Trimming the UCERF3-TD Logic Tree: Model Order Reduction for an Earthquake Rupture Forecast Considering Loss Exceedance

Keith Porter,[a] **M.EERI, Kevin Milner,**[b] **and Edward Field**[c] **M.EERI**

The Uniform California Earthquake Rupture Forecast version 3-Time Dependent depicts California's seismic faults and their activity. Its logic tree has 5,760 leaves. Considering 30 more model combinations related to ground motion produces 172,800 distinct models representing so-called epistemic uncertainties. To calculate risk to a portfolio of buildings, one also considers millions of earthquakes and spatially correlated ground-motion variability. We offer a tree-trimming technique that retains the probability distribution of portfolio loss. We applied it to a California statewide building portfolio and various levels of nonexceedance probability between 1 in 100 and 1 in 2,500. We trimmed the logic tree from 172,800 leaves to as few as 15. The result: a supercomputer that would otherwise run 24 hours to estimate the distribution of 1-in-250-year loss can calculate it in moments with the reduced-order model. Others can use the reduced-order model to calculate risk to different California portfolios, and scientists can prioritize study to reduce the remaining epistemic uncertainty.

INTRODUCTION

Why the size of the UCERF3-TD logic tree matters. 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. Branches have as few as two and as many as five discrete possible values. The choices allow for 5,760 combinations. Counting three more logic-tree branches for aspects of ground motion prediction, UCERF3-TD with a full ground-shaking model has 172,800 combinations of 11

[a] Institute for Catastrophic Loss Reduction, 800 Collip Cir, London ON N6G 4M3, Canada, kporter@iclr.org
[b] Southern California Earthquake Center, 3651 Trousdale Pkwy Rm 169, Los Angeles CA 90089
[c] U.S. Geological Survey, 1711 Illinois St, Golden, CO 80401

model elements. One must choose one option for each model element before one can calculate loss in a single earthquake. Each set of choices can produce a different value of loss.

With 172,800 choices, each with an associated probability of being the right choice and each capable of giving a different answer, the loss takes on a probability distribution. Its range of possible values spans an order of magnitude, i.e., a multiplicative error of 3 or more either way. Multiply by the so-called aleatory uncertainties of the between-events ground-motion variability, spatially correlated within-event ground-motion variability, and approximately 6,000,000 possible earthquake ruptures, and one can glimpse how robust calculation of the risk to a large portfolio of properties can grow prohibitively computationally expensive for anyone without access to a supercomputer. How large can a portfolio get? We estimate that the state of California has on the order of 10 million buildings.

State policymakers and insurance executives might want to know the monetary loss or number of fatalities with some specified rare but inevitable likelihood, such as the loss with 1 chance in 500 of happening next year. Few people have the resources to calculate the probability distribution of loss without making simplifying assumptions that might lead to a gross over- or under-estimate of the value with 0.2% chance of happening next year.

Some decision-makers can tolerate an answer that could be low or high by a factor of 3, but many cannot. Large insurers must buy reinsurance to be confident that a rare earthquake has a low chance of bankrupting them. Reinsurance can represent half of an insurer's annual budget (California Earthquake Authority 2022). If they buy 1/3rd as much as they need, they risk ruining themselves and their insureds through their inability to pay claims. If they buy three times too much, they must double the premiums they charge policyholders, which merely accelerates their bankruptcy when insureds cancel their policies. To make the right choice with confidence requires knowing the probability distribution of loss.

In two prior works that we discuss later, we offer new methods for trimming an earthquake rupture forecast logic tree to reduce computational effort without reducing uncertainty or biasing an estimate of expected annualized loss. In the present work, we revisit those methods with a similar goal, but considering large, rare losses rather than average annualized losses.

Objectives. The branches of the logic tree contribute unequally to loss uncertainty. Some contribute greatly to uncertainty, some do not. If one can find out which is which, one can fix the less-important modeling choices to a single value. If one can eliminate a branch with three choices, one reduces the size of the model by 3 times. Fix two logic-tree branches and the problem gets smaller by 9 times, requiring 1/9th the computational effort. Fix another and the problem is smaller by 27 times, requiring only 4% of the computational effort as before.

Mathematicians call that process “model order reduction.” Refer toSchilders et al. (2008) for general treatment. The goal of model order reduction is to find and fix as many branches of the logic tree as possible without changing the probability distribution of loss. Mathematicians have developed a rich body of model order reduction techniques. Most of them only work with scalar random variables, i.e., one-dimensional numbers that have scale such as the maximum earthquake magnitude that can occur away from a mapped fault. But between UCERF3-TD’s native earthquake-rupture branches and the additional ground-motion model elements, 7 of 11 logic-tree branches for a statewide risk calculation are nominal random variables.

A nominal random variable can take on values with no scale or order, no average, no standard deviation. For example, in one branch of UCERF3-TD, one chooses between five models of the relationship among slip length, rupture area, and magnitude. Each model corresponds to a different scholarly article. One chooses between the five articles. There is no sense in which the articles have a meaningful order or scale. Most existing model-order-reduction techniques do not apply to models with nominal random variables.

Here, we seek to select a single option for as many branches of UCERF3-TD plus the added ground motion uncertainties as we can, without changing the probability distribution of statewide portfolio loss each year with 1 chance in 100, 1 in 250, 1 in 400, 1 in 550, and 1 in 2,500. How does doing so help anyone? Our goals are twofold:

(1) Make the calculation of portfolio loss easier for other people who have different portfolios and no supercomputer. If they can ignore some branches, they can perform robust risk calculations that would otherwise take too long. That is, we aim to find a subset of logic tree branches in the present work that other people can use in their loss estimates so that they do not have to model all the logic tree branches of UCERF3-TD.

(2) Find the UCERF3-TD model variables that contribute most to uncertainty. Further study of those branches might yield new knowledge and reduce epistemic uncertainty.

Although we apply our solution to UCERF3-TD, it could apply to other problems: to future California earthquake hazard risk models, to earthquake models outside of California, to catastrophe risk models for other perils, and perhaps to other models with a mixture of nominal and scalar random variables.

LITERATURE REVIEW

UCERF3-TD logic tree. Let us first review the UCERF3-TD logic tree, then review model order reduction techniques. Field et al. (2013) offer a new earthquake rupture forecast for California: the Uniform California Earthquake Rupture Forecast version 3, Time-Independent, or UCERF3-TI. It has seven uncertain model components arranged in a logic tree. Each branch has two to five choices, each with a weight (a degree of belief or Bayesian probability). Field et al. (2015) add an eighth element to model aperiodicity in earthquake recurrence that makes the model time-dependent (hence the name Uniform California Earthquake Rupture Forecast version 3, Time-Dependent, or UCERF3-TD). Refer to Figure 1. Of the eight uncertain parameters, only three are scalar: total event rate of earthquakes of magnitude 5 or greater, maximum off-fault magnitude, and aperiodicity. We detail UCERF3-TD later.

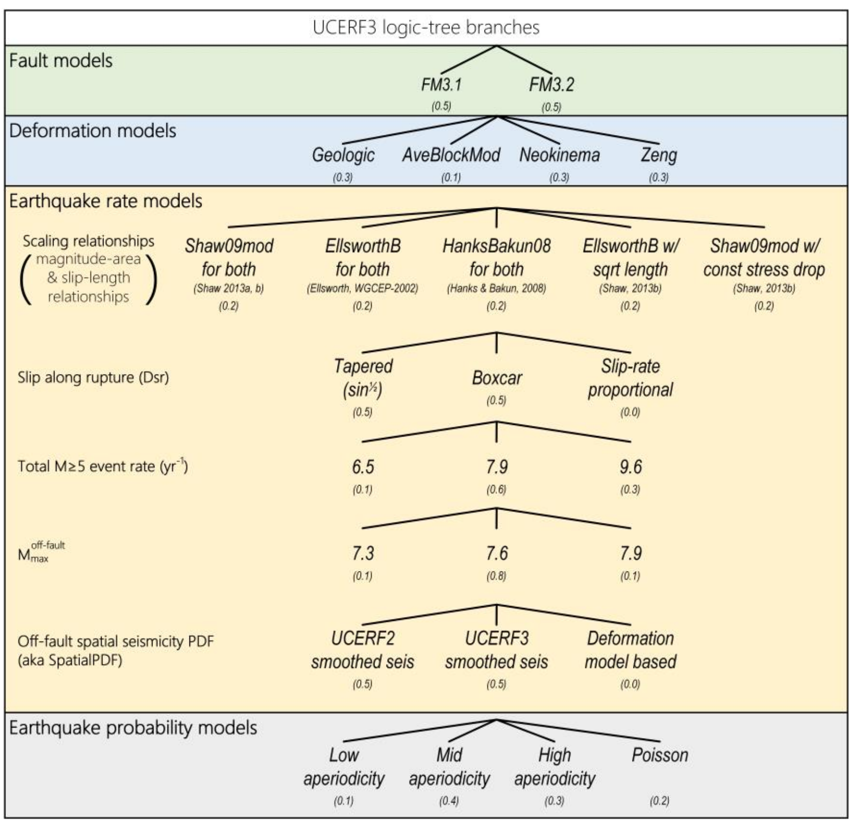

Figure 1. UCERF3-TD logic tree. Each branching point represents an uncertain variable; each branch a possible value.

Model order reduction techniques. Size limitations prevent a thorough review of model order reduction techniques, but a summary seems useful. They fall into five classes: proper orthogonal decomposition, reduced bias, simplified physics, nonlinear dimensionality reduction, and balancing methods. Proper orthogonal decomposition (e.g., Loeve 1955) requires one to evaluate and operate on a covariance matrix and find a smaller number of eigenvalues and eigenvectors, essentially changing n potentially correlated random variables into fewer than n uncorrelated ones. But there is no such thing as a correlation matrix for nominal random variables. The reduced-bias technique (e.g., Prud'homme et al. 2002) operates on linear functions of elliptic and parabolic partial differential equations; again, only scalar variables. A simplified-physics approach replaces a complex model with a simpler one using physical insight or empirical observation, which seems unhelpful to choosing between the modeling options considered here, which are already physically based and empirically supported. Balancing methods involve diagonalization of positive definite matrices (e.g., Antoulas 2005), again a problem limited to scalar values. Some nonlinear dimensionality reduction techniques might accommodate nominal variables: Graeme Weatherill (GFZ German Research Centre for Geosciences, written commun., November 8, 2023) suggests that t-distributed stochastic neighbor embedding (t-SNE) and uniform manifold approximation and projection (UMAP) could accommodate nominal variables. One can encode the nominal dimensions numerically such as with binary variables representing each category. Refer to e.g., McInnes et al. (2020) appendix C and Awan (2023). These methods might work, but the one we have in mind seems simpler, at least to us, and we know it works with OpenSHA (Field et al 2005), the software that encodes UCERF3-TD.

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 attempt 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 achieved large reductions in the required number of simulations for the mean and confidence limits of the conditional loss distribution (the loss distribution given some loss in a specific year), but only a threefold reduction for the unconditional, annual-loss distribution.

Kotha et al. (2018) offer a method to select an event set by matching the mean hazard at selected locations. Using six small portfolios of buildings in the San Franisco Bay Area, they show that they can reasonably reproduce average annualized losses and the loss exceedance curves generated by a catalog that represents a reduced version of the UCERF2 earthquake rupture forecast (Field et al. 2007). Event sets can reduce computational effort, but they shed no light on which branches of the logic tree matter to the distribution of loss, a central objective of the present work.

In prior work (Porter et al. 2012) we applied a deterministic sensitivity analysis technique called tornado-diagram analysis meant to identify the important variables in UCERF2 (Field et al. 2007). In Porter et al. (2017), we offer new a model order reduction technique that works on models with nominal random variables. We applied it to UCERF3-TD, using the expected annualized loss, *EAL*, to a proxy for the California Earthquake Authority's (CEA) statewide insurance portfolio. (*EAL* measured ground-up repair cost rather than insured loss after deductibles and limits.) We found a reduced-order model that required evaluating 60 leaves out of 57,600. Why not 172,800? Because in that work, we ignored a variable called added epistemic uncertainty recommended by Atik and Youngs (2014).

METHODOLOGY

Evaluate the model output for one logic-tree leaf. One begins by selecting an asset portfolio and evaluating the portfolio loss exceedance curve for one logic-tree leaf. That is, fix every branch to one value and evaluate loss in each rupture in the UCERF3-TD model. Calculate the loss exceedance curve as follows. Let

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; refer to equation (1)

$$V = \sum_{a=0}^{N_a-1} V_a \qquad (1)$$

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

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

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

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

$f_{Xa|k}(x)$ = probability density function of $X_{a|k}$, evaluated at x, given by the ground-motion-prediction equation, as in equation (2), in which ϕ denotes the Gaussian probability density function. Ground-motion-prediction equations generally assume lognormally distributed ground motion conditioned on rupture and site parameters, and provide a median and logarithmic standard deviation, denoted here by θ_{X_a} and β_{X_a}.

$$f_{X_a|k}\left(x\right)=\phi\left(\frac{\ln\left(x/\theta_{X_a}\right)}{\beta_{X_a}}\right) \tag{2}$$

$y_a(x)$ = mean repair cost as a fraction of replacement cost for asset a, given ground motion x. This quantity is evaluated using a vulnerability function (e.g., Porter 2009a, b, and 2010).

$\mu_{L,|k}$ = expected value of portfolio loss L given rupture k. For portfolios with assets that are spaced less than a few kilometers apart, within-event spatial correlation of ground motion matters. One can sample over N_τ values of the between-event ground-motion variability and N_f spatially correlated random fields of within-event ground-motion variability, and apply equation (3). In the equation, i is an index to between-event values, j is an index to stochastic simulations of within-event variability, $x_{ai,j}$ is the ground motion at asset a given between-event term i and within-event simulation j, and $w_{\tau i}$ denotes the weight applied to between-event value i. Refer to Porter et al. (2024) for details of the spatially correlated ground motions and for a simplification to equation (3).

$$\mu_{L|k}=\sum_{i=0}^{N_\tau-1}\sum_{j=0}^{N_f-1}\sum_{a=0}^{N_a-1}V_a y_a\left(x_{ai,j}\right)w_{\tau_i}\frac{1}{N_f} \tag{3}$$

L = uncertain portfolio loss

l = a value of L

$\delta_{L|k}$ = coefficient of variation of portfolio loss in rupture k. Refer to Porter et al. (2024) 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 equation (4). In the equation, the coefficient 1000/V normalizes the mean loss in terms of loss per \$1000 of replacement cost, a loss measure sometimes used in the catastrophe-risk modeling industry. Porter et al. (2024) presents a regression analysis that suggests the following values for c_1 and c_2. The resulting curve gradually drops from 2 (at low portfolio loss) to 0.5 (at high portfolio loss).

c_1 = a parameter for estimating $\delta_{L|k}$ = 0.9832

c_2 = a parameter for estimating $\delta_{L|k}$ = -0.117

$$\delta_{L|k} \approx c_1 \cdot \left(\frac{1000}{V} \mu_{L|k} \right)^{c_2} \tag{4}$$

$\theta_{L|k}$ = median value of portfolio loss L given rupture k, assuming that L is approximately lognormally distributed; refer to equation (5). Porter et al. (2024) offers evidence.

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

$\beta_{L|k}$ = standard deviation of the natural logarithm of portfolio loss L given rupture k, assuming that L is approximately lognormally distributed. Refer to equation (6).

$$\beta_{L|k} = \sqrt{\ln\left(1+\left(\delta_{L|k}\right)^2\right)} \tag{6}$$

r_k = rate at which rupture k occurs, given the choice of logic tree leaf. The earthquake rupture forecast (e.g., Field et al. 2015) provides r_k. The reader may wonder how rate comes from a time-dependent model. Here, r_k is the equivalent Poisson rate implied by the chosen start date and duration of the forecast.

$G(l)$ = 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, as shown in equation (7).

$$G(l) = \sum_{k=0}^{N_k-1} r_k \cdot \left(1 - \Phi\left(\frac{\ln\left(l/\theta_{L|k}\right)}{\beta_{L|k}} \right) \right) \tag{7}$$

L_p = loss with a specified exceedance rate, rather than the exceedance rate of some value of loss. It is the inverse of the loss exceedance curve evaluated at p, as shown in equation (8)

$$L_p = G^{-1}(p) \tag{8}$$

Evaluate the cumulative distribution function of the full model output. Let

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. The earthquake rupture forecast (e.g., Field et al. 2015) specifies leaf weights.

$L_{p,j}$ = loss associated with exceedance frequency p in logic-tree leaf j, from equation (8). Note that each leaf j can have a different loss exceedance curve and therefore a different 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}. It will not be necessary to assume a parametric form of the distribution of L_p such as normal or lognormal.

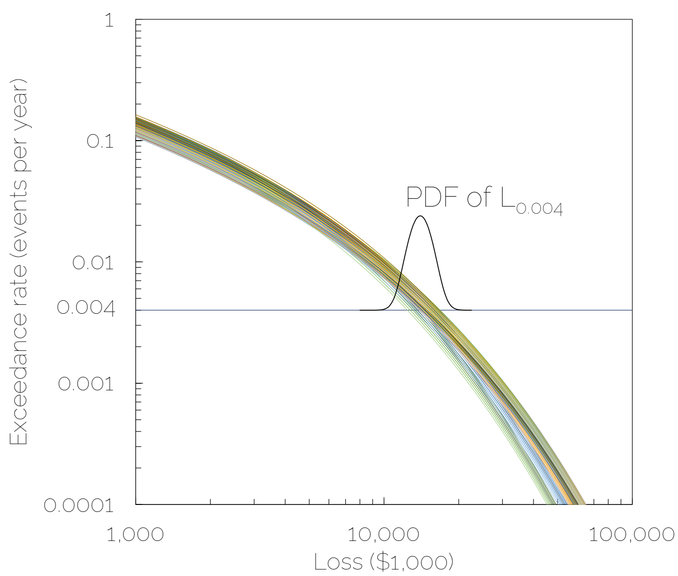

Figure 2. Illustration of the probability density function (PDF) of L_p. The colored curves represent loss-exceedance curves for different logic-tree leaves. The present model-order-reduction effort aims to reduce the number of possible loss exceedance curves (thereby simplifying the model and reducing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

$F_{L_p}(l)$ = cumulative distribution function for L_p in the full model using equation (9), in which H is the Heaviside function, as shown in equation (10). $F_{L_p}(l)$ has a mean value given by equation (11), variance by equation (12), and coefficient of variation by equation (13).

$$F_{L_p}(l) = \sum_{j=0}^{Z-1} w_j \cdot H\left(l - L_{p,j}\right) \tag{9}$$

$$H(x) = 0 \qquad x < 0$$
$$= 0.5 \qquad x = 0 \tag{10}$$
$$= 1 \qquad x > 0$$

$$\mu_{L_p} = \sum_{j=0}^{Z-1} L_{p,j} \cdot w_j \tag{11}$$

$$\sigma_{L_p}^2 = \left(\sum_{j=0}^{Z-1} \left(L_{p,j}\right)^2 \cdot w_j \right) - \left(\mu_{L_p}\right)^2 \tag{12}$$

$$\delta_{L_p} = \frac{\sigma_{L_p}}{\mu_{L_p}} \tag{13}$$

Evaluate the loss exceedance curve for a reduced-order model. Here is how to evaluate the exceedance curve for a reduced model and to measure the error in loss with a specified exceedance rate 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 model, meaning the number of leaves in it, by equation (14).

$$z = \sum_{j=0}^{Z-1} I_j \tag{14}$$

c_0 = normalizing constant for weights in the reduced-order model, using equation (15).

$$c_0 = \sum_{j=0}^{Z-1} w_j \cdot I_j \tag{15}$$

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, by equation (16), which has an expected value given by equation (17), variance given by equation (18), and coefficient of variation given by equation (19).

$$\hat{F}_{L_p}(l) = \frac{1}{c_0} \cdot \sum_{j=0}^{Z-1} w_j \cdot I_j \cdot H\left(l - L_{p,j}\right) \tag{16}$$

$$\hat{\mu}_{L_p} = \frac{1}{c_0} \sum_{j=0}^{N_j-1} L_{p,j} \cdot w_j \cdot I_j \tag{17}$$

$$\hat{\sigma}_{L_p}^2 = \left(\frac{1}{c_0} \sum_{j=0}^{Z-1} \left(L_{p,j}\right)^2 \cdot w_j \cdot I_j \right) - \left(\hat{\mu}_{L_p}\right)^2 \tag{18}$$

$$\hat{\delta}_{L_p} = \frac{\hat{\sigma}_{L_p}}{\hat{\mu}_{L_p}} \tag{19}$$

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_{L_p}. One calculates the maximum difference in the cumulative distribution functions, D_n, as in equation (20), and checks that satisfies inequality (21). To apply the two-sample Kolmogorov-Smirnov goodness-of-fit test at the 1% significance level, use c_{ks} = 1.63; at the 5% significance level, c_{ks} = 1.36. It is also desirable to ensure that errors in the mean and coefficient of variation of L_p, defined by equations (22) and (23) respectively, are both less than some reasonable limit, say 5%; refer to inequalities (24) and (25). If the reduced model passes the test specified in equation (21), we can reject at the 1% significance level that the two distributions differ. If it fails equation (24), 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 (25), the reduced model is (probably) getting too certain, even if the Kolmogorov-Smirnov test says it is drawn from the same distribution.

$$D_n = \max_l \left(\left| F_{L_p}(l) - \hat{F}_{L_p}(l) \right| \right) \tag{20}$$

$$D_n \leq c_{ks} \sqrt{\frac{z+Z}{z \cdot Z}} \tag{21}$$

$$\varepsilon_\mu = \frac{\hat{\mu}_{L_p} - \mu_{L_p}}{\mu_{L_p}} \tag{22}$$

$$\varepsilon_\delta = \frac{\hat{\delta}_{L_p} - \delta_{L_p}}{\delta_{L_p}} \tag{23}$$

$$\left| \varepsilon_\mu \right| \leq 0.05 \tag{24}$$

$$\left| \varepsilon_\delta \right| \leq 0.05 \tag{25}$$

Path search. With the foregoing equations, we can apply the path-search technique from Porter et al. (2017) to model order reduction for L_p.

1. Evaluate $F_{L_p}(l)$, μ_{L_p}, and δ_{L_p} for the full model as shown in equations (1) through (13).

2. Let *a* denote an index to independent variables and *b* denote an index to their possible values. 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 (20), (22), and (23), 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 inequality (21) and inequalities (24) and (25). 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 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 ($c = 2, 3, \ldots N_c - 1$) where N_c is the number of branches in the logic tree.

APPLICATION TO UCERF3-TD TREE TRIMMING PROBLEM

Independent variables: branches of UCERF3-TD plus three ground-motion branches. To estimate ground motion, we add three uncertainties not shown in Figure 1: site characteristics (which model of Vs30—average shear-wave velocity in the upper 30 m of soil—to use), which of five ground-motion-prediction equations to use, and how much epistemic uncertainty to add. Note that Field et al. (2020) suggest that added epistemic uncertainty is improperly posed and may exert a large, unjustified influence on results, but we still included it here.

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-TD what each variable represents. 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.

Table 1. Independent variables, variable types, possible values, weights, and descriptions

A	Variable (branch) name	Type	*b*	Possible value[1]	*w*	Description. See Field et al. (2013) Table 15 for maps of size and extent of effects.
0	Fault model	N	0	FM 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.
			1	FM 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	

A	Variable (branch) name	Type	*b*	Possible value[1]	*w*	Description. See Field et al. (2013) Table 15 for maps of size and extent of effects.
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 SHAW 09m 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 spatl seism 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.
			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. Mid to high coefficient of variation (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 30 m of soil using 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. BSSA2014 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 NGAWest-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 per Field et al. (2013)

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 or for a loss exceedance curve also requires inputs that one could consider independent variables:

Portfolio. We considered a portfolio of buildings similar in composition, value, and geographic distribution to the one insured by the California Earthquake Authority, the state's largest insurer of earthquake risk to residences. The portfolio represents an estimate of the assets exposed to risk. Each asset is 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 (Federal Emergency Management Agency 2012), 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 California Earthquake Authority's market penetration rate—that is, the fraction of homes they insure. 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). Refer to the research data statement for the portfolio data.

Vulnerability functions. These relate ground motion to mean repair cost (and sometimes variability) as a fraction of replacement cost new. We used Hazus-based vulnerability functions from 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 trimming the UCERF3-TD logic tree.

RESULTS FOR LOSS *L* WITH VARIOUS EXCEEDANCE PROBABILITIES

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 an insurer with several lines of business in several states, which provide diversification benefits. The California Earthquake Authority is different for exactly these reasons: one line of business, one state, all catastrophe risk. The California Earthquake Authority's current risk-transfer strategy approved by its board (and revealed 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 $p \in \{0.01, 0.004, 0.0025, 0.0018, 0.0004\}$. Refer to Porter et al. (2024) for more details.

Table 2 summarizes results. Columns reflect probability levels. Rows show independent variables organized from least to most important. The least important can be trimmed from all models without significantly affecting the probability distribution of the dependent variable. Where a variable can be trimmed, the table shows the value to which it can be set. Some variables always strongly influence the dependent variable. Some only affect the dependent variable for some probabilities. 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, called "additional epistemic uncertainty" cannot be trimmed at all without greatly disturbing the dependent variables. It seems improperly posed and may exert an unjustified influence on results.

Table 2 shows that the optimal trimmed logic tree differs depending on exceedance probability level. So how can one get value from it in practice? We suggest a pragmatic approach: use the 1/250 choices regardless of the probability level of interest. Its choices share parameter values most common to all five probability levels. It greatly reduces the computational effort, but neither by the most nor the least, a sort of golden mean for model order reduction. And 1/250 may be the most common point insurers and reinsurers consider on the loss exceedance curve. However, this is just a suggestion; other opinions may differ.

Table 3 summarizes the size of each reduced-order model. Columns indicate the dependent variable for which the model was trimmed. Rows show the size of the full and reduced models.

Table 2. 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	(app 4)
Maximum Off-Fault Magnitude	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
Earthquake 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
Added Epist Uncertainty						

* We recommend using the 1/250 results in general for reasons explained in the text

Table 3. Summary of the degree of model order reduction

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

SUMMARY AND CONCLUSIONS

We identify a reduced-order model for the UCERF3-TD logic-tree model using a subset of 11 independent variables that reproduces the probability distribution of an important dependent variable: loss at a low nonexceedance probability. 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.

Our model order reduction technique starts by evaluating the probability distribution of the full model's dependent variable. It trims one independent variable at a time, setting it to one possible value and tests whether the probability distribution of the dependent variable significantly changes or its first two moments significantly change relative to the full model. The reduced-order model with the smallest change is preferred. One iterates until reaching the smallest model that preserves the probability distribution of the dependent variable (passing a two-sample Kolmogorov-Smirnov test at 1% significance) and the dependent variable's first two moments within ±5%. We applied the technique to the loss exceedance curve.

At loss-exceedance probabilities generally used by insurers and the California Earthquake Authority in particular (1/250 to 1/550), one can trim six to nine of UCERF3-TD's 11 independent variables, reducing the model by 99.7% to 99.991%. We recommend fixing six parameters as shown in Table 2 for the California Earthquake Authority's 1/250-year loss. Doing so reduces the model size and computational effort by 99.7%. A hypothetical risk calculation that takes 24 hours for the full model can be reduced one that takes seconds.

This technique can handle a model that produces a scalar dependent variable that depends on scalar and nominal independent variables. It allows for interaction between independent variables. 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, since the problems differ mostly in the choice of the dependent variable. The technique reduced the loss model from 172,800 leaves to 15 leaves in the cases of the 400- and 550-year repair cost.

Which independent variables can be trimmed depends on the choice of dependent variable. The preferred value of the trimmed 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. With greater study of those two uncertainties, researchers might reduce them. Doing so would thin the upper tail of the loss distribution. It would save insurers on reinsurance. And it would save policyholders on premium costs that help pay for reinsurance.

With some limitations discussed next, the present model order reduction technique seems applicable to future earthquake rupture forecasts and other risk models that share the features of UCERF3-TD: a combination of independent (or transformable to independent) scalar and nominal uncertain variables, and probably ordinal variables as well.

All studies are limited. Good ones raise interesting questions. Here are some limitations and some questions. First, we applied the technique only to a single deterministic statewide portfolio. Would other portfolios have different results? We suspect they will be like the differences between columns in Table 2, sharing many common choices.

We did not account for uncertainty in the vulnerability functions. How important is that? Nor did we account for other uncertainties in the portfolio. For example, how important is uncertainty in the assignment of 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.

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, as demonstrated by Field et al. (2017).

Can one identify *a priori* the branches of the complete logic-tree that contribute much less to the uncertainty than others, without first computing the losses for each combination? Tornado-diagram analysis examines the effect of each branch separately; it would be interesting to check whether the approach reliably predicts that the same variables matter.

Our method operates on one dependent variable. What if the model has more? Here are two options: (1) Produce a separate reduced-order model for each dependent variable, or (2) In step 3 of the path search, calculate D_n for each dependent variable and trim branches by selecting the (a, b) pair with the smallest value of the *sum* of D_n values where *each individual* D_n satisfies the goodness-of-fit test in inequality (21) and inequalities (24) and (25).

ACKNOWLEDGEMENTS AND DISCLAIMERS

The California Earthquake Authority and the U.S. Geological Survey funded this work. The authors have no conflict of interest. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. We thank reviewers E. Bilderback, J. Carter, R. Gold, K. Jaiswal, S.R. Kotha, and G. Weatherill.

RESEARCH DATA AND CODE AVAILABILITY

Find the SA10 random fields at https://doi.org/10.25810/xf0m-m080, the building portfolio at https://doi.org/10.25810/094s-mp33, and OpenSHA code at https://github.com/opensha/.

REFERENCES CITED

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

Antoulas, A.C. (2005). Approximation of Large-Scale Dynamical Systems. Society of Industrial and Applied Mathematics, Philadelphia, 487 p.

Atik, L.A., and Youngs, R.R. (2014). Epistemic uncertainty for NGA-West2 models. *Earthquake Spectra* 30 (3), https://doi.org/10.1193/062813EQS173M

Awan, A.A. (2023). Intro to t-SNE. *DataCamp.* https://www.datacamp.com/tutorial/introduction-t-sne

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

California Earthquake Authority (2022). *The Strength to Rebuild: Financial Foundations of the California Earthquake Authority*. Sacramento, CA

Chiou, B.S.J., and Youngs, R.R (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.

Federal Emergency Management Agency (2012). *Multi-hazard Loss Estimation Methodology Earthquake Model Hazus®-MH 2.1 Technical Manual*. Washington, DC, 718 pp.

Field, E.H., Gupta, N., Gupta, V., Blanpied, M., Maechling, P., and Jordan, T.H (2005). Hazard calculations for the WGCEP-2002 forecast using OpenSHA and distributed object technologies. Seismological Research Letters 76: 161-167.

Field, E.H., T.E. Dawson, K.R. Felzer, A.D. Frankel, V. Gupta, T.H. Jordan, T. Parsons, M.D. Petersen, R.S. Stein, R.J. Weldon II, and C.J. Wills, 2007. The Uniform California Earthquake Rupture Forecast, Version 2 (UCERF 2). U.S. Geological Survey Open File Report 2007-1437, Reston VA, 96 p.

Field, E.H., Arrowsmith, R.J., Biasi, G.P., Bird, P., Dawson, T.E., Felzer, K.R., Jackson, D.D., Johnson, K.M., Jordan, T.H., Madden, C., Michael, A.J., Milner, K.R., Page, M.T., Parsons, T., Powers, P.M., Shaw, B.E., Thatcher, W.R., Weldon II, R.J., and Zeng, Y. (2013). Uniform California Earthquake Rupture Forecast, Version 3 (UCERF3)—the Time-Independent Model. U.S. Geological Survey Open-File Report 2013–1165, Reston VA, 115 p.

Field, E.H., Biasi, G.P., Bird, P., Dawson, T.E., Felzer, K.R., Jackson, D.D., Johnson, K.M., Jordan, T.H., Madden, C., Michael, A.J., Milner, K.R., Page, M.T., Parsons, T., Powers, P.M., Shaw, B.E., Thatcher, W.R., Weldon II, R.J., 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.

Field, E.H., Porter, K.A., and Miller, K.R. (2017). A prototype operational earthquake loss model for California based on UCERF3-ETAS – a first look at valuation. *Earthquake Spectra* 33 (4)

Field, E.H.., Milner, K.R.., and Porter, K.A. (2020). Assessing the value of removing earthquake-hazard-related epistemic uncertainties, exemplified using average annual loss in California. *Earthquake Spectra* 36 (4), https://doi.org/10.1177/8755293020926185

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

Loeve, M. (1955). *Probability Theory.* Van Norstrand Co Inc., New York, 516 pp.

McInnes, L., Healy, J., and Melville, J. (2020). *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction.* arVix, https://doi.org/10.48550/arXiv.1802.03426

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://doi.org/10.1061/40687(2003)87

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.

Porter, K.A. (2009a). Cracking an open safe: HAZUS vulnerability functions in terms of structure-independent spectral acceleration. *Earthquake Spectra* 25 (2): 361-378.

Porter, K.A. (2009b). Cracking an open safe: more HAZUS vulnerability functions in terms of structure-independent spectral acceleration. *Earthquake Spectra* 25 (3): 607-618.

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

Porter, K.A., Field, E.H., and Milner, K. (2012). Trimming the UCERF2 hazard logic tree. *Seismological Research Letters*, 83 (5): 815-828.

Porter, K.A., Milner, K., and Field, E.H. (2024). Trimming the UCERF3-TD Logic Tree : Model Order Reduction for an Earthquake Rupture Forecast Considering Loss Exceedance, Supplemental Material. CU Scholar…

Prud'homme, C., Rovas, D.V., Veroy, K., Machiels, L., Maday, Y., Patera, A.T., and Turinici, G. (2002). Reliable real-time solution of parametrized partial differential equations: Reduced-basis output bound methods. *Journal of Fluids Engineering* 124(1): 70-80.

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

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

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

Wills, C.J., Gutierrez, C.I., Perez, F.G., and Branum, D.M. (2015). A next generation VS30 map for California based on geology and topography. *Bull. Seism. Soc. of America* 105 (6): 3083–3091.