SESM 16-05

Trimming the UCERF3-TD Hazard Tree with a New Probabilistic Model-Reduction Technique

By

Keith A. Porter
Edward Field
Kevin Milner

Sep 2016

Structural Engineering and Structural Mechanics Program
Department of Civil Environmental and Architectural Engineering
University of Colorado
UCB 428
Boulder, Colorado 80309-0428

Contents

Index of Figures

Index of Tables

Trimming the UCERF3-TD Hazard Tree with a New Probabilistic Model-Reduction Technique

Keith Porter,[a] Edward Field,[b] and Kevin Milner[c]

The size of the logic tree within the Uniform California Earthquake Rupture Forecast Version 3, Time-Dependent (UCERF3-TD) poses a challenge to risk analyses of large portfolios, especially when multiplied by multiple ground-motion prediction equations and site-effect models. An insurer or catastrophe risk modeler concerned with rare, catastrophic losses to a portfolio of California assets would today have to evaluate a portfolio 57,600 times to create a loss exceedance curve that explores the entire possibility space. Which branches matter most, and which can be ignored? We employed two model-reduction techniques to find a subset of UCERF3-TD parameters that must vary and fixed baseline values for the remainder such that the reduced model produces approximately the same distribution of loss that the full model does. The two techniques are (1) a tornado-diagram approach we employed previously for UCERF2, and (2) an apparently novel probabilistic sensitivity approach that appears better suited to functions of nominal random variables. The newer approach produces a smaller reduced model with only 60 leaves. Results can be used to reduce computational effort in loss analyses by several orders of magnitude.

INTRODUCTION

Probabilistic seismic hazard analyses (PSHA) and probabilistic seismic risk analyses (PSRA) have grown in complexity since their introduction by Esteva (1967) and Cornell (1968). They can involve many analytical stages, each with multiple competing models that attempt to idealize nature, such as how to relate rupture area to magnitude. One cannot be sure

[a] University of Colorado Boulder, 428 UBC, Boulder CO 80309-0438
[b] U.S. Geological Survey, 1711 Illinois St., Golden CO 80401
[c] Southern California Earthquake Center, 3651 Trousdale Pkwy #169, Los Angeles, CA 90089

which competing model best reflects reality, so analysts sometimes accept multiple models and arrange them in a logic tree, as in the case of the Uniform California Earthquake Rupture Forecast versions 2 (UCERF2, Field et al. 2009) and its time-dependent version-3 successor, UCERF3-TD (Field et al. 2015).

In such a logic tree, each modeling choice represents an independent variable—one subject to the user's choice—and is used as an input parameter for the evaluation of an dependent variable. If the logic tree is used in PSHA or PSRA, the dependent variable typically measures hazard (such as the exceedance frequency of various levels of peak ground acceleration) or risk (such as the exceedance frequency of various levels of building repair cost). Each possible value of each independent variable is assigned a weight (the term used by people who hold a frequentist or classical viewpoint of probability) or probability (from the Bayesian viewpoint; we will use the Bayesian terminology here). Probabilities of all the allowable values of one independent variable sum to unity. The dependent variable is evaluated for every allowable combination of values of the independent variables, that is, for every possibility in the possibility space. The probability of each possibility is taken as the product of the probabilities of the independent variable values that led to that value of the dependent variable. One can think of each such combination as a leaf on the logic tree, each leaf associated with one value of the dependent variable and one probability that it is the correct value. One can then evaluate the cumulative distribution function (CDF) of the dependent variable at any given value by sorting the leafs in increasing order and summing the probabilities of outcomes less than or equal to that value.

Because the number of leafs in the logic tree increase exponentially with the number of independent variables, a large model can make the evaluation of hazard or risk computational costly, even on a supercomputer. We address here the question of whether and how one can reduce the computational effort without significantly changing the estimate of hazard or risk. Here are some reasons to believe one can do so. Not all branches equally affect the CDF of the dependent variable—some branches matter less than others. In the cases of UCERF2 and UCERF3-TD, some model elements matter less than others because they deal with geographically localized issues that don't matter much at the statewide level. Some values of some independent variables matter less because they have less weight attached to them, that is, because modelers believe those values are less likely to be correct or applicable. Some branches

matter less because, despite qualitative differences, alternative values produce only slightly different values of the dependent variable.

To the extent that one can prune the logic tree of these low-impact branches, one may be able to greatly reduce the size of the logic tree and thus the computational effort with little impact on the resulting hazard or risk estimates. That is our goal for UCERF3-TD: to find a subset of its parameters—its independent variables—that have the least overall impact on statewide risk, and trim them from the UCERF3-TD logic tree, leaving a smaller model comprising only the important parameters, a smaller model that requires less computational effort for a PSHA or PSRA.

MODEL ORDER REDUCTION TECHNIQUES

Let us begin by examining the literature pertaining to the general problem of reducing the complexity and large size (that is, the dimension) of mathematical simulations of real-life, nonlinear processes. The general problem is called model order reduction. Model order reduction addresses many different situations. We focus on a subclass of problems whose state space—that is, whose degrees of freedom or independent variables—is entirely composed on all nominal numbers, that is, quantities without scale or order.

To explain nominal numbers: the student identification numbers of children in a classroom are examples; they are merely labels. A second class of numbers called ordinal tell of order but not scale. For example, if we were to assign the number 1 to the tallest child in the classroom, the number 2 to the second tallest, etc., the number assigned to any given child is an ordinal number. A third class called cardinal numbers measure relative quantity, such as the height of the same schoolchildren in centimeters. The distinction will be important because most model-order-reduction techniques deal with functions of uncertain cardinal numbers, and rely on mathematical operations that are only defined for cardinal numbers.

UCERF3-TD represents a model whose state space is comprised of all nominal numbers: all of its logic-tree branches represent choices among competing models that we identify merely with labels such as the names of publications that offer relationships between rupture area and earthquake magnitude. This particular class of problem is less commonly dealt with than models whose state space is all comprised of cardinal numbers.

There are at least eight common approaches to model order reduction that involve projection or transformation of the state space to fewer degrees of freedom, but all are limited

to models of cardinal numbers. The eight include: (1) proper orthogonal decomposition (POD), also known as principal component analysis (PCA; see Chatterjee 2000); (2) balanced truncation and (3) approximate balancing by iteration (e.g., Gugercin and Antoulas 2004 and Sorensen and Antoulas 2002 respectively); matrix interpolation (e.g., Amsallem and Farhat 2008); (4) matrix interpolation (Benner et al. 2015); (5) transfer function interpolation (also Benner et al. 2015); (6) Loewner framework (Mayo and Antoulas 2007); (7) cross Gramian (e.g., Antoulas 2009); (8) Krylov subspace techniques (Bai 2002). All apply to linear functions of real numbers, and are therefore inapplicable to UCERF3-TD.

These approaches are not applicable to models whose independent variables are nominal numbers because the independent variables cannot be arranged into a vector space that one can then project into an alternate, reduced-order space. All eight approaches require the application of differential calculus or other mathematical operations that are only defined for cardinal numbers to project the independent variables into a different, lower-order space. The only cardinal number we have to operate on here is the dependent variable, so the math all has to operate on the (cardinal) dependent variable rather than on the (nominal) independent variables.

Instead of a projection approach, we seek instead to separate the independent variables into those that contribute strongly to the dependent variable and those that do not, and to find baseline values of the latter at which one may fix them to reduce computational effort in future loss estimates. Let us refer to the first group (the independent variables that contribute strongly and we therefore allow to vary) as the varying set, the latter group (the ones we will fix) as the fixed set, and the combination as the full set. Let us refer to a model that allows the full set to vary as the full model, and let us refer to a model that allows the varying set to vary and fixes the fixed set at its baseline values as the reduced-order model.

Let us refer to the number of combinations of model-parameter values as the size of the model. (One usually refers to the number of parameters as the dimension of the model, but size is more useful when the model has a discrete number of possible values of each parameter, and when that number varies from parameter to parameter). Let us refer to any particular combination of values of the independent variables as a leaf in the logic tree. The number of leaves is the size of the model. We care less about the reduction in dimension—how many independent variables there are—than about the size of the model. For example, if we can fix one parameter with five possible values at one of those values, that is more desirable than fixing

two parameters each with two possible values to one value each. The former is preferable because it reduces the size of the model (the computational effort) by a factor of 5, the latter by 4, even though the former reduces the model dimension by 1 and the latter by 2.

We found no literature on model order reduction techniques that apply to state spaces whose dimensions are all nominal numbers. In past research (Porter et al. 2012), we modified a method of deterministic sensitivity study called a tornado-diagram analysis that seems to originate in the field of decision analysis (Howard 1988). We used the modified approach as a model order reduction technique to reduce the size of the Uniform California Earthquake Rupture Forecast version 2 (Working Group on California Earthquake Probabilities [WGCEP], 2007; Field et al., 2009). See Porter et al. (2012) for details of the modified tornado-diagram analysis, which we briefly recap later, along with an apparently new approach that we present and apply here for the first time.

UCERF3-TD

Let us now describe the problem that motivated this research. The ongoing Working Group on California Earthquake Probabilities (WGCEP) is responsible for developing authoritative earthquake forecasts for California on behalf of the United States Geological Survey (USGS), the Southern California Earthquake Center (SCEC), and the California Geological Survey (CGS). The most recent WGCEP model is known as the third Uniform California Earthquake Rupture Forecast (UCERF3), which was developed with support from California Earthquake Authority. As with our analysis of the previous model, UCERF2, the objective here is to identify the logic-tree parameters that contribute most strongly to uncertainty in statewide loss estimates.

The long-term, time-independent model (UCERF3-TI; Field et al., 2015 and references therein), provides the long-term rate of all possible earthquakes throughout the region (at some level of discretization and above magnitude 5.0). The primary achievements for UCERF3-TI were a relaxation of fault-segmentation assumptions and the inclusion of multi-fault ruptures, both of which were acknowledged limitations of the previous model (UCERF2). The rates of all earthquakes were solved for simultaneously, and from a broader range of data, using a system-level "grand inversion" that is both conceptually simple and extensible. This new approach is more derivative and less prescriptive than that taken previously; for example, rather than assuming a magnitude-frequency distribution (MFD) on most faults, the inversion solves

for the MFD that is most consistent with available data. The inverse problem is generally large and underdetermined, so a range of solutions was sampled using an efficient simulated annealing algorithm. The model also made more explicit use of geodetic data via three new deformation models, which not only provide alternative fault slip-rate constraints, but also enabled the inclusion of 150 fault sections that were previously excluded due to lack of geologic data. These additions served to fill out and expose the interconnectivity of the fault system, thereby revealing more multi-fault rupture possibilities. For example, the number of fault-based ruptures increased from 10,000 in UCERF2 to more than 250,000 in UCERF3.

Overall, UCERF3-TI has a lower rate of *M* 6.5-7.0 earthquakes, reflecting an explicit regional MFD constraint added to avoid a UCERF2 over-prediction at these magnitudes. The rate of larger, multi-fault ruptures generally increased as a consequence, reflecting a tradeoff that was effectively brokered by the grand inversion in satisfying all data constraints. In other words, the previous over-prediction problem in UCERF2 was turned into part of the solution with respect to the rate of multi-fault ruptures, although rates of the latter were also effectively minimized in UCERF3-TI, so if anything, it under predicts the frequency of such events (Field et al., 2015). Following extensive review, UCERF3-TI was used in the 2014 update of the USGS National Seismic Hazard Maps (Petersen et al., 2014; Powers and Field, 2015).

Building on UCERF3-TI, the WGCEP subsequently defined a time-dependent model (UCERF3-TD) that uses renewal models to represent elastic-rebound-implied rupture probabilities (Field et al., 2015, and references therein). A new methodology was developed, which solved applicability issues in the previous UCERF2 approach with respect to unsegmented models. The new algorithm also supports magnitude-dependent aperiodicity in the renewal model, and accounts for the historic open interval on faults that lack a date-of-last-event constraint. In general, UCERF3-TD probabilities are relatively low compared to UCERF3-TI on faults where a recent large event has occurred, and relatively high where the time since last event is greater than roughly half the average recurrence interval.

A number of modeling choices are represented in UCERF3-TD, as shown in the logic tree depicted in Figure 1. The two alternative fault models represent uncertainty in the existence and/or geometry of several faults. The four deformation models represent alternative sets of fault slip rates, reflecting both methodological differences and the degree of influence from geodetic versus geologic constraints. The five scaling relationships represent viable relationships between average slip versus fault length and/or magnitude versus area, where

differences reflect uncertainties with respect to the depth penetration of large earthquakes and/or whether surface displacement observations are representative of slip at mid-seismogenic depths (~7 km). The branches labeled "slip along rupture" represent the average slip as you go from one end of a fault rupture to the other, with one option applying a taper and the other keeping slip uniform (on average, meaning averaged over many realizations of a rupture). The total $M \geq 5$ event rate defines the total long-term rate of earthquakes in the entire forecast region (California plus a buffer zone), with the three options reflecting uncertainties in historical- and instrumental-catalog inferences due to potential intrinsic rate variability, the influence of aftershocks, and catalog magnitude-completeness questions. $M_{max}^{off-fault}$ represents the largest magnitude that can occur in the background-seismicity component of the model (off explicitly modeled faults), with the alternative values representing expert judgment. The branches labeled "off-fault spatial seis PDF" reflect the degree of smoothing applied to historical seismicity in defining the spatial distribution of background seismicity rates, where the "UCERF3 smoothed seis" option reflecting much less smoothing than the "UCERF2 smoothed seis" alternative, the latter of which was applied exclusively in UCERF2. Finally, the earthquake probability models reflect the degree of elastic-rebound predictability in the model, as defined by the aperiodicity of the renewal model; the "low", "med" and "high" choices reflect increasing predictability, with the "Poisson" option implying none.

All combined, UCERF3-TD has 5,760 logic-tree branches, each representing a viable model with an associated weight, or probability, of being correct. A variety of logic-tree sensitivity analyses, as well as comparisons with UCERF2, are given in the UCERF3-TI and UCERF3-TD reports for several different evaluation metrics (e.g., the probability of occurrence for one or more $M \geq 6.7$ earthquakes; 2%-in-50-year ground-motion exceedance). However, and as emphasized repeatedly in those reports, the relative importance of logic-tree branches depends strongly on the chosen evaluation metric and the geographic location of the site(s) of interest. This, coupled with assumptions and approximations in the model, means that UCERF3 applicability should be evaluated on a case-by-case basis. The purpose of this paper is to explore logic-tree sensitivity with respect to statewide loss estimates, with the hope of prioritizing which uncertainties should get focused on in future studies.

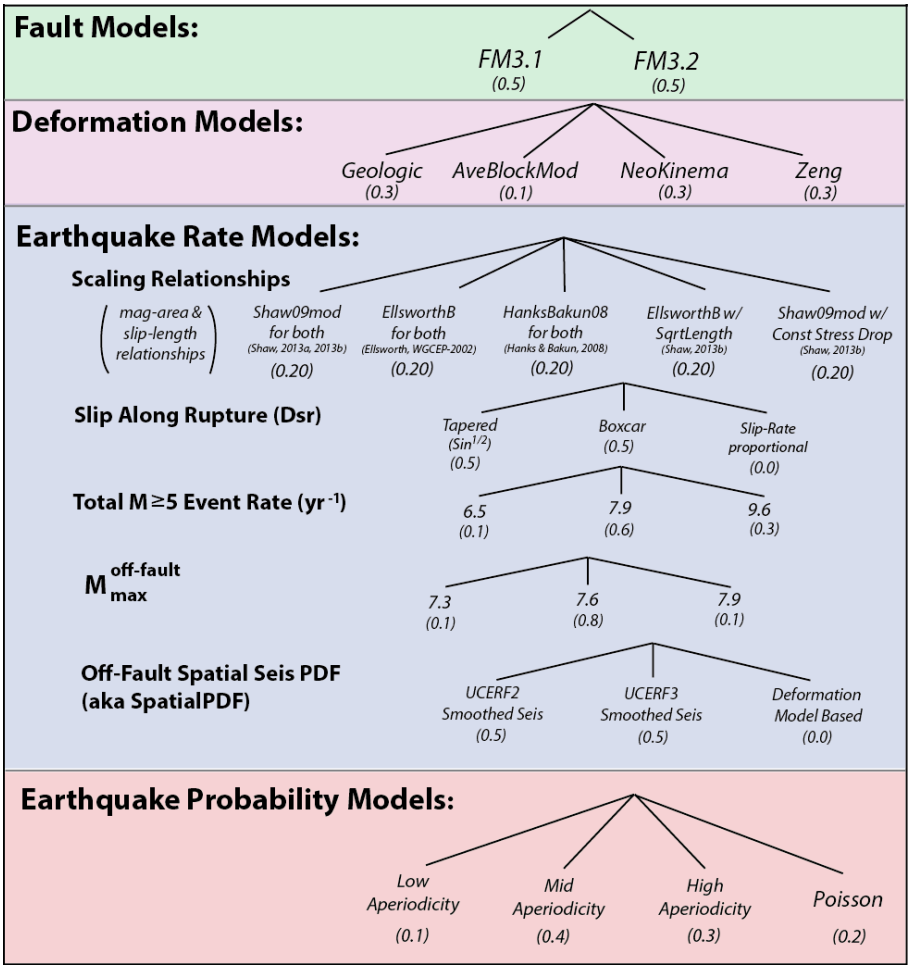

Figure 1. UCERF3 logic-tree branches

OBJECTIVES

To find a reduced-order UCERF3-TD model we use a statewide loss measure: expected annualized loss (EAL) in terms of cost to repair earthquake damage to a building portfolio that approximates the California building stock. We could have considered other portfolios, for example, portfolios representing buildings in a smaller geographic region such as Los Angeles County, or buildings sharing a particular set of features such as highrise pre-1994 welded steel moment-frame buildings, or assets other than buildings such as human casualties. Other portfolios could produce different results, such as displaying as greater sensitivity to UCERF3 branches that particularly affect Los Angeles or that particular affect long-period motion. The methodology presented here can be readily applied to those other portfolios, but we have selected this one for its statewide relevance.

We seek to find a reduced-order model that approximates the full model such that the probability distribution of EAL with the reduced-order model adequately approximates the

probability distribution of EAL with the full model. In particular, we will require that the reduced model and full model pass a two-sample Kolmogorov-Smirnov test at the 1% significance level as in Equation (1). We will also apply what one might call a main-effects test and a variance test, where the mean and coefficient of variation of the reduced model differs from those of the full set by less than 5%, as in Equations (2) and (3):

$$D_n \le 1.63 \cdot \sqrt{\frac{N_B^i + N_B}{N_B^i \cdot N_B}} \tag{1}$$

$$\left|e_\mu\right| \le 0.05 \tag{2}$$

$$\left|e_\delta\right| \le 0.05 \tag{3}$$

The error terms in Equations (1) through (3) are defined in Equations (4) through (6):

$$D_n = \max\left(\left|F_{EAL}^i\left(eal\right) - F_{EAL}\left(eal\right)\right|\right) \tag{4}$$

$$e_\mu = \frac{\mu^i - \mu}{\mu} \tag{5}$$

$$e_\delta = \frac{\delta^i - \delta}{\delta} \tag{6}$$

In the equations, D_n is referred to as the Kolmogorov-Smirnov statistic, $F_{EAL}(eal)$ denotes the cumulative distribution function (CDF) of EAL under the full model evaluated at any particular value *eal*, as in Equation (7). $F^i_{EAL}(eal)$ is that of the reduced-order model after fixing i of its parameters, N_B is the size of the full model (the number of its leafs, which here is 57,600), N^i_B is the size of the reduced-order model, μ denotes the expected value of the *EAL* of the full model per Equation (8), μ^i denotes that of the reduced-order model, δ is the coefficient of variation of the full model per Equation (9), δ^i is that of the reduced-order model, and e_μ and e_δ are error terms for the mean and coefficient of variation. Equation (1) is not exactly how the Kolmogorov-Smirnov test works because samples in the test are supposed to be equiprobable and here they are not. Still, Equation (1) seems close enough for practical purposes.

$$F_{EAL}\left(eal\right) = \sum_{b=0}^{N_B-1} I\left(eal - EAL_b\right) \cdot P\left[b\right] \tag{7}$$

$$\mu = \sum_{b=0}^{N_B-1} EAL_b \cdot P\left[b\right] \tag{8}$$

$$\delta = \frac{\sqrt{\left(\sum_{b=0}^{N_B - 1} \left(EAL_b \right)^2 \cdot P[b] \right) - \mu^2}}{\mu} \tag{9}$$

In Equations (7) through (9), $I(x) = 0$ if $x \leq 0$, 1 if $x > 0$ and $P[b]$ is the weight (the Bayesian probability) of leaf b, evaluated as the product of all the conditional probabilities that lead to it (here, the probabilities in). When one fixes a branch in the reduced model, its conditional probability is taken as 1.0 rather than the weight shown in Figure 1. When applied to the reduced-order model, the summations in Equations (7) through (9) only include the leafs in that model.

METHOD

COMPUTING THE DEPENDENT VARIABLE

Let us first define the terms in Equation (10) then detail the modeling assumptions.

$$EAL_b = \sum_{i=0}^{I-1} V_i \cdot y_i(s) \left| \frac{dG_{i,b}(s)}{ds} \right| ds \tag{10}$$

In the equation, EAL_b denotes the expected value (the average) loss in a year. Here, we acknowledge that *EAL* can be uncertain. It depends in part on one's assumptions about seismic hazard, which leaf b of the UCERF3 logic tree one is using to calculate *EAL*. The parameter i is an index to a particular asset—in this case, a group of collocated buildings having a common vulnerability function, that is, a common relationship between shaking intensity (denoted by s and measured using any convenient intensity measure) and loss. The portfolio contains I assets. The parameter V_i denotes the value of asset i, which we measure here as the estimated replacement cost new (RCN, a term of art borrowed from property valuation for tax assessors), of the group of buildings. We denote by $y_i(s_i)$ the seismic vulnerability function for asset i evaluated at the shaking it experiences, s_i. In a portfolio risk analysis, one almost always employs a set of seismic vulnerability functions that define assets in terms of a small set of readily observable features such as structural material, lateral force resisting system, range of heights, and era of construction. The term $G(s)$ here denotes the hazard curve, by which we mean the mean rate at which the location of asset i experiences shaking equal to or greater than s, in events per year.

Let us approach the problem of finding the reduced-order model two ways: first, the same as in Porter et al. (2012)—by tornado-diagram analysis—and also a new way. Let us refer to the new way as a probabilistic model-order-reduction search.

MODEL ORDER REDUCTION METHOD 1, TORNADO DIAGRAM

As described by Howard (1988), a tornado diagram analysis operates on a function of real numbers, meaning numbers that have order and scale. One estimates a best-estimate, lower-bound, and upper-bound value of each input parameter. The vector of best-estimate parameter values is referred to as the baseline vector. One calculates the dependent variable by evaluating the function at the baseline vector, then twice for each independent variable: the baseline vector except that one independent variable at its lower-bound value and again the baseline vector except with that one independent variable at its upper-bound value. The absolute value of the difference between the dependent variable at the lower- and upper-bound value for a given independent variable is a measure of the sensitivity of the dependent variable to that independent variable. It is referred to as the swing associated with that independent variable. Independent variables are sorted in decreasing order of swing and a horizontal bar chart is created. The horizontal axis measures the dependent variable. Each bar corresponds to one independent variable. The topmost bar represents the independent variable with the highest swing, and the others arranged below in decreasing order of swing. The left and right ends of each bar are placed at the values of the dependent variable corresponding to the lower- and upper-bound values of the independent variable that was varied. A vertical line is drawn through the value of the dependent variable corresponding to the baseline vector.

In the case of a function of nominal numbers (as opposed to real numbers), Howard's (1988) approach does not work: without order or scale, the input parameters have no lower or upper bound or best estimate. One has to find a different way to select the baseline vector and the bounds of each independent variable. In our adaptation of Howard's approach, one calculates the dependent variable at all positions in the possibility space, associates each with a probability (the products of the Bayesian probability of each parameter value), calculates the cumulative distribution function of the dependent variable under the full model, and selects as the baseline vector the position in the possibility space that produces a value of the dependent variable that is closest to its the mean value. Notice how we select the baseline vector based on the dependent variable rather than based on the independent variables, which have no means.

One then fixes all parameter values but one at their baseline value, varies the remaining parameter through all its possible values, and finds the difference between the maximum and minimum value of the performance measure in this small subset. That difference is taken as the swing to represents the sensitivity of the dependent variable to the varied parameter. Recall that under Howard's approach one would have selected upper- and lower-bound values of each independent variables and then evaluated the swing in the dependent variable as the difference in the function evaluated under each bound. Instead, one evaluates the dependent variable at all the possible values of one independent variable, finds the minimum and maximum values of the dependent variable, and identifies the values of the independent variable that produced those extrema. One applies the minimum and maximum functions, defined only for ordinal and cardinal numbers, on the dependent variable rather than independent variables.

One repeats the process for all of the independent variables, each time varying only one and leaving all the others at their baseline value, and determines the swing associated with each parameter. As before, independent variables are ordered in decreasing swing and a horizontal bar chart is constructed, each horizontal bar bounded by the minimum and maximum values of the dependent variable associated with varying that particular independent variable. The bar with the greatest swing is topmost, the one with the smallest swing at the bottom.

One then constructs cumulative distribution functions of the dependent variable allowing only the topmost independent variable to vary, then allowing the top 2 to vary, then the top 3, etc. The probability mass of each leaf is taken as the product of the Bayesian probabilities of just the values of the independent variables that are allowed to vary. The aim is to find the smallest set of independent variables for which the dependent variable's cumulative distribution function resembles that of the full model. The resemblance is adequate when the mean and coefficient of variation of the reduced-order model are within some small difference of those of the full model, say 5%, and the two CDFs pass a Kolmogorov-Smirnov goodness of fit test at say the 5% significance level.

In our analysis of UCERF2, we produced the tornado diagram duplicated in Figure 2A and selected a reduced-order model that varies only the top 3 independent variables, producing the cumulative distribution functions in Figure 2B. The figures show that when we applied our modification of Howard's tornado-diagram method to UCERF2, we found that if one were interested in the PDF of statewide annualized economic loss or loss of life, a PSRA that varies only the probability model, ground motion prediction equation, and magnitude-area

relationship (representing only 2% of the branches) produced essentially the same result as using the full tree, in terms of the mean and standard deviation of the PDF of loss.

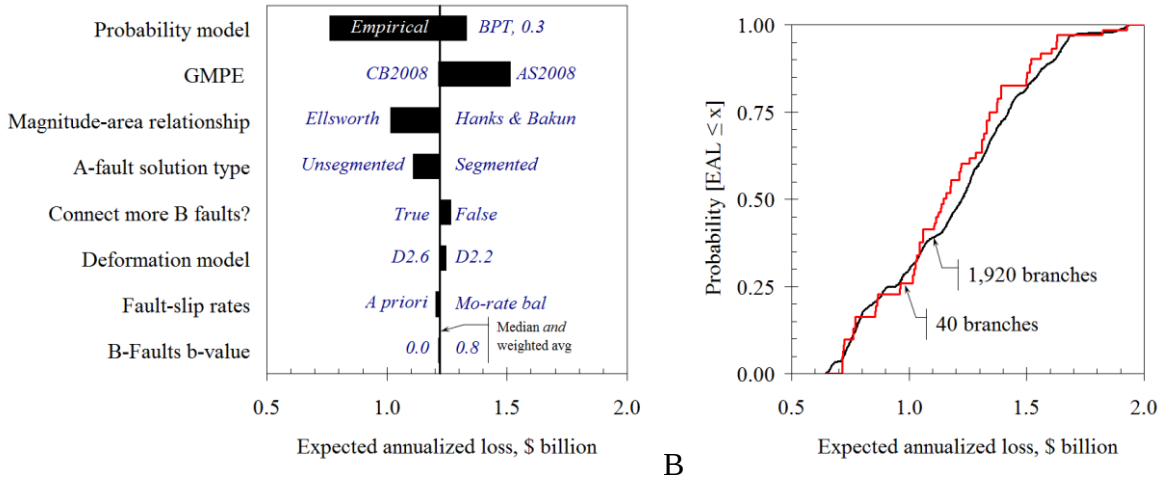

Figure 2. A. Example tornado diagram. B. K-S test (Porter et al. 2012) showing expected annualized loss (EAL) under the full model (labeled "1,920 branches") and the reduced-order model (labeled "40 branches")

MODEL ORDER REDUCTION METHOD 2, PATH SEARCH

We now introduce a new search algorithm. It is intended for a model whose parameters are all nominal numbers. It is also intended to select the baseline vector more systematically than a tornado diagram analysis with all nominal numbers. The new algorithm starts by varying all independent variables, then fixes one, then another, etc., until one has fixed all but one independent variable. This approach resembles so-called one-at-a-time (OAT) sampling, but with a deliberate and ordered selection of the independent variables to be fixed. The sequence of independent variables that are fixed is called the path. One selects the smallest reduced model along that path that still satisfies Equations (1) through (3). One does not fix the last independent variable because with only one leaf, one cannot satisfy Equation (3). Here is how the path search works.

1. Evaluate the full model at each leaf in the logic tree.
2. Evaluate the CDF, mean, and coefficient of variation for the full model. Here, that means Equations (7) through (9).
3. List the independent variables. Fix the first at one if its possible values, allowing other independent variables to vary. Evaluate Equations (7) through (9) of the reduced model. Let i denote an index to independent variables and j denote an index to possible values of an independent variable. For each (i, j) pair, calculate $D^{i,j}_{max}$,

defined as in Equation 12, as well as e_μ and e_δ. We operate on the maximum difference rather than some average because we aim to satisfy Equation (1), that is, because we are applying the Kolmogorov-Smirnov goodness-of-fit test.

$$D_{\max}^{i,j} = \max\left(\left|F_{EAL}^{i,j}\left(eal\right) - F_{EAL}\left(eal\right)\right|\right) \tag{11}$$

4. Select the (*i*, *j*) pair with the lowest value of $D^{i,j}{}_{max}$. Fix independent variable *i* at value *j*.
5. Repeat steps 2 through 4 starting with reduced model, successively fixing each of the still-varying parameters at each possible value.
6. Repeat step 5 until all but one branch is fixed.
7. Find the smallest model (i.e., the one with the fewest varying parameters) that satisfies the objectives, which here means Equations (1) through (3).

MODEL ORDER REDUCTION METHOD 3, GRID SEARCH

One could in principle try a third option for finding the best reduced-order model: one could search every possible reduced-order model and select the one with the leaves that still satisfies the objectives, which here means satisfying Equations (1) through (3). Such a grid search would require evaluating *N* reduced-order models:

$$N = 1 + \left(\sum_{b=1}^{B} n_b\right) + \left(\sum_{b=1}^{B}\left(n_b \cdot \sum_{c=b+1}^{B} n_c\right)\right) + \left(\sum_{b=1}^{B}\left(n_b \cdot \sum_{c=b+1}^{B}\left(n_c \cdot \sum_{d=c+1}^{B} n_d\right)\right)\right) + \ldots \tag{12}$$

where n_i is the number of possible values of parameter *i* and *B* is the number of parameters. The first summand (1) represents the number of models with 0 fixed parameters, that is, with all parameters allowed to vary; it is the number of full models in the full model, which is one. The second summand represents the number of models with 1 fixed parameter and all the others allowed to vary. If we only fix parameter *b*, there are n_b reduced-order models to evaluate. The third summand represents the number of models with 2 fixed parameters, and so on.

APPLICATION

TWO MORE HAZARD PARAMETERS, SOFTWARE, AND PORTFOLIO

In addition to the UCERF3 logic-tree branches, we considered two models of Vs30: that of Wills and Clahan (2006), which is based on Vs30 measurements of various geologic units, and that of Wald and Allen (2007), which is based on an empirical relationship between topographic slope and Vs30. We considered five NGAWest-2 ground motion prediction

equations: Abrahamson et al. (2014, abbreviated here as "ASK2014"), Boore et al. (2014, "BSSA2014"), Campbell and Bozorgnia (2014, "CB2014"), Chiou and Youngs (2014, "CY2014"), and Idriss (2014, "IDR2014").

We used the OpenSHA EAL calculator to evaluate Equation (10) for a portfolio of California buildings, measuring loss in terms of building repair cost. The calculator estimates the average loss in a year to a specific set of buildings (catastrophe risk modelers usually refer to the set of buildings as a portfolio). Let us look back again at Equation (10), and discuss modeling assumptions involved in evaluating I, V_i, $y_i(s)$, and $G_{i,b}(s)$. We will not vary the portfolio here or otherwise address any uncertainty in I, V, or y because we are primarily interested in learning more about hazard $G_{i,b}(s)$.

We use a portfolio that approximates most (though not all) of the building stock in California. The portfolio was constructed using FEMA's Hazus-MH 2.1 software, the same portfolio used in Porter et al. (2012), with some updates. That earlier portfolio was based on the Hazus-MH 2002 inventory and reflected 2002 prices. We factored up values to account for population growth. California's population increased between 2010 and 2013 from 35,116,033 to 38,041,430, an increase of 8.3%. We also factored up values to account for price changes. RSMeans' 30-city historical cost index in January 2013 was 197.6. Its January 2002 value was 126.7, suggesting a price increase over 2002 of 56%. Thus, the portfolio was scaled up from 2002 to approximate 2013 values by 69% ($1.56 \cdot 1.083 = 1.69$). The update does not reflect changes in the geographic distribution of the population. To reduce computational effort, we omitted from analysis any asset (combination of census tract and vulnerability function) with a value less than $1 million, which reduces the number of assets and the computational effort by 90% and the total value by 10%. We used the vulnerability functions $y_i(s)$ presented in Porter (2009).

COMPUTATIONAL EFFORT

The calculation of EAL for the CEA proxy portfolio for all branches including four probability models, five ground-motion prediction equations, and two site-class models took 275,000 CPU hours on TACC Stampede (150 wall clock hours). Note that the analysis does not scale with the number of probability models. The software calculates expected portfolio loss on a rupture-by-rupture basis and postmultiplies by event rates. This new efficient approach only duplicates work when key rupture properties (magnitude, area due to aseismicity, and rake) change. Only five branch levels affect these quantities: fault model,

deformation model, scaling relationships, ground-motion prediction equation, and site-class model. All of the other branch levels are in effect free (not requiring additional calculation of shaking or portfolio loss) as they only affect rupture rates, which is handled in post processing.

Although computational time is not a simple function of logic tree branches in the most recent version of the software, it is worthwhile measuring reductions in computation effort for the brute-force method of treating each branch independently, since that is what many UCERF3 users are likely to do.

FINDINGS

GENERAL FINDINGS

Hazard uncertainties make EAL appear approximately lognormal. Uncertainties in UCERF3, selection of ground motion prediction equation, and choice of Vs30 model make statewide EAL uncertain. Its uncertain quantity is reasonably approximated by a lognormal random variable, as illustrated in Figure 3. The lognormal distribution has a median value of $4.21 billion and the natural logarithm of EAL has a standard deviation equal to 0.21. The $4.2 billion figure generally agrees with the FEMA (2008) figure of $3.5 billion per year. The higher value is largely attributable to the fact that we have updated the portfolio value to account for population growth and inflation to 2013, and to differences in the hazard models.

(Here and henceforth the EAL to which we refer represents loss to 90% of the estimated statewide building replacement cost, so a better estimate of statewide earthquake repair cost EAL would be a lognornal random variable with median value of $4.67 billion and logarithmic standard deviation of 0.21, but let us leave that detail aside. This paper is about UCERF3, not EAL, and the constant underestimate by a factor of 0.9 seems unimportant.)

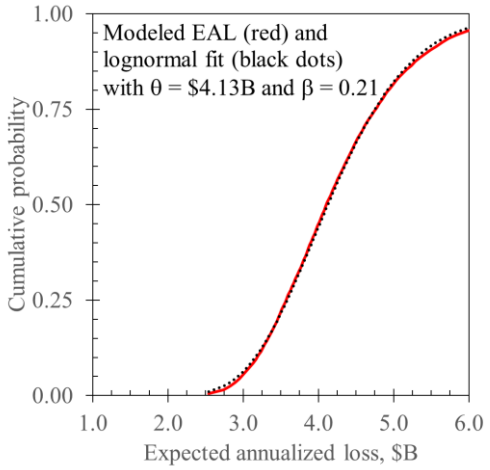

Figure 3. Marginal cumulative distribution function (red) is approximately lognormal (black), passing the Lilliefors goodness-of-fit test at the 5% significance level.

Background sources contribute substantially to loss. UCERF3 fault-based sources (as opposed to gridded background seismicity) contribute an estimated 73% to 89% of EAL, depending on logic-tree leaf, with a weighted average contribution of 82% and no obviously similar parametric probability distribution. See Figure 4. The implication is that background sources contribute significantly to EAL and with an uncertain degree of contribution. Background sources must be included in the full model.

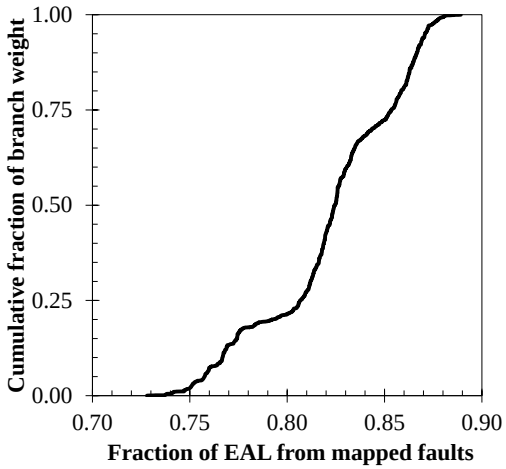

Figure 4. Contribution to EAL from fault-based sources (excluding background gridded seismicity).

FINDINGS OF THE TORNADO-DIAGRAM ANALYSIS

The tornado-diagram analysis indicates that the top contributors to uncertainty in statewide EAL are the total rate of M 5 earthquakes, the probability model, the scaling relationship, and selection of ground motion prediction equation, at least when one selects the baseline values shown in Figure 5. Baseline values are shown in italics. The least important uncertainties with

these baseline values are the fault model, the distribution of slip along rupture, the deformation model, and the spatial seismicity probability density function. Yellow italics (baseline values of total M 5 rate, probability model, and deformation model) indicate that the baseline is the parameter value with the highest Bayesian probability in the logic tree weighting—its probability in Bayesian terms. Other baseline values have probability approximately equal to their alternatives.

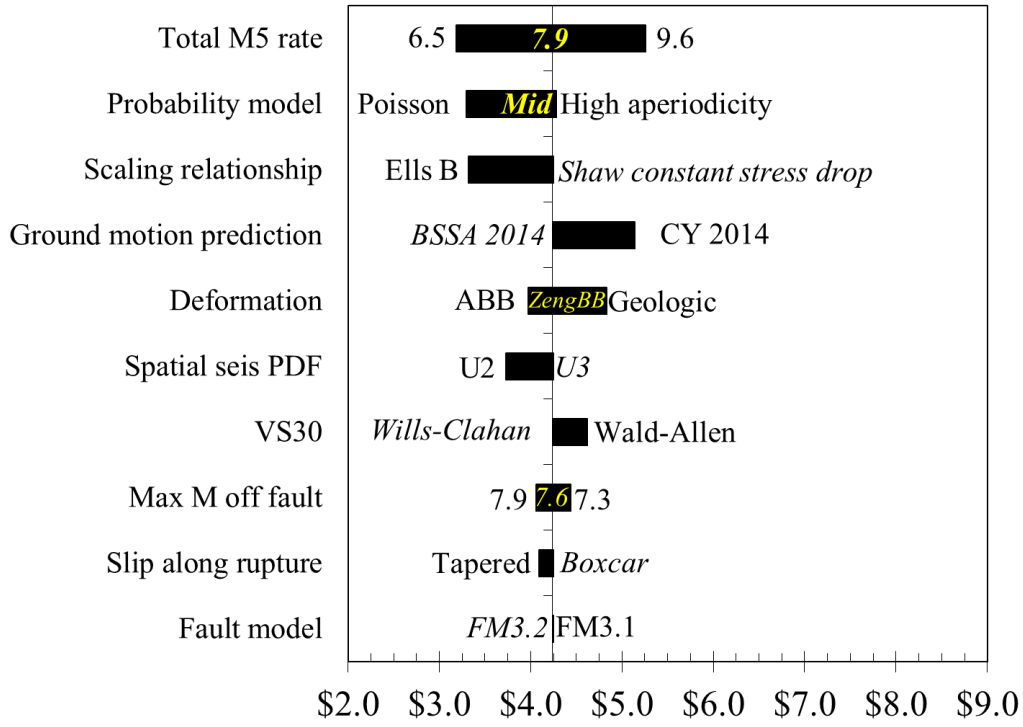

Figure 5. Tornado diagram suggests the top contributors to uncertainty in expected annualized loss are total magnitude 5 rate, ground motion prediction equations, scaling relationship, and probability model.

If one allows only the top 5 independent variables in Figure 5 to vary according to the Bayesian probabilities, fixing the other 5 at their baseline value with 100% Bayesian probability, one can evaluates the resulting cumulative distribution function of EAL and compare it with that of the full tree. See Figure 6. In the figure, θ and β denote the parameters of the lognormal distribution (median and standard deviation of the natural logarithm respectively), and μ and δ denote the mean and coefficient of variation respectively. The reduced model exhibits error terms D_n = 0.046, e_μ = 2% and e_δ = -4%, satisfying Equations (1) through (3). All of which suggests low bias and reasonable agreement between this particular subset model and a full model, for 2% of the computational effort.

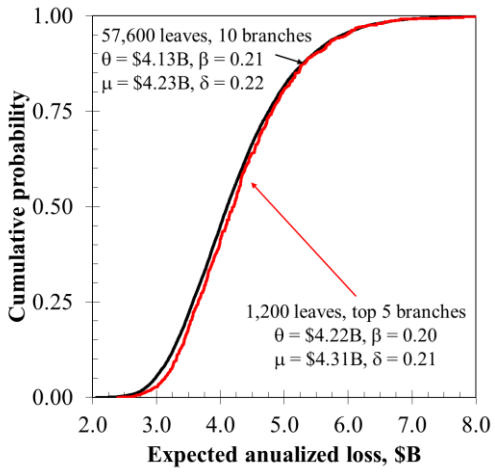

Figure 6. Cumulative distribution function of EAL with the full model (black, smooth curve) and the reduced model (red, steps). The reduced model passes a two-parameter Kolmogorov-Smirnov goodness of fit test for agreement with the full model, despite the slight bias high.

Allowing additional independent variables to vary does not necessarily cause e_μ to asymptote to zero. Notice how the bar for the spatial PDF in Figure 5 has its baseline value coincident with its upper bound value, and how the swing is nearly as high as that of the next-higher independent variable. These observations suggest that if we allow the spatial PDF to vary, e_μ should decrease. It does, from +2% to -4%, as shown Table 1. Allowing additional independent variables to vary does seem to cause e_δ to approach asymptote to zero as shown in Table 1.

Why do the two bias measures not monotonically asymptote to zero as more parameters are allowed to vary? Two explanations suggest themselves: (1) particular parameter values will invariably produce higher or lower results than do other values of the same parameter, or they will produce higher or lower variability. One might fix one parameter at a value that tends to produce higher mean results, and another at a value that tends to produce lower mean results. The two fixed parameters would tend to offset each other, resulting in a low mean bias. If one then frees one of those two parameters to vary, the result will tend to shift higher or lower, increasing mean bias. A similar statement can be made about uncertainty bias. (2) More generally, parameters might interact, so allowing more of them to vary could produce unpredictable interactions. Finding a good fixed set and good baseline values for it—values that result in low mean bias and low uncertainty bias—can require some luck.

Table 1. Allowing more parameters to vary does not cause bias to monotonically decrease

	Top 1	Top 2	Top 3	Top 4	Top 5	Top 6	Top 7	Top 8	Top 9	Full set
Leaves	3	12	60	300	1,200	2,400	4,800	14,400	28,800	57,600
Effort	0.01%	0.02%	0.10%	0.52%	2.1%	4.2%	8.3%	25.0%	50.0%	100%
E[EAL]	$4.44	$4.23	$3.86	$4.22	$4.31	$4.04	$4.27	$4.27	$4.23	$4.23
σ	$0.62	$0.70	$0.67	$0.76	$0.89	$0.88	$0.95	$0.95	$0.92	$0.91
δ	0.14	0.17	0.17	0.18	0.21	0.22	0.22	0.22	0.22	0.22
θ	$4.40	$4.17	$3.80	$4.15	$4.22	$3.95	$4.17	$4.18	$4.13	$4.14
β	0.14	0.17	0.17	0.18	0.20	0.21	0.21	0.21	0.21	0.21
D_n	0.45	0.91	0.28	0.05	0.05	0.11	0.02	0.02	0.02	0.00
e_μ	5%	0%	-9%	0%	2%	-4%	1%	1%	0%	0%
e_δ	-36%	-23%	-20%	-16%	-4%	0%	3%	3%	1%	0%

Perhaps because parameters interact, the order of contributors from top to bottom in a tornado diagram displays sensitivity to the selection of the baseline vector values. We checked several reasonable baseline vectors from samples whose EAL was near the mean of the distribution shown in Figure 3. See two of the resulting alternative tornado diagrams in Figure 7. The alternative diagrams show that the order of parameters, while it does vary from baseline vector to baseline vector, does tend to have the same top parameters and the same bottom parameters.

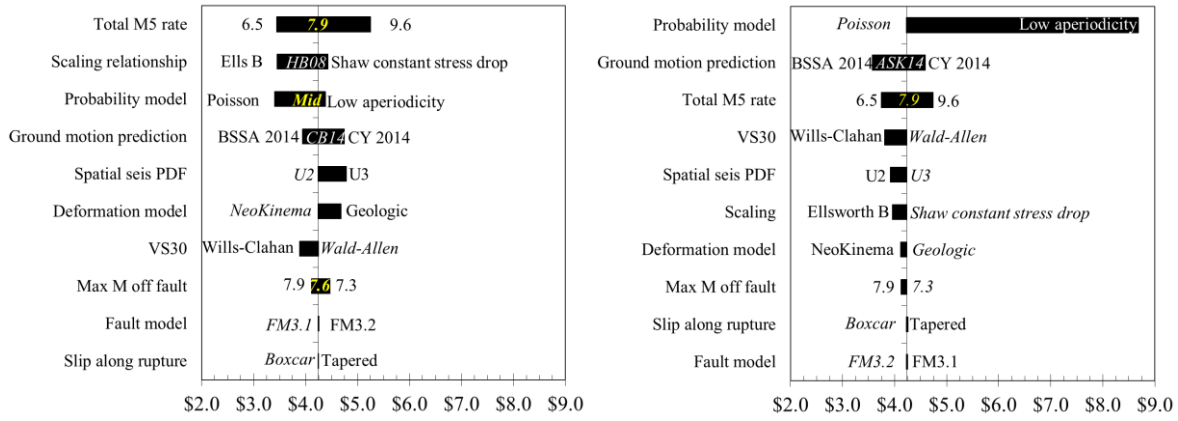

Figure 7. Tornado diagrams are sensitive to the baseline vector. Shown here are two alternative tornado diagrams selected from leaves whose EAL was near the median

The difference between these diagrams begs the question, is there a better way to select the baseline vector? "Better" would be that it produces a reduced model whose bias more reliably asymptotes to zero as more parameters are allowed to vary, or whose baseline vector can be selected more systematically. The question led to our developing the probabilistic model-reduction search algorithm, whose results are discussed next.

FINDINGS OF THE PATH SEARCH

The smallest reduced model that satisfies all three of the similar-distribution, main-effects, and variance tests in Equations (1) through (3) fixes six of ten parameters. The parameters were fixed in the following order, and at these values:

1. Max M off fault: 7.6 (solution 1) or 7.9 (solution 2, better)
2. Fault model: FM3.1
3. Slip along rupture: boxcar
4. Deformation model: NeoKinema
5. Ground motion prediction: ASK2014
6. Probability model: mid

Fixing these parameters allows the following to vary:

7. Scaling relationship: 5 values
8. Total $M \geq 5.0$ rate: 3 values
9. Spatial PDF: 2 values
10. Vs30: 2 values

The reduced model has 60 leaves (0.1% of the size of the full model), $D_n = 0.14$, $e_\mu = 3.5\%$ and $e_\delta = 4.9\%$, satisfying Equations (1) through (3). It also passes the two-parameter K-S test at both the 1% and 5% significance levels. The resulting CDF is shown in Figure 8A. The solution has a slight bias toward higher EAL. Is it necessarily the best 60-leaf solution? Probably not, since the search was not exhaustive, which the grid search would have been. In fact, we found a slightly better solution by examining the tornado diagrams (Figure 7) and noticing that mean EAL can be reduced by selecting maximum magnitude off fault = 7.9 rather than 7.6. That solution has the same number of leaves (60) but lower error terms: $D_n = 0.051$, $e_\mu = 1.3\%$ and $e_\delta = 3.8\%$. See Figure 8B for its CDF.

We considered the grid-search option after completing the incremental search. In light of the apparent success of the incremental search, and considering the computational effort required for the grid search, we did not pursue the grid-search option. We may pursue it in future research.

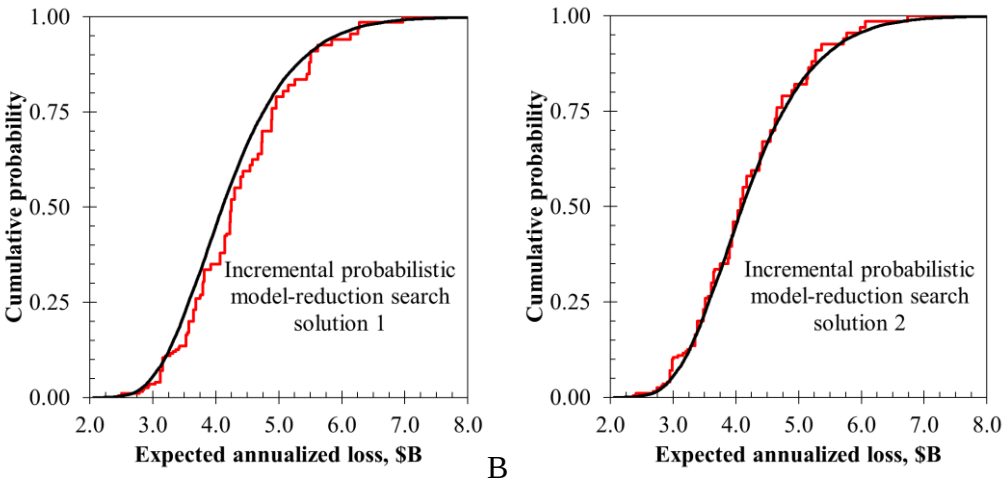

Figure 8. Cumulative distribution function of the full model (black) and reduced models (red) by incremental probabilistic model-reduction search: (a) solution 1, and (b) slightly superior solution 2

FINDINGS OF THE GRID SEARCH

In the present problem, the grid search would require examining N = 1,166,402 reduced-order models. That is, to find the best reduced-order model, one would evaluate the CDF, mean, and coefficient of variation of EAL for more than a million reduced-order models and choose the smallest model that satisfies Equations (1) through (3). We have identified this third option, shown how to calculate the size of the problem, and calculated it for the present application, all of which seem useful. We have not yet performed the grid search, reserving that work for a later study, when we will have leisure to address a number of related questions, such as, if the path search produces the same answer as the grid search, why? If not, why not? Under what conditions will the two produce the same or different answers?

CONCLUSIONS

The UCERF3 logic tree contains eight uncertain modeling decisions, often referred to as branches, each of which involves a choice between two or more discrete model elements. Each option of each branch has an associated weight (or probability in Bayesian terms) assigned by the UCERF3 developers. The combinations of branch choices—one can think of them as leaves in the logic tree—total 5,760. If one wishes to use the UCERF3 model to estimate seismic hazard (in the sense of exceedance frequency of various levels of ground motion at a particular location), one most likely must also choose a ground motion prediction equation and a model of the average shearwave velocity in the upper 30 meters of soil, denoted here by Vs30. (An alternative is to use physics-based modeling, which requires a 3-D crustal velocity model rather

than a ground motion prediction equation like those of NGAWest-2.) If one selects among five NGAWest-2 ground motion prediction equations and two models of Vs30, then one must deal with a total of 10 branches and 57,600 leaves.

The size of the full UCERF3 model and the added Vs30 and ground motion prediction equation uncertainties creates a problem. To perform a probabilistic seismic hazard analysis of even a few sites and account for these 10 modeling uncertainties can take a prohibitive amount of time. To perform a probabilistic risk analysis of even a modest portfolio of assets can similarly take an impractical amount of time.

We therefore sought to reduce the model (that is, to trim the UCERF3-TD hazard tree) to find just those parameters that really matter, allowing them to vary and fixing the other parameters at a single baseline value each. By "really matter" we mean those that contribute significantly to uncertainty in hazard. We do not mean anything about their scientific importance. In particular, we set out to find the smallest subset of varying parameters that, when the other parameters are fixed at baseline values, produces a reduced model that closely resembles the full model, at least in one uncertain output. In our case, we selected as the output measure the expected annualized building repair cost to an inventory that approximates all the buildings in California. We applied three tests to check the similarity of the reduced model to the full model: a main-effects test to ensure low bias in the mean of the reduced model relative to the full model; a variance test to ensure low bias in the coefficient of variation of the reduced model relative to the full model, and a test of the similarity of distribution.

We used a statewide inventory derived from the Hazus-MH estimate of 2002, updated to 2013 to account for population growth and inflation. We used a set of vulnerability functions derived from the Hazus-MH model. All of the EAL calculations were performed in an OpenSHA calculator, not in Hazus-MH.

We employed two alternative model-reduction methods: a tornado-diagram analysis, which is a kind of deterministic sensitivity study, and a new incremental probabilistic model-reduction search algorithm (generically called a one-at-a-time or OAT approach), which we introduce here. Our model-reduction algorithm is unusual in that it works for a model that comprises only nominal parameters, that is, parameters that are neither cardinal (having measure) nor ordinal (having order), but are merely labels (for example, fault models 3.1 and 3.2). The algorithm appears to be new; at least we could not find a similar one in the literature.

The new algorithm produced a smaller model with just four varying parameters and six fixed ones, for a total of 60 leaves (0.1% of the size of the full model of 57,600 leaves), compared with the five-parameter, 1,200-leaf model that we found using a tornado diagram. These parameters vary:

1. Scaling relationship: 5 values
2. Total $M \geq 5.0$ rate: 3 values
3. Spatial PDF: 2 values
4. Vs30: 2 values

These parameters are fixed:

5. Max M off fault: 7.9
6. Fault model: FM3.1
7. Slip along rupture: boxcar
8. Deformation model: NeoKinema
9. Ground motion prediction: Abrahamson et al. (2014)
10. Probability model: mid

The reduced model has a Kolmogorov-Smirnov statistic $D_n = 0.051$ (less than the allowable maximum of 0.17 for samples of 60 and 57,600 at the 5% significance level), mean bias (difference between mean loss under the reduced model as a fraction of that of the full model) $e_\mu = 1.3\%$ and bias in the coefficient of variation (again as a fraction of that of the full model) $e_\delta = 3.8\%$. See Figure 8B for a comparison of the two CDFs.

LIMITATIONS

We trimmed the UCERF3-TD logic tree using a statewide loss measure. The trimmed tree may only be appropriate for portfolios that are distributed similarly to the statewide population. We make no value judgments about the scientific importance of the fixed parameters, only about how they affect statewide building-repair-cost EAL. For simplicity, we apply the two-sample Kolmogorov-Smirnov test to CDFs of the full population of two uncertain discrete variables, whereas it is meant to test CDFs constructed from equiprobable samples of continuous variables. We have not increased the uncertainty in any model element (e.g., ground motion prediction equations) to account for the possibility that nature may actually behave outside the bounds of the modeled options. We do not speculate on the possible consequences of ignoring such exogenous uncertainties.

ACKNOWLEDGMENTS

This research was supported by the Southern California Earthquake Center. SCEC is funded by NSF Cooperative Agreement EAR-0529922 and USGS Cooperative Agreement 07HQAG0008. Computation for the work described in this paper was supported by The University of Texas at Austin (http://www.tacc.utexas.edu), use of which was provided through the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1053575. Any use of trade, firm, or product names is for descriptive purposes only, and does not imply endorsement by the U.S. Government.

REFERENCES

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

Amsallem, D., and C. Farhat, 2008. Interpolation method for adapting reduced-order models and application to aeroelasticity. *AIAA journal*, 46 (7): 1803-1813.

Antoulas, A.C., 2009. An overview of model reduction methods and a new result. *Proceedings of the 48th IEEE Conference on Decision and Control*, Shanghai, 15-18 Dec 2009, p. 5357-5361.

Bai, Z., 2002. Krylov subspace techniques for reduced-order modeling of large-scale dynamical systems. *Applied Numerical Mathematics*, 43 (1): 9-44.

Benner, P., S. Gugercin, and K. Willcox, 2015. A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM Review*, 57 (4): 483-531.

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

Campbell, K.W. and Y. Bozorgnia, 2014. NGA-West2 ground motion model for the average horizontal components of PGA, PGV, and 5% damped linear acceleration response spectra. *Earthquake Spectra* 30 (3), 1087-1115

Chatterjee, A., 2000. An introduction to the proper orthogonal decomposition. *Current Science*, 78 (7), 808-817

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

Cornell, C.A., 1968. Engineering seismic risk analysis. *Bulletin of the Seismological Society of America*. 58 (5), 1583-1606.

Esteva, L., 1967. Criteria for the construction of spectra for seismic design. In *Third Panamerican Symposium on Structures, Caracas, Venezuela* (vol. 1082).

FEMA, 2008. *HAZUS-MH Estimated Annualized Earthquake Losses for the United States*. Report FEMA 366, Washington DC, 53 pp.

Field, E. H., R. J. Arrowsmith, G. P. Biasi, P. Bird, T. E. Dawson, K. R. Felzer, D. D. Jackson, K. M. Johnson, T. H. Jordan, C. Madden, 2015. Long-term, time-dependent probabilities for the third Uniform California Earthquake Rupture Forecast (UCERF3), Bull. Seismol. Soc. Am. 105, 511–543, doi: 10.1785/0120140093.

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, 2009. Uniform California Earthquake Rupture Forecast, Version 2 (UCERF 2). *Bulletin of the Seismological Society of America*, 99 (4), 2053–2107

Gugercin, S., and A.C. Antoulas, 2004. A survey of model reduction by balanced truncation and some new results. *International Journal of Control*, 77 (8): 748-766

Howard, R.A., 1988. Decision analysis: practice and promise. *Management Science* 34 (6), 679-695. http://goo.gl/0LVJaA [accessed 22 Aug 2016]

Idriss, I.M., 2014. An NGA-West2 empirical model for estimating the horizontal spectral values generated by shallow crustal earthquakes. *Earthquake Spectra* 30 (3), 1155-1177

Mayo, A.J., and A.C. Antoulas, 2007. A framework for the solution of the generalized realization problem. *Linear Algebra and its Applications*, 425 (2): 634-662.

Michael, A.J., K.R. Milner, M.T. Page, T. Parsons, P.M. Powers, B.E. Shaw, W.R. Thatcher, R.J. Weldon, and Y. Zeng, 2014. Uniform California Earthquake Rupture Forecast, version 3 (UCERF3): The time-independent model, *Bull. Seismol. Soc. Am*. 104, 1122–1180, doi: 10.1785/0120130164.

Petersen, M. D., M. P. Moschetti, P. M. Powers, C. S. Mueller, K. M. Haller, A.D. Frankel, Y. Zeng, S. Rezaeian, S. C. Harmsen, O. S. Boyd, N. Field, R. Chen, K. S. Rukstales, N. Luco, R. L. Wheeler, R. A. Williams, and A. H. Olsen, 2014. Documentation for the 2014 update of the United States national seismic hazard maps, U.S. Geol. Surv. Open-File Rept. 2014-1091, 243 pp., 10.3133/ofr20141091.

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

Porter, K.A., J.L. Beck, and R.V. Shaikhutdinov, 2002. Sensitivity of building loss estimates to major uncertain variables. *Earthquake Spectra*, 18 (4), 719-743

Powers, P.M. and E.H. Field, 2015. Update to the National Seismic Hazard Model in California. Earthq. Spectra. 31, S177-S200.

Sorensen, D.C., and A.C. Antoulas. 2002. The Sylvester equation and approximate balanced reduction. *Linear Algebra and its Applications*, 351: 671-700.

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

Wills, C.J. and Clahan, K.B., 2006. Developing a map of geologically defined site-conditions categories for California. *Bulletin of the Seismological Society of America*, 96 (4A), pp. 1483-1501.

Working Group on California Earthquake Probabilities (WGCEP), 2007. *The Uniform California Earthquake Rupture Forecast, Version 2 (UCERF 2)*, U.S. Geol. Surv. Open-File Rept. 2007-1437 and California Geol. Surv. Special Rept. 203, http://pubs.usgs.gov/of/2007/1437/ (last accessed December 2011).