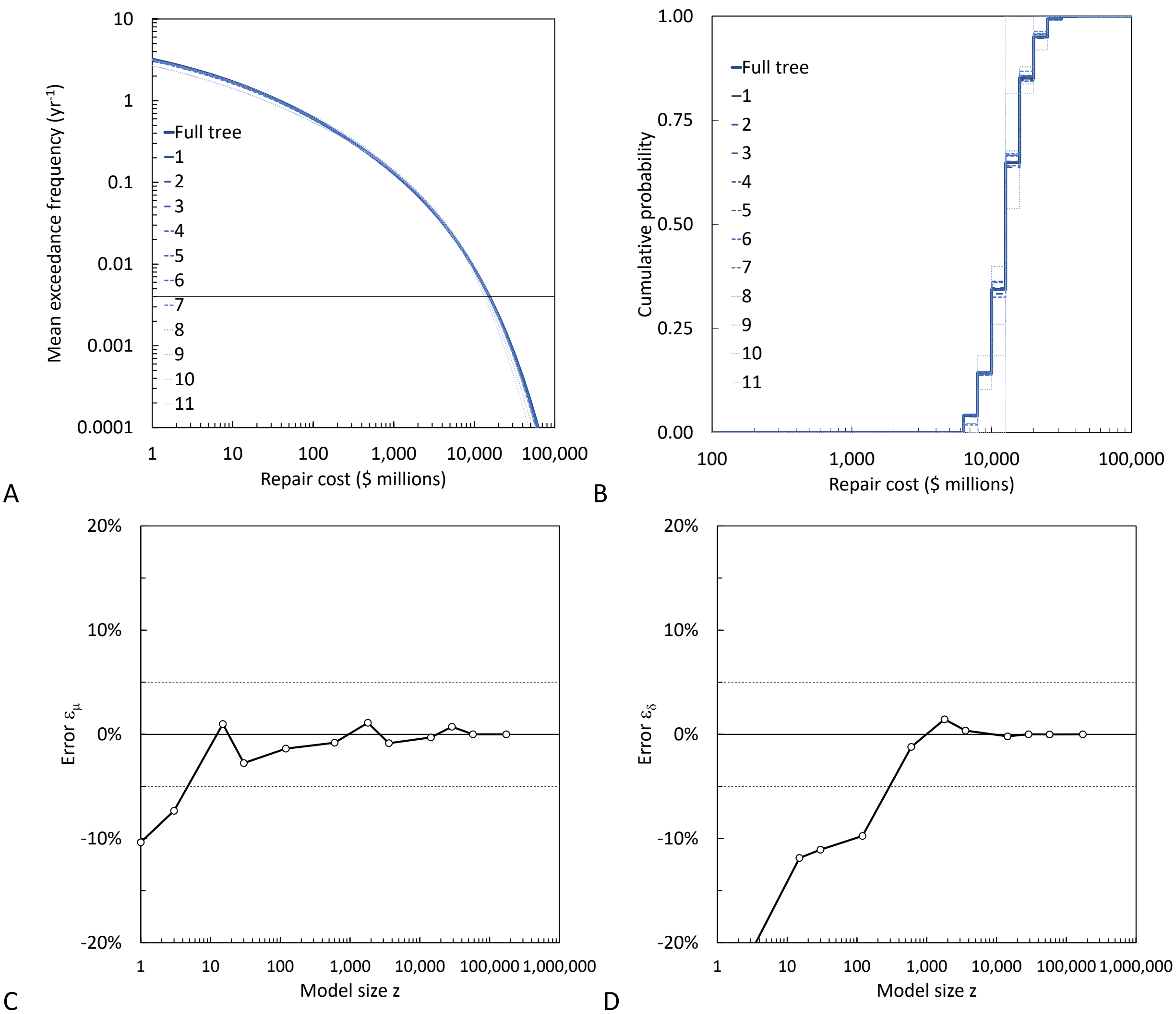


Figure 5. UCERF3-TD model trimmed to 250-year loss for CEA-proxy portfolio (A) loss exceedance curve; (B) cumulative distribution function of 250-year loss; (c) mean error ε_μ versus model size z; (D) coefficient of variation error ε_δ versus model size z

	Trimmed branch	z	μ_{Lp}	δ_{Lp}	D_n	D_{nmax}	ε_μ	ε_δ	Pass
0	Full Tree = N/A	172,800	$18,820	0.34					
1	MMax Off Fault = 7.6	57,600	$18,822	0.34	0.000	0.008	0%	0%	TRUE
2	Scaling Relationship = EllB M(A) & Shaw12 Sqrt Length D(L)	11,520	$18,766	0.34	0.013	0.016	0%	1%	TRUE
3	Fault Model = Fault Model 3.1	5,760	$18,640	0.34	0.012	0.022	-1%	1%	TRUE
4	Vs30 Model = Wald & Allen (2007)	2,880	$18,827	0.34	0.015	0.031	0%	1%	TRUE
5	Total Mag 5 Rate = 7.9	960	$18,561	0.34	0.023	0.053	-1%	0%	TRUE
6	Slip Along Rupture Model (Dsr) = Uniform	480	$18,959	0.34	0.028	0.075	1%	0%	TRUE
7	Spatial Seismicity PDF = UCERF2	240	$18,288	0.34	0.020	0.105	-3%	0%	TRUE
8	Deformation Model = Neokinema	60	$17,640	0.33	0.058	0.210	-6%	-2%	FALSE
9	ERF Probability Model = High COV Values	15	$18,759	0.33	0.119	0.421	0%	-3%	TRUE
10	Ground Motion Model = Boore, Stewart, Seyhan & Atkinson (2014)	3	$16,840	0.28	0.184	0.941	-11%	-16%	FALSE
11	GMM Additional Epistemic Uncertainty = None	1	$16,227	0.00	0.369	1.630	-14%	-100%	FALSE

Table 5. Trimming path for 400-year loss (p = 0.0025)

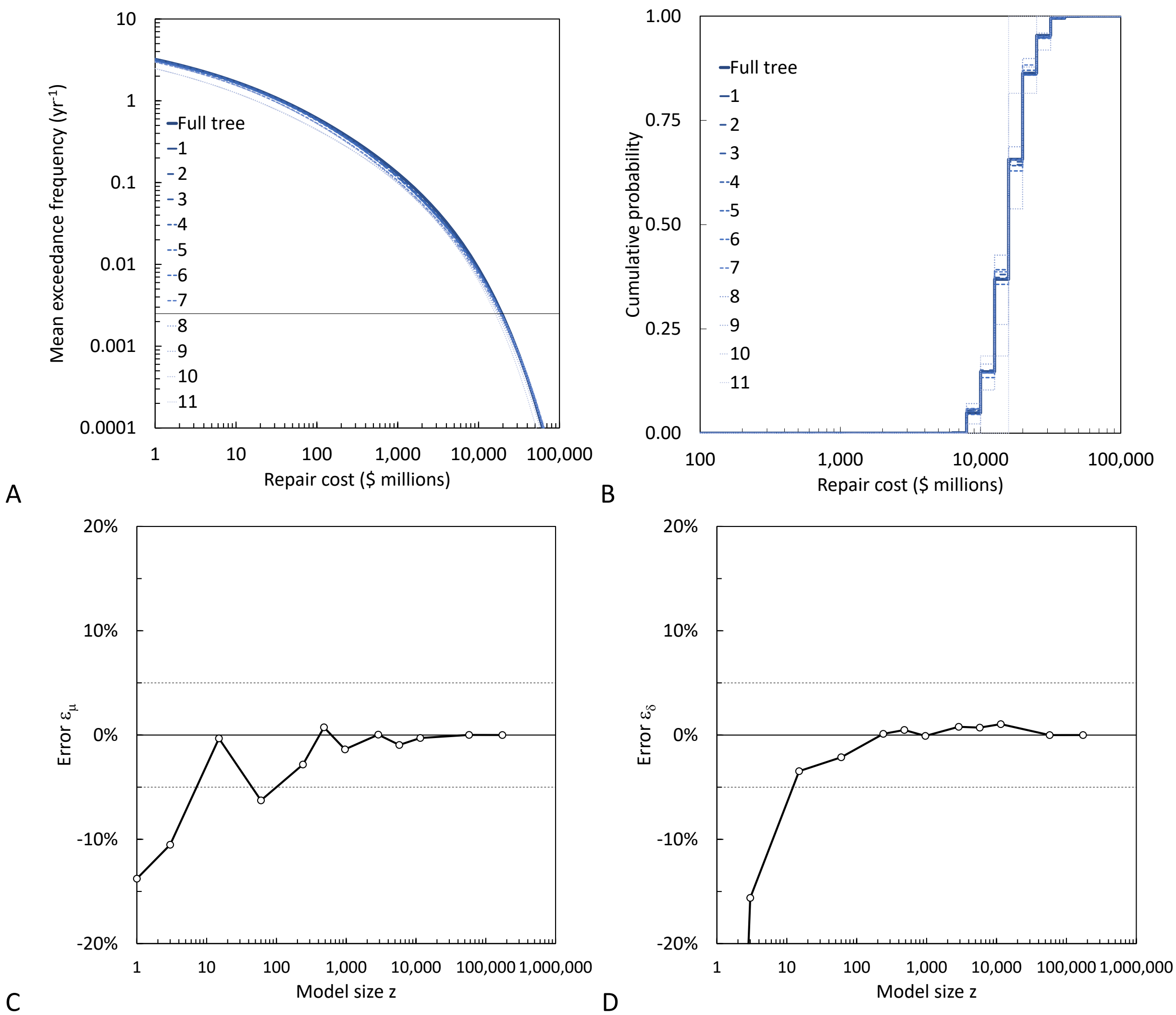


Figure 6. UCERF3-TD model trimmed to 400-year loss for CEA-proxy portfolio (A) loss exceedance curve; (B) cumulative distribution function of 400-year loss; (c) mean error ε_μ versus model size z; (D) coefficient of variation error ε_δ versus model size z

This information is preliminary or provisional and is subject to revision. It is being provided to meet the need for timely best science. The information has not received final approval by the U.S. Geological Survey (USGS) and is provided on the condition that neither the USGS nor the U.S. Government shall be held liable for any damages resulting from the authorized or unauthorized use of the information.

	Trimmed branch	z	μ_{Lp}	δ_{Lp}	D_n	D_{nmax}	ε_μ	ε_δ	Pass
0	Full Tree = N/A	172,800	$21,672	0.34					
1	MMax Off Fault = 7.6	57,600	$21,675	0.34	0.000	0.008	0%	0%	TRUE
2	ERF Probability Model = Mid COV Values	14,400	$21,381	0.34	0.009	0.014	-1%	0%	TRUE
3	Fault Model = Fault Model 3.2	7,200	$21,571	0.33	0.004	0.020	0%	0%	TRUE
4	Vs30 Model = Wills et al. (2015)	3,600	$21,375	0.33	0.010	0.027	-1%	0%	TRUE
5	Slip Along Rupture Model (Dsr) = Uniform	1,800	$21,788	0.34	0.019	0.039	1%	0%	TRUE
6	Scaling Relationship = Shaw (2009) Modified	360	$21,396	0.32	0.007	0.086	-1%	-5%	TRUE
7	Total Mag 5 Rate = 7.9	120	$21,264	0.31	0.012	0.149	-2%	-6%	FALSE
8	Spatial Seismicity PDF = UCERF2	60	$20,545	0.31	0.045	0.210	-5%	-8%	FALSE
9	Deformation Model = Geologic	15	$22,216	0.32	0.080	0.421	3%	-5%	TRUE
10	GMM Additional Epistemic Uncertainty = None	5	$21,436	0.15	0.202	0.729	-1%	-56%	FALSE
11	Ground Motion Model = Boore, Stewart, Seyhan & Atkinson (2014)	1	$19,392	0.00	0.493	1.630	-11%	-100%	FALSE

Table 6. Trimming path for 550-year loss (p = 0.0018)

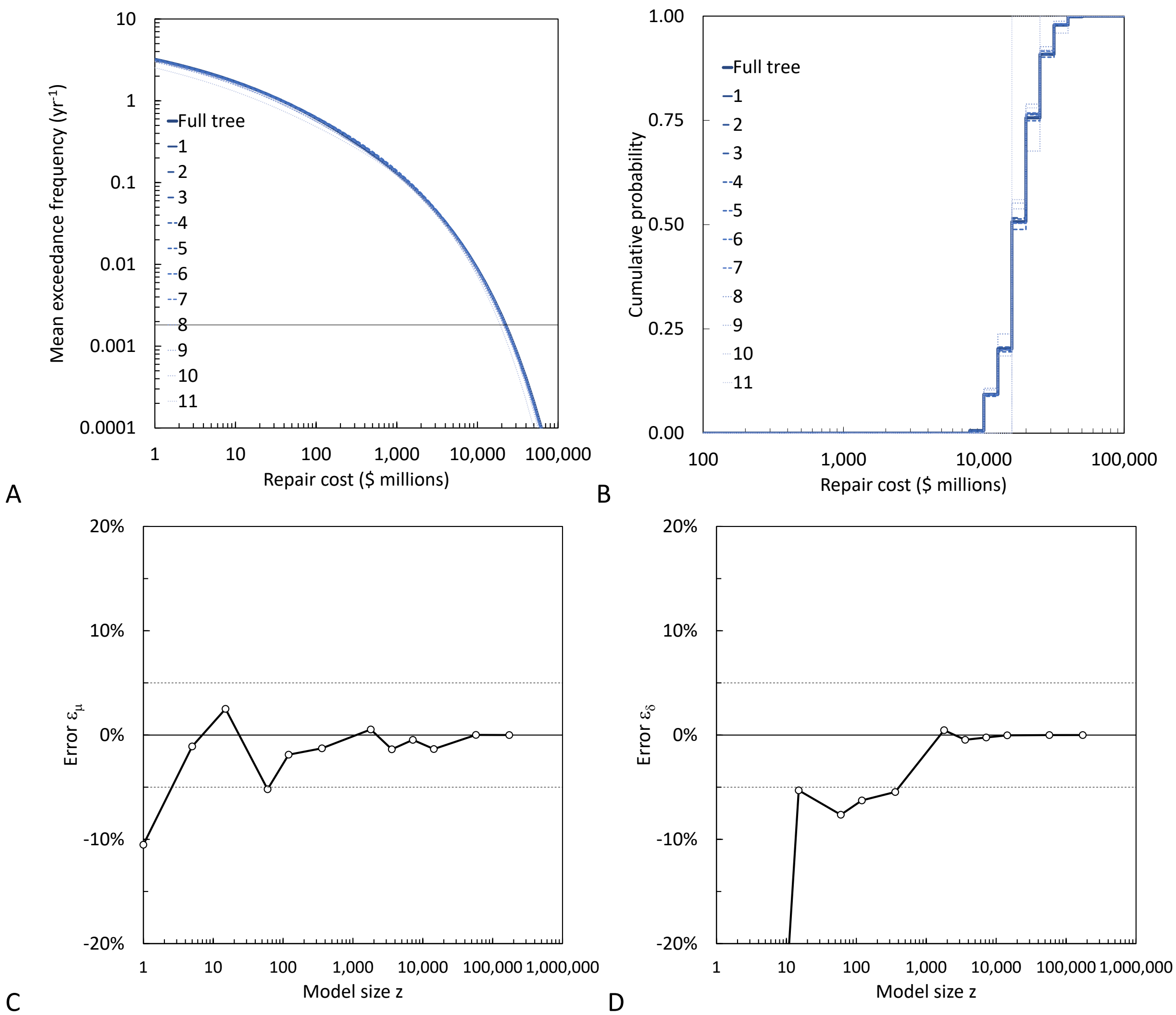


Figure 7. UCERF3-TD model trimmed to 550-year loss for CEA-proxy portfolio (A) loss exceedance curve; (B) cumulative distribution function of 550-year loss; (c) mean error ε_μ versus model size z; (D) coefficient of variation error ε_δ versus model size z

	Trimmed branch	z	μ_{Lp}	δ_{Lp}	D_n	D_{nmax}	ε_μ	ε_δ	Pass
0	Full Tree = N/A	172,800	$37,465	0.33					
1	MMax Off Fault = 7.6	57,600	$37,470	0.33	0.000	0.008	0%	0%	TRUE
2	Total Mag 5 Rate = 7.9	19,200	$37,417	0.33	0.006	0.012	0%	0%	TRUE
3	Fault Model = Fault Model 3.2	9,600	$37,731	0.33	0.009	0.017	1%	0%	TRUE
4	Vs30 Model = Wills et al. (2015)	4,800	$37,405	0.33	0.005	0.024	0%	0%	TRUE
5	Slip Along Rupture Model (Dsr) = Uniform	2,400	$37,964	0.33	0.017	0.034	1%	0%	TRUE
6	Deformation Model = Neokinema	600	$37,257	0.32	0.013	0.067	-1%	-3%	TRUE
7	Scaling Relationship = Shaw (2009) Modified	120	$36,485	0.30	0.025	0.149	-3%	-7%	FALSE
8	Spatial Seismicity PDF = UCERF2	60	$35,773	0.31	0.037	0.210	-5%	-6%	FALSE
9	ERF Probability Model = High COV Values	15	$37,554	0.31	0.117	0.421	0%	-6%	FALSE
10	Ground Motion Model = Boore, Stewart, Seyhan & Atkinson (2014)	3	$34,062	0.27	0.184	0.941	-9%	-19%	FALSE
11	GMM Additional Epistemic Uncertainty = None	1	$32,991	0.00	0.369	1.630	-12%	-100%	FALSE

Table 7. Trimming path for 2,500-year loss (p = 0.0004)

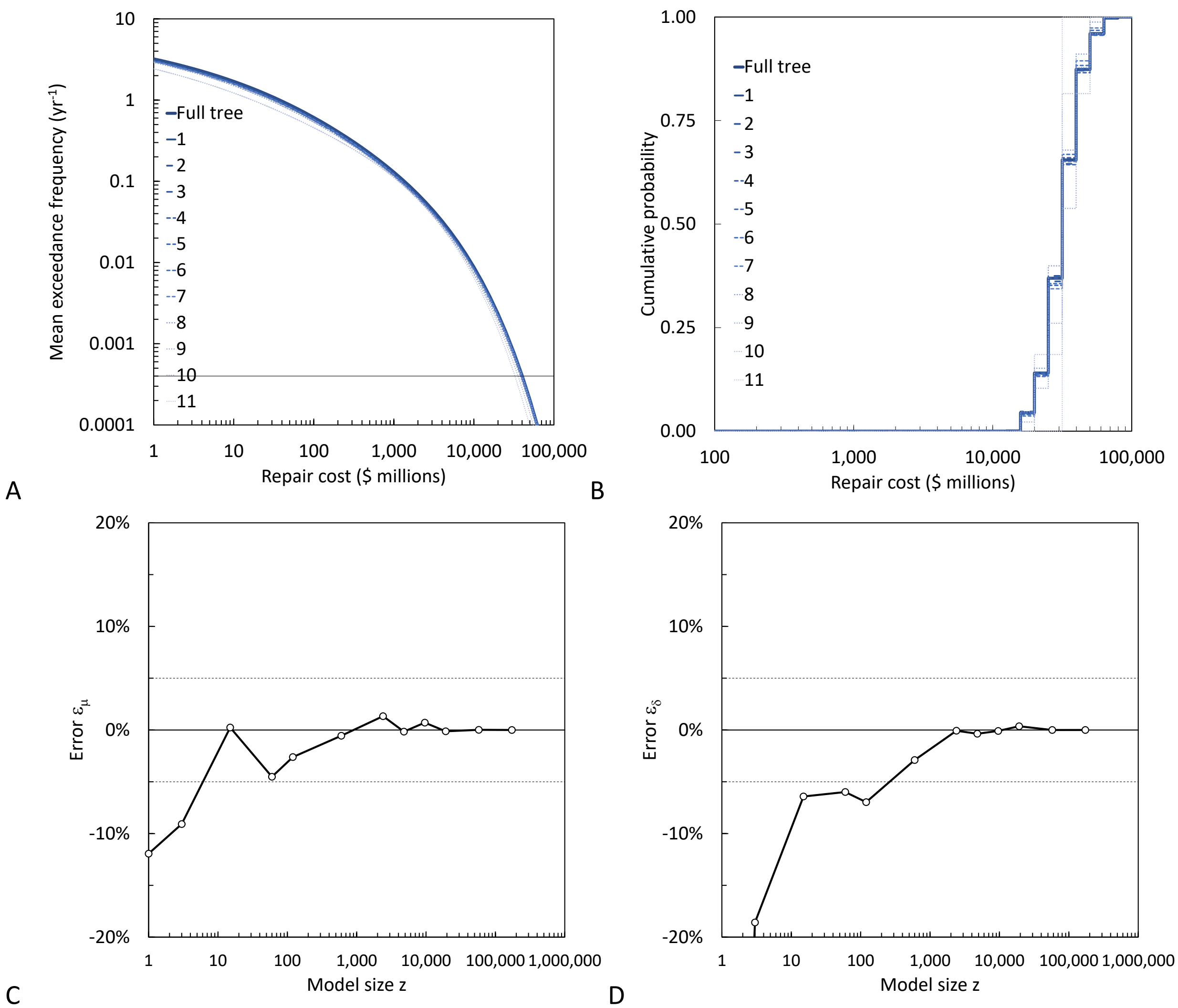


Figure 8. UCERF3-TD model trimmed to 2500-year loss for CEA-proxy portfolio (A) loss exceedance curve; (B) cumulative distribution function of 2500-year loss; (c) mean error ε_μ versus model size z; (D) coefficient of variation error ε_δ versus model size z

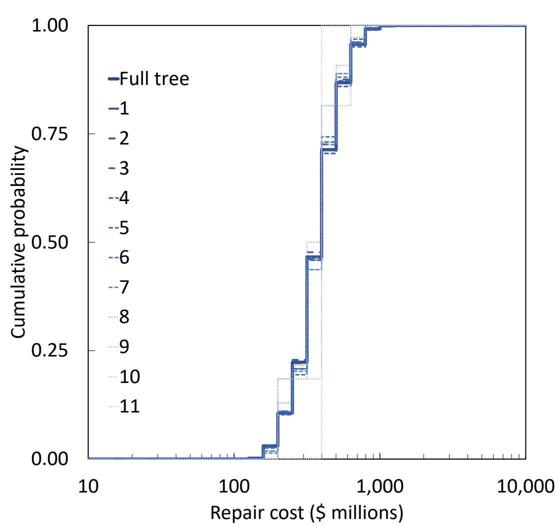
4.3 Results with expected annualized loss (*EAL*) as the dependent variable

Table 8 shows the UCERF3-TD extended logic tree trimmed to maintain the distribution of expected annualized loss (*EAL*). The table shows that, just as with the loss exceedance curves, the distribution of *EAL* is most sensitive to GMM additional epistemic uncertainty, but interestingly, not the ground motion model, which can be trimmed and set to that of Abrahamson et al. (2014). The model can be reduced to 720 leaves out of the 172,800 total (4% of its original size) and retain the probability distribution of *EAL*. If the user can tolerate an 8% underestimate of the coefficient of variation, the model can be trimmed to just 18 leaves, varying total magnitude 5 rate, spatial seismicity probability density function, and ground motion model addition epistemic uncertainty.

Figure 9A shows that the cumulative distribution functions of the full and reduced-order models of *EAL* are virtually indistinguishable until the last three branches are trimmed. Figure 9B and Figure 9C show that error in the expected value and coefficient of variation of *EAL* remain within 10% bounds until $z <$ 18 branches. As with earlier, similar plots, only the dots in these two figures are meaningful; the lines connecting them just make it easier to see the trend of greater agreement with greater model size.

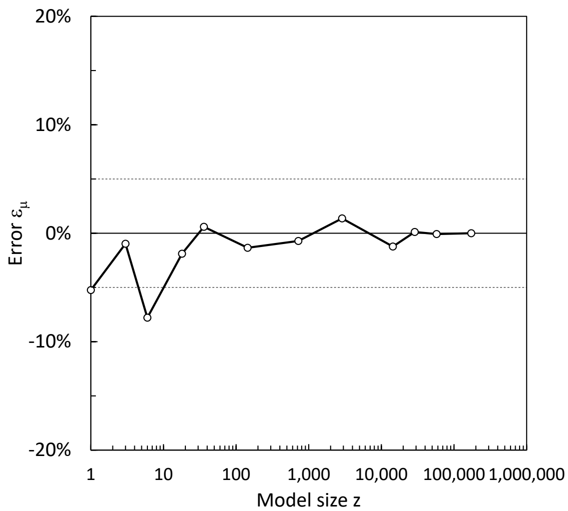
	Trimmed branch	z	μ_{EAL}	δ_{EAL}	D_n	D_{nmax}	ε_μ	ε_δ	Pass
0	Full Tree = N/A	172,800	$441	0.39					
1	MMax Off Fault = 7.6	57,600	$441	0.39	0.001	0.008	0%	0%	TRUE
2	Fault Model = Fault Model 3.1	28,800	$442	0.39	0.004	0.010	0%	0%	TRUE
3	Slip Along Rupture Model (Dsr) = Tapered Ends	14,400	$436	0.38	0.012	0.014	-1%	-1%	TRUE
4	Ground Motion Model = Abrahamson, Silva & Kamai (2014)	2,880	$447	0.39	0.015	0.031	1%	1%	TRUE
5	Deformation Model = Zeng B-Fault Bounded	720	$438	0.37	0.018	0.061	-1%	-4%	TRUE
6	Scaling Relationship = Shaw (2009) Modified	144	$436	0.36	0.030	0.136	-1%	-8%	FALSE
7	ERF Probability Model = Mid COV Values	36	$444	0.36	0.029	0.272	1%	-8%	FALSE
8	Vs30 Model = Wills et al. (2015)	18	$433	0.36	0.029	0.384	-2%	-8%	FALSE
9	Total Mag 5 Rate = 7.9	6	$407	0.32	0.102	0.665	-8%	-17%	FALSE
10	Spatial Seismicity PDF = UCERF3	3	$437	0.31	0.281	0.941	-1%	-19%	FALSE
11	GMM Additional Epistemic Uncertainty = None	1	$418	0.00	0.466	1.630	-5%	-100%	FALSE

Table 8. Trimming path for expected annualized loss EAL

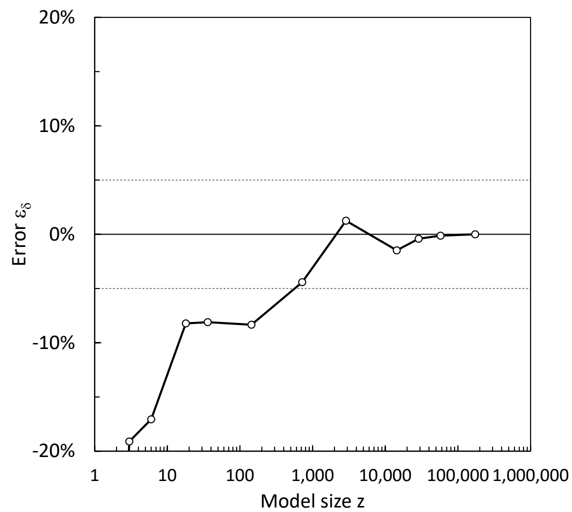


A

This information is preliminary or provisional and is subject to revision. It is being provided to meet the need for timely best science. The information has not received final approval by the U.S. Geological Survey (USGS) and is provided on the condition that neither the USGS nor the U.S. Government shall be held liable for any damages resulting from the authorized or unauthorized use of the information.



B



C

Figure 9. UCERF3-TD model trimmed to expected annualized loss for CEA-proxy portfolio (A) cumulative distribution function of EAL; (B) mean error ε_μ versus model size z; (C) coefficient of variation error ε_δ versus model size z

5. Summary and conclusions

5.1 Summary

The Uniform California Earthquake Rupture Forecast, Time Dependent, version 3 is sometimes depicted as a logic tree with eight modeling uncertainties, also called epistemic uncertainties, that one can think of as independent variables in a model of earthquake risk. To calculate risk can also require adding three independent variables to the logic tree: the selection of Vs30 model, the selection of ground motion model, and the degree of ground motion model additional epistemic uncertainty, for a total of 11 independent variables. To evaluate earthquake loss to a building portfolio in every allowable combination of the 11 variables requires calculating loss for 172,800 combinations or leaves in the logic tree and each of 6 million earthquake ruptures—a computationally demanding task even for a supercomputer. (We leveraged the fact that some branches only affect earthquake occurrence rates rather than ground motion fields and portfolio losses, but still the effort is huge.)

The present study seeks to reduce that computational burden by identifying a reduced-order model using a subset of the 11 independent variables that reproduces the probability distribution of an important dependent variable. We considered six dependent variables related to the building repair cost for a statewide portfolio of buildings that approximates that of the California Earthquake Authority's insurance portfolio of insured single-family dwellings. The dependent variables are the total repair cost in a single earthquake with each of five exceedance probabilities, plus expected annualized loss.

We applied a recently developed model-order reduction technique that starts by evaluating the probability distribution of the dependent variable for the full model, then stepwise trims one independent variable at a time, setting its value to one of its possible values, and testing whether the probability distribution of the dependent variable significantly changes, or its mean and coefficient of variation significantly changes relative to those of the full model. The reduced-order model with the smallest change is preferred, and the process iterates until one reaches the smallest possible model that preserves the probability distribution of the dependent variable to the satisfaction of the two-sample Kolmogorov-Smirnov goodness-of-fit test at the 1% significance level and the dependent variables' mean and coefficient of variation within ±5%.

The model order reduction technique allows one to trim the UCERF3-TD model by as few as three and as many as nine of its 11 independent variables. But computational effort scales exponentially with the number of independent variables. If one independent variable out of the 11 has five possible values, eliminating it reduces the number of logic-tree leaves by a factor of 5—an 80% reduction in computational effort for a 9% reduction in the number of variables.

Table 9 summarizes the degree of success of the model-order reduction technique. The columns refer to the dependent variables for which the model was trimmed: the repair cost in one year associated with each of five probability levels, and expected annualized loss. The rows show the size of the full model in terms of number of independent variables and logic-tree leaves, the same quantities for the smallest reduced-order model that passes goodness-of-fit and error tests, and the ratio of the latter to the former.

Table 9. Summary of the degree of model order reduction for the dependent variables considered here

Model size		Repair cost L_p with exceedance probability p =					EAL
		1/100	1/250	1/400	1/550	1/2500	
Full model	Independent variables	11	11	11	11	11	11
	Logic-tree leaves	172,800	172,800	172,800	172,800	172,800	172,800
Reduced order	Independent variables	8	5	2	2	5	6
	Logic-tree leaves	7,200	600	15	15	600	720
Reduced ÷ full	Independent variables	73%	45%	18%	18%	45%	55%
	Logic-tree leaves	4%	0.3%	0.009%	0.009%	0.3%	0.4%

At the annual loss-exceedance probabilities commonly used by earthquake insurers in general and the California Earthquake Authority in particular (1/250 to 1/550), the model allows one to trim between six and nine of UCERF3-TD's 11 independent variables, reducing the number of logic-tree leaves between 99.7% and 99.991%. A hypothetical risk calculation that takes 24 hours for the full model (evaluating all 172,800 leaves) can be reduced to a calculation that takes 8 seconds (for a reduced-order model with 15 leaves).

Some variables always strongly influence the dependent variable, some one or two, some three or four, and some all five cases. Table 10 recaps this story, listing independent variables from least to most important, in the sense that the least important can be trimmed from all models without affecting the probability distribution of the dependent variable. The maximum off-fault earthquake magnitude can be set to 7.6 in all cases. The fault model can also be fixed in all cases, but the preferred value is FM3.1 in some cases and FM3.2 in others. One variable, ground motion model additional epistemic uncertainty, cannot be trimmed from the logic tree without greatly disturbing the probability distribution of any of the dependent variables. Others can be trimmed from the logic tree in some but not all cases. Where a variable can be trimmed, the table shows the value to which it can be set without greatly disturbing the probability distribution of the dependent variable.

Table 10. Variables that can be trimmed from the logic tree and set to a deterministic value

Variable	Preferred value of trimmed variable for exceedance probability p =					EAL
	1/100	1/250	1/400	1/550	1/2500	
MMax Off Fault	7.6	7.6	7.6	7.6	7.6	7.6
Fault Model	3.1	3.2	3.1	3.2	3.2	3.1
Total Mag 5 Rate		7.9	7.9	7.9	7.9	7.9
ERF Probability Model	Mid COV	Mid COV	High COV	Mid COV		
Vs30 Model		W2015	WA2008	W2015	W2015	
Slip Along Rupt Mod (Dsr)		Uniform	Uniform	Uniform	Uniform	
Deformation Model			Neokinema	Avg Block	Neokinema	ZengBB
Scaling Relationship			ELL B SQL	Shaw 09m		
Spatial Seismicity PDF			UCERF2	UCERF2		
Ground Motion Model						ASK2014
GMM Added Epist Uncertainty						

Table 11 recaps the order of variable trimmed for each of the dependent variables, sorted by the order for the 550-year loss. The order shows some consistency between loss-exceedance cases, with the least influential variables (the ones trimmed earliest) generally including MMax off fault, ERF probability model, fault model, and Vs30 model. The most influential variables are generally the spatial seismicity probability density function, deformation model, and the two variables related to ground motion models. The slip along rupture, scaling relationship, and total magnitude-5 rate generally get trimmed toward the middle of the model-order-reduction process. The order for *EAL* looks very different.

Table 11. Recap order of trimmed variables

	Variable	Order of trimmed variable for exceedance probability p =					Ditto, EAL
		1/100	1/250	1/400	1/550	1/2500	
← More *influence* Less →	MMax Off Fault	1	1	1	1	1	1
	ERF Probability Model	3	4	9	2	9	7
	Fault Model	2	2	3	3	3	2
	Vs30 Model	5	3	4	4	4	8
	Slip Along Rupture Model (Dsr)	6	5	6	5	5	3
	Scaling Relationship	4	7	2	6	7	6
	Total Mag 5 Rate	8	6	5	7	2	9
	Spatial Seismicity PDF	7	9	7	8	8	10
	Deformation Model	9	8	8	9	6	5
	GMM Additional Epistemic Uncertainty	11	11	11	10	11	11
	Ground Motion Model	10	10	10	11	10	4

5.2 Conclusions

(1) This model order reduction technique can handle a model that produces a scalar dependent variable that is a function of both scalar and nominal independent variables. It resembles a probabilistic sensitivity test that allows for interaction between independent variables.

(2) This is the first time this technique was applied to large rare losses (points on the loss exceedance curve) in a large building portfolio. An earlier application of the technique only examined expected

annualized loss. The technique worked as expected. That should be unsurprising since a value on the loss exceedance curve (denoted here by L_p) is not fundamentally different from expected annualized loss (*EAL*) in that like *EAL*, L_p is merely a scalar function of a suite of the same scalar and nominal independent variables.

(3) The technique reduced the loss model from 172,800 leaves to 15 leaves in the cases of the 400- and 550-year repair cost, or to 720 leaves in the case of EAL.

(4) Which independent variables can be trimmed depends on the choice of dependent variable. The preferred value of those trimmed independent variables can also depend on which dependent variable one cares about. Only two variables cannot be trimmed from the logic tree for any of the dependent variables considered here: ground-motion-model additional epistemic uncertainty and ground motion model. Greater study of those two uncertainties might reduce them. Doing so would thin the upper tail of the loss distribution, save insurers on reinsurance costs, and indirectly save policyholders on premium costs that help pay for reinsurance.

5.3 Limitations and open questions

(1) The technique was applied only to a single deterministic statewide portfolio that approximates that of the California Earthquake Authority. We have not demonstrated that the same trimmed model will also reasonably approximate *EAL* or the loss exceedance curve for other insurance portfolios. Smaller regional portfolios might be more strongly affected by uncertainties that have a spatially concentrated effect, such as the difference between the two fault models. Such a test would be straightforward but of limited interest to the California Earthquake Authority.

(2) We did not account for uncertainty in the seismic vulnerability functions that relate loss to ground motion given model building type and replacement cost.

(3) We did not treat uncertainty in the assignment of model building type to individual assets, or uncertainty in asset replacement cost. Both *EAL* and L_p would scale linearly with an across-the-board under- or over-estimation of asset replacement cost, but the uncertainty might not work that way.

(4) We did not consider the effects of spatiotemporal clustering (e.g., large damaging aftershocks), which can have a larger influence on expected annual losses than all the uncertainties considered here.

6. References cited

Abrahamson, NA, Silva, WJ, and Kamai, R (2014). Summary of the ASK14 ground motion relation for active crustal regions. *Earthquake Spectra* 30 (3): 1025-1055

Boore, DM, Stewart, JP, Seyhan, E, and Atkinson, GM (2014). NGA-West2 equations for predicting PGA, PGV, and 5% damped PSA for shallow crustal earthquakes. *Earthquake Spectra* 30 (3): 1057-1085.

Chiou, BSJ, and Youngs, RR (2014). Update of the Chiou and Youngs NGA model for the average horizontal component of peak ground motion and response spectra. *Earthquake Spectra*, 30(3): 1117-1153.

Field, EH, Arrowsmith, RJ, Biasi, GP, Bird, P, Dawson, TE, Felzer, KR, Jackson, DD, Johnson, KM, Jordan, TH, Madden, C, Michael, AJ, Milner, KR, Page, MT, Parsons, T, Powers, PM, Shaw, BE, Thatcher, WR, Weldon II, RJ, and Zeng, Y (2014). Uniform California earthquake rupture forecast, version 3 (UCERF3)—the time-independent model. *Bulletin of the Seismological Society of America*, 104 (3): 1122-1180, https://doi.org/10.1785/0120130164 [accessed July 13, 2020].

Field, EH, GP Biasi, GP, Bird, P, Dawson, TE, Felzer, KR, Jackson, DD, Johnson, KM, Jordan, TH, Madden, C, Michael, AJ, Milner, KR, Page, MT, Parsons, T, Powers, PM, Shaw, BE, Thatcher, WR, Weldon II, RJ, and Zeng, Y (2015). Long-term time-dependent probabilities for the third uniform California earthquake rupture forecast (UCERF3). *Bulletin of the Seismological Society of America*, 105 (2A): 1-33, April 2015, doi: 10.1785/0120140093

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

Perkins, D, and Taylor, C (2003). Earthquake occurrence modeling for evaluating seismic risks to roadway systems. *Sixth U.S. Conference and Workshop on Lifeline Earthquake Engineering (TCLEE) 2003, August 10-13, 2003, Long Beach, California, United States*. https://ascelibrary.org/doi/abs/10.1061/40687(2003)87 [accessed April 12, 2020]

Porter, K, Field, E, and Milner, K (2017). Trimming a hazard logic tree with a new model-order-reduction technique. *Earthquake Spectra*, 33(3): 857–874. https://doi.org/10.1193/092616eqs158m [accessed April 12, 2020]

Porter, KA (2009a). Cracking an open safe: HAZUS vulnerability functions in terms of structure-independent spectral acceleration. *Earthquake Spectra* 25 (2): 361-378, http://www.sparisk.com/pubs/Porter-2009-Safecrack-Casualty.pdf

Porter, KA (2009b). Cracking an open safe: more HAZUS vulnerability functions in terms of structure-independent spectral acceleration. *Earthquake Spectra* 25 (3): 607-618, http://www.sparisk.com/pubs/Porter-2009-Safecrack-MDF.pdf

Porter, KA (2010). Cracking an open safe: uncertainty in HAZUS-based seismic vulnerability functions. *Earthquake Spectra*, 26 (3): 893-900 http://www.sparisk.com/pubs/Porter-2010-Safecrack-COV.pdf

Porter, KA, Field, EH, and Milner, K (2012). Trimming the UCERF2 hazard logic tree. *Seismological Research Letters*, 83 (5): 815-828, https://www.sparisk.com/pubs/Porter-2012-SRL-Tree-trim.pdf [accessed April 12, 2020]

Schilders, WH, Van der Vorst, HA, and Rommes, J (2008). Model Order Reduction: Theory, Research Aspects and Applications. Berlin: Springer, 471 p.

Shaw, BE (2009). Constant stress drop from small to great earthquakes in magnitude-area scaling, *Bulletin of the Seismological Society of America*, 99: 871, doi:10.1785/0120080006.

Wald, DJ, and Allen, TI (2007). Topographic slope as a proxy for seismic site conditions and amplification. *Bulletin of the Seismological Society of America*, 97: 1379-1395

Wills, CJ, Gutierrez, CI, Perez, FG, and Branum, DM (2015). A next generation VS30 map for California based on geology and topography. *Bulletin of the Seismological Society of America* 105 (6): 3083–3091

Appendix 1. Lognormally distributed portfolio loss

Propagating uncertainty in between-event ground motion variability and spatially correlated within-even ground motion variability can involve hundreds of realizations of e_τ and e_ϕ for each logic-tree leaf and rupture. We assumed that 5 realizations of e_τ and 100 realizations of e_ϕ would suffice to approximate the distribution of portfolio loss. The distribution would possibly converge using a smaller number of realizations, but we do not know what that number is. Anyhow, the computational effort of evaluating each rupture 500 times for each branch of the UCERF3-TD logic tree was too demanding, so we sought an approximation.

It is comparatively easy to estimate the mean value of portfolio loss for a given rupture and logic-tree leaf using equation (2). It takes advantage of the fact that the expected value of a sum equals the sum of the expected values. The portfolio loss is taken as the sum of the losses to the individual assets. Even if the asset losses are correlated because of τ and ϕ, the expected value of portfolio loss is the sum of the expected values of loss to the individual assets, correlation associated with τ and ϕ notwithstanding.

We therefore seek to estimate the distribution of portfolio loss conditioned on the mean portfolio loss. We do so with a large sample of portfolio losses, as follows.

We examined many ranges of loss, referred to here as loss bins, logarithmically equally spaced from $1 million to over $10 billion. For each combination of loss bin *i*, ground motion model *b*, ground motion model added epistemic uncertainty *c*, and Vs30 model *d*, we found the rupture *k* with the largest occurrence rate r_k. Let us refer to that rupture as the modal rupture.

For each such modal rupture, we evaluate 500 ground motion fields, one for each combination of the five samples of e_τ and the 100 samples of e_ϕ. We evaluate the portfolio loss conditioned on the resulting ground motion field using equation (25) and the portfolio loss using equation (26):

$$L = \sum_a V_a y_a \left(x_o \right) \tag{26}$$

where *a* denotes an index to portfolio assets, V_a is the replacement cost new of asset a, $y_a(x_o)$ denotes the repair cost as a fraction of replacement cost new for asset *a*, and x_o is the ground motion at the location of asset *a*. The mean and coefficient of variation of the 500 sample of portfolio loss are calculated for each modal rupture and each combination of {*i, b, c, d*} and a curve fit to the data as shown in Figure 10. Weights $w_{e\tau}$ are shown in Table 2. Weights $w_{e\phi}$ are all 0.01.

$$\mu_{L|k} = \sum_{e_\tau} \sum_{e_\phi} L \cdot w_{e_\tau} \cdot w_{e_\phi} \tag{27}$$

$$\sigma^2_{L|k} = \left(\sum_{e_\tau} \sum_{e_\phi} L^2 \cdot w_{e_\tau} \cdot w_{e_\phi} \right) - \mu^2_{L|k} \tag{28}$$

$$\delta_{L|k} = \frac{\sigma_{L|k}}{\mu_{L|k}} \tag{29}$$

The relationship between $\delta_{L|k}$ and $\mu_{L|k}$ shows some structure: there appear to be upper and lower branches. They seem to result from the different ground motion models. Data on the upper branch tend to come from using the Abrahamson et al. (2014) and Chiou and Youngs (2014) ground-motion-prediction equations, although not exclusively, and some of the data in the lower branch also use those ground-motion-prediction equations. We checked that the mean portfolio losses calculated by summing mean asset losses ("sum of means") with equation (2) equals the average portfolio loss among the 500 simulations of between- and within-event ground motion variability ("mean of sum") calculated by equation (27), as shown in Figure 11.

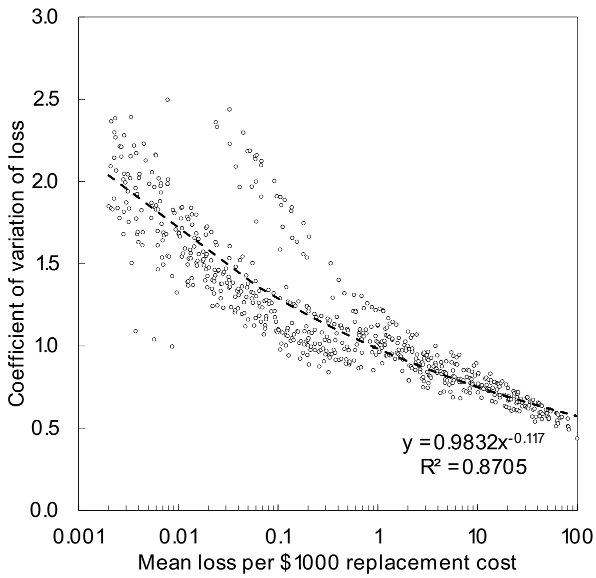


Figure 10. Coefficient of variation of portfolio loss δ_L as a function of mean loss μ_L.

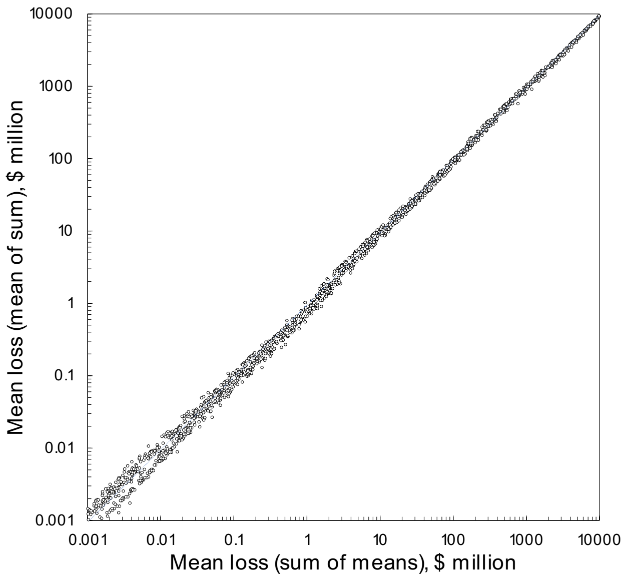


Figure 11. Checking that mean portfolio loss using simulated ground motion fields equals mean portfolio loss summing over mean asset losses.

We also checked that portfolio loss tends to be lognormally distributed, considering all 24 combinations of ground-motion-prediction equation, added epistemic uncertainty, and Vs30 model for the modal rupture in the $10 billion loss bin. All 24 samples passed a Lilliefors goodness of fit test at the 5% significant level. Figure 12 illustrates two of these checks.

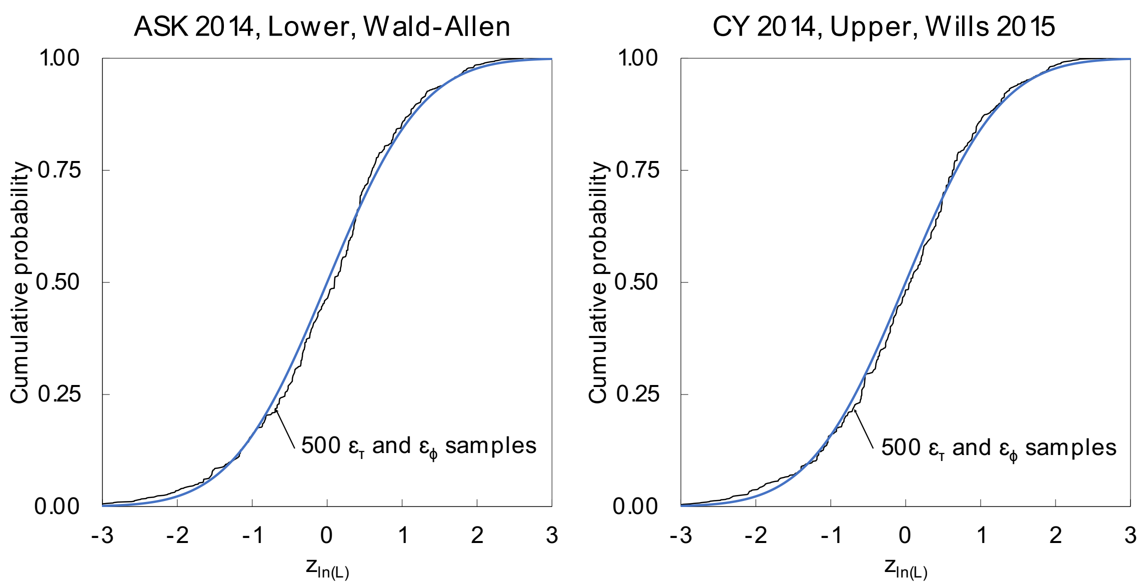


Figure 12. Two sample checks that portfolio loss is approximately lognormally distributed