16th World Conference on Earthquake Engineering, 16WCEE 2017
Santiago Chile, January 9th to 13th 2017
Paper N°1934
Registration Code: S-F1461943996

DEAGGREGATION OF UNCERTAINTY IN SEISMIC HAZARD

P.M. Powers[(1)], N. Shome[(2)], and M.D. Petersen[(3)]

[(1)] *Research geophysicist, U.S. Geological Survey, pmpowers@usgs.gov*
[(2)] *Risk modeler, Risk Management Solutions, Inc., nilesh.shome@gmail.com*
[(3)] *Research geophysicist, U.S. Geological Survey, mpetersen@usgs.gov*

Abstract

The U.S. National Seismic Hazard Model (NSHM) is composed of multiple logic trees that capture the epistemic uncertainty inherent in any estimate of seismic hazard at a site. Logic trees and associated weights describe, but are not limited to, variability of fault geometry, maximum magnitude, and regional seismicity rate. They are also used to represent different models of ground motion. With each new edition of the NSHM, the number of logic trees and associated branches grows. While this does not pose insurmountable computational demands, the computation of mean hazard maps does grow proportionally, particularly when computing associated uncertainties on a map scale, and it is difficult to determine which branches are the most influential. It is also increasingly important that the NSHM program produce maps of uncertainty; to this end, any reductions in computation time are a benefit. Using the California component of the NHSM, the Uniform California Earthquake Rupture Forecast version 3 (UCERF3), as an example, we examine uncertainty when computing seismic hazard at a site. Whereas analyses of uncertainty commonly use Monte Carlo sampling to compute fractiles about a mean hazard curve, for this analysis we iterate over the entire logic tree. Although this approach maintains correlations between branches, we are able track individual logic tree branch choices through a hazard calculation that would otherwise be obscured. We then deaggregate the range of possible ground motions or return periods and use tornado diagrams and stacked histograms to illustrate the relative influence of each logic tree branch on hazard at a site. In the case of UCERF3, fault slip-rate and ground motion models are the most influential, but models of fault slip distribution and the maximum magnitude of background seismicity sources are not.

Keywords: uncertainty; seismic hazard; seismic risk; PSHA; deaggregation

1. Introduction

Probabilistic seismic hazard analyses typically employ logic trees to represent epistemic uncertainty in earthquake source and ground motion models in order to capture the center, body, and range of earthquake hazard or ground shaking at a site. The earthquake rate model for the California portion of the 2014 U.S. Geological Survey (USGS) National Seismic Hazard Model (NSHM) is based on the long-term, time-independent component of the Uniform California Earthquake Rupture Forecast, version 3.3 (UCERF3; [1]) developed by the Working Group on California Earthquake Probabilities (WGCEP, e.g. [2]), which is responsible for creating and updating earthquake forecasts for California using best-available science. To compute hazard, the 2014 NSHM uses the new NGA-West2 ground motion models (GMMs; [3]).

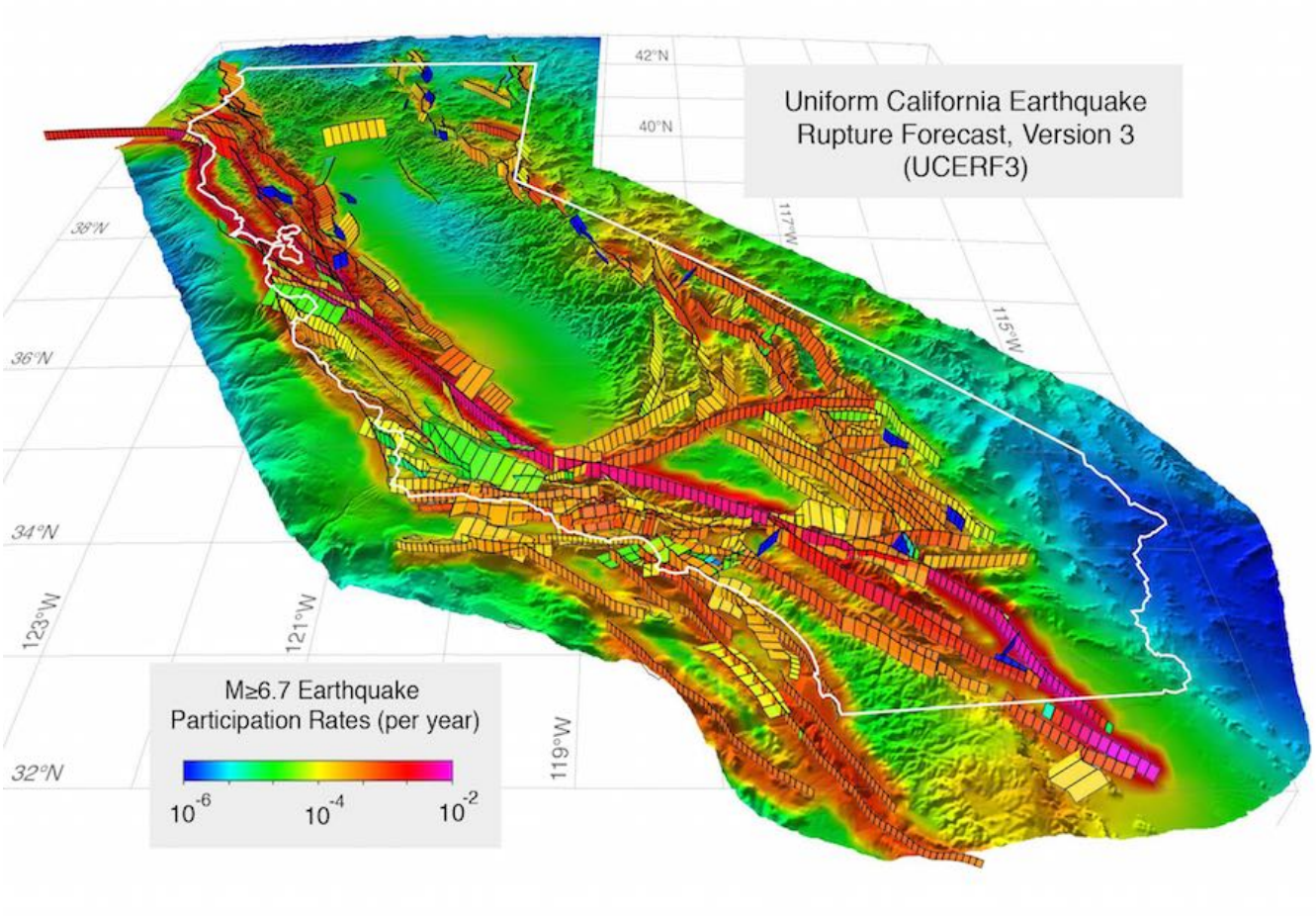

Fig. 1 – Map of **M**≥6.7 earthquake participation rate for the Uniform California Earthquake Rupture Forecast, version 3.3 (UCERF3) model. For the purpose of this illustration, the discretized fault network is projected above the Earth's surface. The color-coded topography indicates the participation rate of gridded, or background, earthquake sources.

UCERF3 (Fig. 1) marks a turning point in the generation of earthquake rate models. The two primary goals of the UCERF3 earthquake-rate model are to relax assumptions of fault-segmentation and include a greater diversity of multi-fault ruptures, both of which were limitations of the previous model (UCERF2, [4]), and the lack of which contributed to an over-prediction of **M**6.5–**M**7 earthquake rates, relative to the long-term statewide historical rate. To achieve these goals, UCERF3 treats the calculation of earthquake rates as an inverse problem, simultaneously solving a system of equations governing the rates of all possible earthquakes, while

adhering to a variety of data constraints. The inversion approach eliminates the apparent over-prediction and permits inclusion of multi-fault ruptures similar to recently observed earthquakes, for example, the 2002 moment-magnitude (**M**) 7.9 Denali, Alaska earthquake.

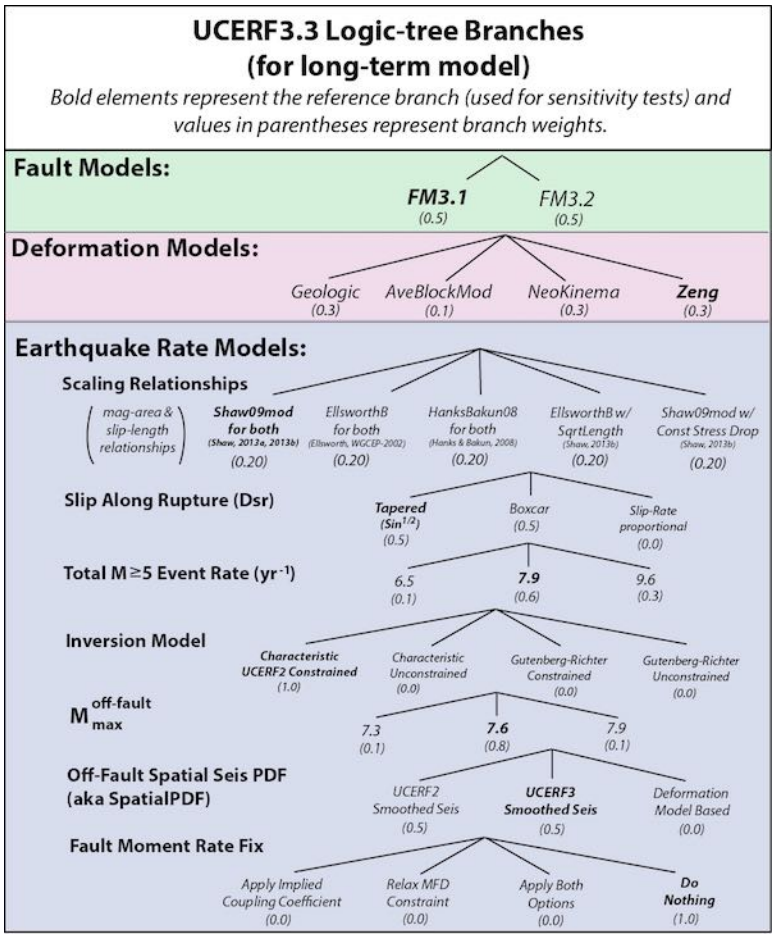

Fig. 2 – The UCERF3 logic tree showing branches and associated weights. Two branches, 'Inversion Model' and 'Fault Moment Rate Fix,' only have a single branch-leaf each with non-zero weight and are not explicitly discussed in the text. See Field et al. [1] for more details. PDF refers to probability density function, MFD to magnitude-frequency distribution, and UCERF2 to version 2 of the Uniform California Earthquake Rupture Forecast.

UCERF3 also introduces multiple new datasets and models. These include four new deformation models that provide fault slip rates: one based on pure geologic data, and three kinematically consistent models derived from geodetic data. The geodetic models provide slip rates on faults where no geologic slip rate data exist; the inclusion of these models yields some of the largest hazard changes relative to UCERF2. The deformation models also constitute one of the largest epistemic uncertainties, with the other important contributors being a new smoothed-seismicity algorithm, two new fault magnitude-area-slip scaling relationships, and alternative values for the total regional rate of **M**5 and larger events. All these options are new, so UCERF3 represents a considerable broadening of epistemic uncertainty, with the complete set yielding 1,440 alternative models (logic tree branches; Figure 2). Considering five new GMMs [3, 5], as well as three branches that expand the epistemic uncertainty in the GMMs, the NSHM logic tree for California consists of 21,600 independent branches.

2. The UCERF3 Logic Tree

In building UCERF3, the WGCEP updated almost all datasets from UCERF2 and developed new fault, deformation, and earthquake rate models, a review of which is provided below, along with details on how the WGCEP assigned branch weights. See Figure 2 for an outline for the foregoing discussion.

2.1 Fault Models

UCERF3 fault models (Figure 2) are based on the Southern California Earthquake Center's Statewide Community Fault Model [6] and define the geometry of large, active, earthquake generating faults. The two alternative fault models differ where there is some ambiguity as to fault structure at depth. Most model differences are focused in coastal southern California, with a few in the vicinity of San Francisco and the northernmost part of the state. Given good community consensus on fault geometries and attendant ambiguities, the WGCEP did not prefer one model to the other and assigned each an equal weight of 0.5. The largest change from UCERF2, affecting both models equally, was the addition of numerous faults with nominal rates of <0.2 mm/yr. These faults are typically Quaternary in age and have the potential to generate or participate in large earthquakes and should be considered in a complete network of active faults.

2.2 Deformation Models

Deformation models (Figure 2) provide fault slip rates and aseismicity factors (percentage of a fault surface that is slipping aseismically, or creeping) for calculating the seismic moment release of large ($\mathbf{M}\gtrsim 6.5$) magnitude events on faults. In UCERF3, and new to the NSHM, geodetic data derived from Global Positioning System (GPS) observations are used in addition to geologic data to estimate fault slip rates. In order to estimate slip rates from geodetic data, UCERF3 considers three kinematically consistent models: 'NeoKinema,' 'Zeng' and an average block model, hereafter referred to as 'AveBlockMod' [7] These models weight geologic and geodetic data in different ways to estimate fault slip rates and still honor observed data. Together they yield a variety of slip rates for each fault model. Among the geodetic models, NeoKinema best fits the observed GPS data. The Zeng model, on the other hand, is explicitly constrained to stay within geologic bounds and satisfy the GPS data as closely as possible. The AveBlockModel is based the average of five kinematically consistent models that assume a block-like geometry. A fourth 'Geologic' model [8] fits geologic observations exactly. Geologic data, however, are limited because only ~170 faults out of ~350 total in UCERF3 are directly constrained by data. To complete the model, any unconstrained fault slip rates had to be inferred.

The WGCEP assigned weights to the different models, favoring geologic (the pure geologic and Zeng models, both with 0.3 weight) over geodetic (NeoKinema with 0.3 weight, and the AveBlockMod with 0.1) models given their relative novelty. Initially, the latter two models were assigned equal weights of 0.2; however, the AveBlockMod contained some alarmingly high slip rates where the model concentrated slip on block boundaries and was down-weighted to 0.1 accordingly. In order to preserve the original balance between geologic and geodetic models, the weight of the NeoKinema model was increased to 0.3.

2.3 Earthquake Rate Models

Earthquake rate models (Figure 2) provide the recurrence rate of all earthquake ruptures under consideration. In UCERF3, rates are calculated for ruptures occurring on known faults, as constrained by the fault and deformation models described above, as well as on unknown faults. Earthquake ruptures on unknown faults are represented as a $0.1^{o}\times0.1^{o}$ geographic grid of sources derived from different models of smoothed seismicity. The calculation of time-independent rates for all earthquake ruptures in UCERF3 depends on the following components: (1) magnitude scaling relationships, (2) slip distribution along ruptures, (3) the total statewide rate of events, (4) the maximum magnitude of off-fault events, and (5) models of smoothed seismicity. UCERF3 integrates these five components as constraints, along with observed paleoseismic event data and the fault and deformation models, when inverting for statewide event rates. Note that two branches, 'Inversion Model' and 'Fault Moment Rate Fix' only have a single branch-leaf each with non-zero weight and are not explicitly discussed here; see Field et al. [1] for additional details.

2.3.1 Magnitude Scaling Relationship

The magnitude of each rupture in UCERF3 is computed from three magnitude-area relationships (Figure 2): 'HanksBakun08' [9], 'EllsworthB' [10], and 'Shaw09mod' [11]. Although the fault-surface area of rupture is generally based on the depth of microseismicity, it is now believed that slip likely extends below such depths for large ruptures associated with high magnitude events. To account for this, the UCERF3 magnitude scaling branch considers two additional slip-length scaling relationships: 'EllsworthB w/ SqrtLength' and 'Shaw09mod w/ ConstStressDrop' [11]. Because observed data do not suggest that any one model should be preferred over another, the WGCEP assigned an equal weight of 0.2 to each.

2.3.2 Slip Along Rupture

In UCERF3, two models capture the epistemic uncertainty in the slip distribution along the strike of faults: (1) 'Tapered' (or square-root-sine model) and (2) Boxcar (or uniform-slip model). Although a tapered-slip model might be preferred on observational grounds, because of how faults are discretized, this model introduces a large number of small magnitude events at the ends of faults. The WGCEP therefore assigned an equal weight of 0.5 to each.

2.3.3 Total Statewide Rate of Events

The total rate of **M**≥5 events per year within the UCERF3 model region, which includes a 100 km buffer around the California state boundary, is calculated from the observed earthquakes for the time span: 1850 to 2011. The data contain both historical and the modern instrumental catalogs. In order to calculate the rates, the state is divided into eight regions and the rates in each are calculated using a magnitude completeness specific to the region. The 1850–2011 rate averages for each of the eight regions were then summed to obtain the statewide average total rate. In addition, a global analogue model was used to calculate temporal rate variations. Considering errors associated with temporal rate variations and the observation that the long-term paleoseismic rate of large earthquakes in California is likely higher than the post-1850 catalog rate [12], the WGCEP assigned an asymmetric distribution of logic tree weights: 0.1 to a total rate of 6.5 **M**≥5 events per year, 0.6 to a total rate of 7.9, and 0.3 to a rate of 9.6.

2.3.4 Maximum off-fault magnitude (M_{max})

The range of maximum off-fault magnitude in UCERF3 is assumed to fall between **M**=7.3 and **M**=7.9, and the WGCEP assigned weights of 0.1 to $\mathbf{M}_{max}$=7.3, 0.8 to $\mathbf{M}_{max}$=7.6 and 0.1 to $\mathbf{M}_{max}$=7.9. The range of $\mathbf{M}_{max}$ for off-fault events is largely substantiated by the 2010 **M**7.2 El Mayor–Cucapah earthquake, which did not occur strictly on a mapped fault and which exceeded the prior UCERF2 value of $\mathbf{M}_{max}$=7 for background seismicity.

2.3.5 Off-Fault Seismicity

The distribution of off-fault seismicity is calculated using two different approaches: (1) fixed-kernel smoothing and (2) adaptive smoothing. The former approach was developed by the USGS [13] and used exclusively in all prior releases of the NSHM. In this approach, estimates of the space-rate-magnitude distribution of **M**≥4 events are computed using an isotropic 2D Gaussian smoothing of the locations of past events by assuming a correlation distance of 50 km. Adaptive smoothing [14, and references therein], on the other hand, estimates the rates of **M**≥2.5 events also using a 2D Gaussian smoothing kernel but with a smoothing distance based on the 8th closest earthquake [15] This latter approach was found to perform well when forecasting **M**≈4 to 5 events in the Regional Earthquake Likelihood Model tests [16] but tends to focus high levels of hazard over very small areas. Recognizing the uncertainty as to which approach is more appropriate for forecasting rates of large magnitude earthquakes (**M**>6), the WGCEP assigned equal weight to each branch.

3. The Ground Motion Model (GMM) Logic Tree

Ground-motion models provide an estimate of ground shaking intensity based on the characteristics of an earthquake source (e.g., magnitude), propagation path of the seismic waves (e.g., distance), the site conditions (e.g., shear wave velocity) and many other parameters. The intensity is quantified by the mean and standard deviation of spectral acceleration, peak ground acceleration or peak ground velocity. Such models are generally

developed using large collections of recorded ground motions, particularly for regions where records are abundant. Recently the Pacific Earthquake Engineering Research center (PEER) developed five so-called next generation attenuation relationships for shallow crustal earthquakes in active tectonic regions (known as NGA-W2; [3]). The five models were developed independently by five research groups using the same database of ground motion records, and were adopted by the USGS for use in the latest update of the NSHM [5]. The models predict different ground motions due to differences in the functional forms, regression techniques, data selection, and model parameterizations, thereby capturing epistemic uncertainty in the prediction of ground motion.

Prior to the release of the 2014 NSHM, the USGS developed criteria for the selection and weighting of GMMs. Based on the criteria, all five NGA-W2 models were assigned equal weight with the exception of the Idriss [17] model due to its lack of detailed modeling features. For instance, Idriss [17] does not distinguish between different fault mechanisms or consider a basin depth term. Because there are relatively fewer data for large magnitude events at short distances, the USGS applied additional epistemic uncertainty to the ground-motion models based on the assumption that the uncertainty is 50% for $M_W \geq 7$ earthquakes at a distance ≤ 10 km. The additional uncertainty was then scaled according to the number of earthquakes used in developing GMMs for different magnitude and distance bins relative to the reference bin, i.e., $M_W \geq 7$ at distance ≤ 10km. In the final application of the GMM logic tree when computing hazard, the mean of each model is assigned a weight of 0.63 and the branches are assigned with additional epistemic uncertainty of 0.185, corresponding to an approximation of a normal distribution.

4. Uncertainty Analysis

The combined earthquake source and ground motion model logic trees for the California portion of the USGS NSHM yield a total of 21,600 unique branches. Whereas the USGS typically only publishes mean hazard curves that are the weighted sum of the 21,600 component curves, in this analysis we preserve the curves from each logic tree branch and compute fractile curves corresponding to 1- and 2-sigma uncertainties in hazard. These analyses show that the uncertainties are typically quite large and consistently larger than the change one might observe at a site from one release of the U.S. hazard model to the next.

4.1 Hazard distribution and tornado diagrams

Although we have performed analyses at a large number of sites throughout California, we restrict our discussion to a Los Angeles site. Fig. 3 shows the distribution of hazard at a site in downtown Los Angeles. Although the 2% in 50-year peak ground acceleration (PGA) changed little from 2008 to 2014, the change is dwarfed by the uncertainty in the 2014 model.

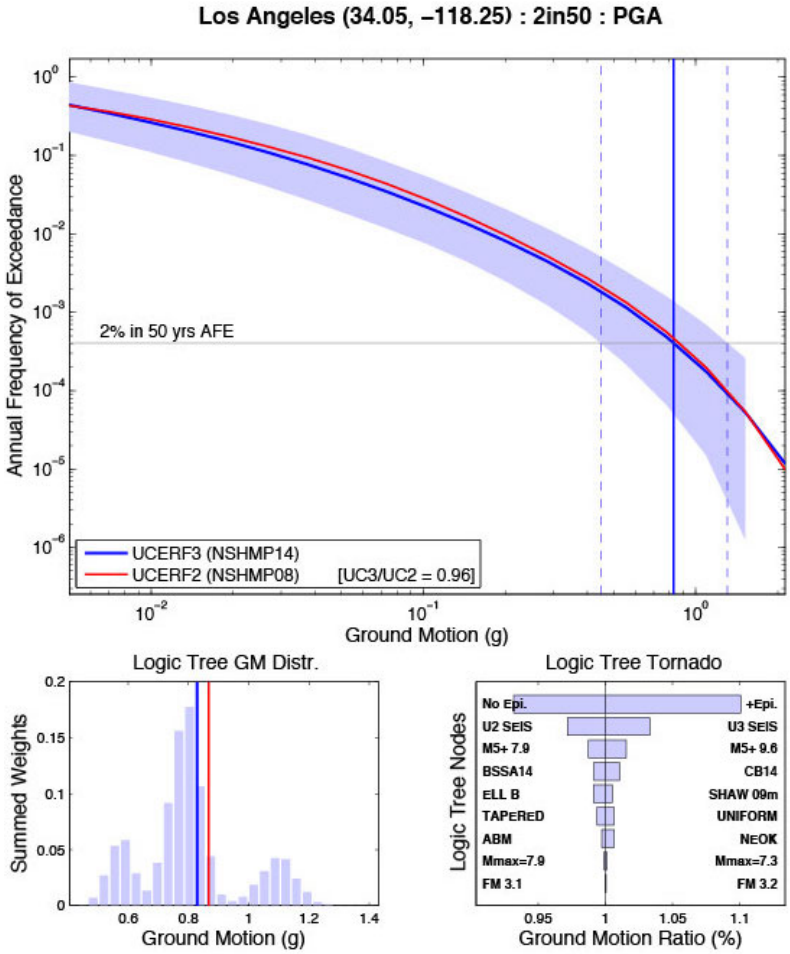

Fig. 3 – The distribution of hazard in downtown Los Angeles, California at the peak ground acceleration expected to be exceeded with a 2% probability in 50 years. The top plot shows the median hazard from UCERF3 (solid blue line) and UCERF2 (solid red line); the light blue shaded area marks the minimum and maximum hazard curves computed from the California NSHMP logic tree. The plot on the lower left shows the distribution of possible ground motions represented by the mean hazard, and the plot on the lower right shows the relative influence of different logic tree branches on total hazard.

4.2 Deaggregation of hazard uncertainty by logic tree branch

While the distribution of ground motions is useful, our analysis preserves all logic tree branch information so it is possible to decompose the distribution of possible ground motions by logic tree branch. Fig. 4 shows the same histogram as that in the lower left plot in Fig. 3, but now recolored by the branches in each logic tree group. For most branch groups, there is little to distinguish one branch from another. However, we know from our tornado diagram that the additional epistemic uncertainty on ground motion is the most influential branch group. Note, then, that the corresponding colored histogram in Fig. 4 clearly shows that the tri-modal form of the distribution is clearly driven by the additional epistemic uncertainty branches of the GMM logic tree. The choice of smoothed seismicity model is the second most influential branch group; note in the histogram corresponding to smoothed seismicity model in Fig. 4 that the UCERF3 smoothed seismicity model (U3 SEIS) is skewed slightly higher than that of UCERF2 (U2 SEIS).

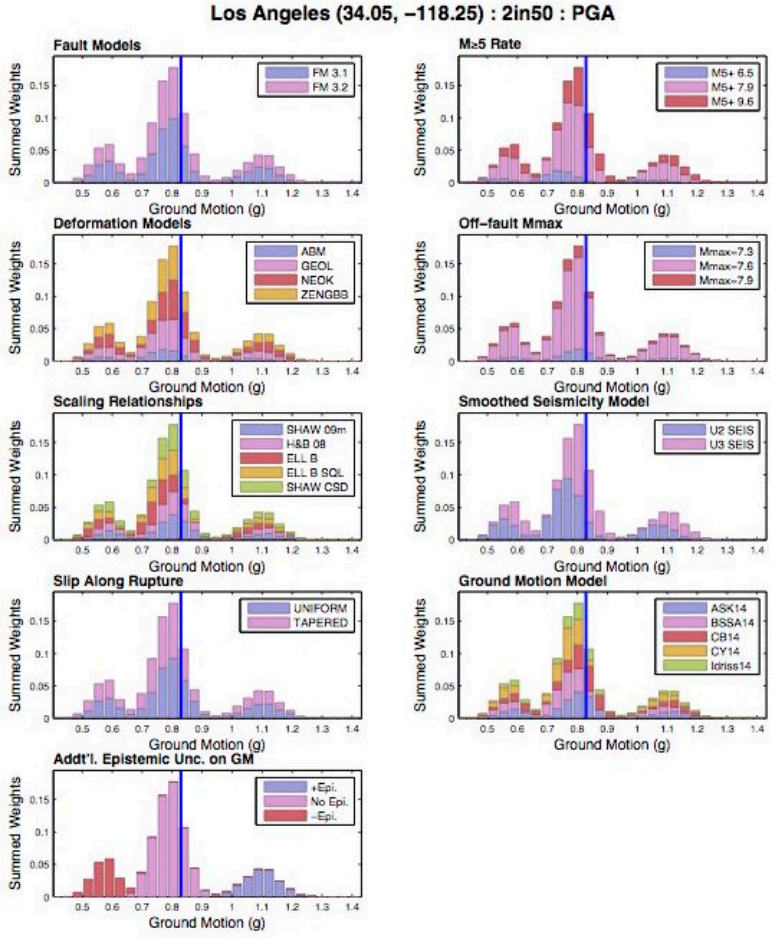

Fig. 4 – Deaggregation of ground motion uncertainty by logic tree branch. Note that all histograms are identical (and identical to that shown in Fig. 3). The relative contributions to the distribution by each group

4.3 Ranking logic tree branches.

As logic trees grow, computational demands can increase significantly and it is therefore useful to have means by which one can evaluate the relative importance of one or more branch groups over the others. Although a more spatially comprehensive approach will be required in the future, we aggregate data from the tornado diagrams computed at 40 sites in California and consider 4 periods (PGA, and 5-Hz, 1-Hz and 4-sec spectral accelerations. Fig. 5 shows the number of times a branch group appears at a particular rank in a tornado diagram.

This exercise shows that, within the limited scope of the analysis at present, the choice of off-fault $\mathbf{M}_{max}$ exerts little influence over hazard, whereas the choice of GMM and the application of additional epistemic uncertainty on ground motions are highly influential. Future work will involve a more spatially comprehensive analysis of logic tree branch choices and comparison of uncertainty estimates computed using the full logic tree with those computed from a reduced logic tree where some branches have been collapsed.

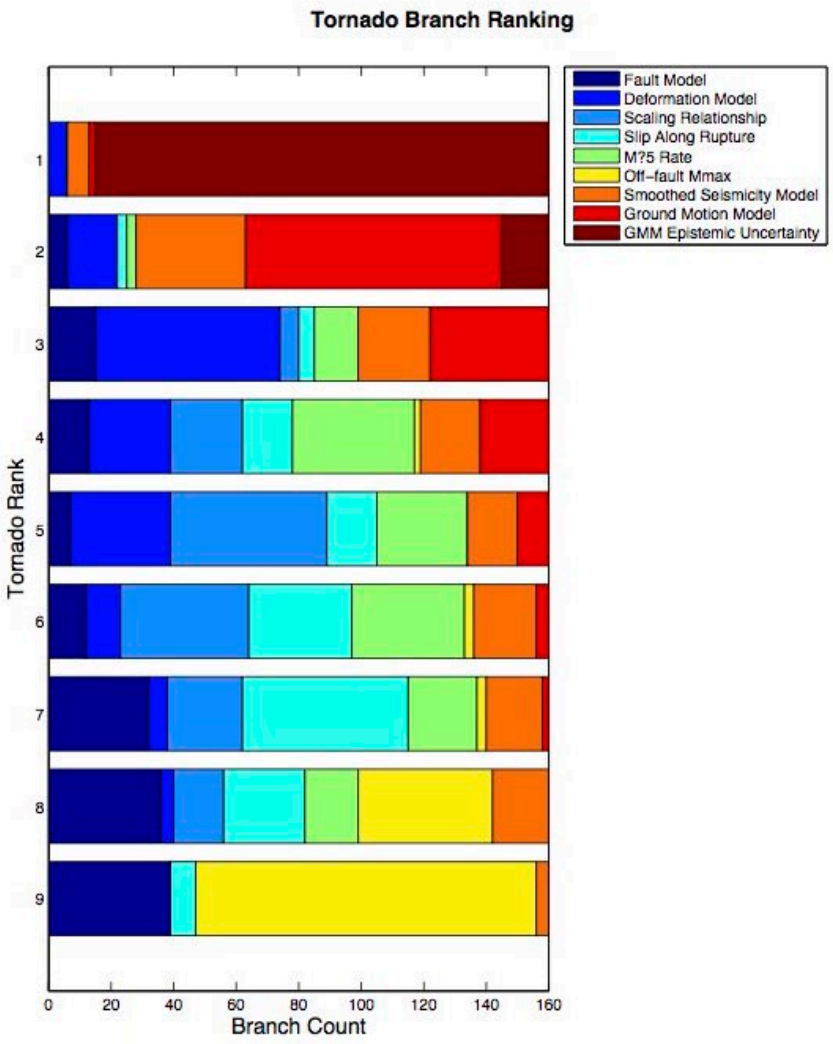

Fig. 5 – Ranking of logic tree branches.

5. Conclusions

Our analyses show that uncertainty in the California NSHMP is quite large and is not adequately reflected in the mean hazard curves that USGS typically produces. Moreover, the uncertainties in hazard are typically much larger than the changes observed by the introduction of new data and models from one release of the NSHM to the next. The California NSHMP logic tree also has many branches, but preliminary studies show that we can probably collapse portions of the logic tree while preserving estimates of uncertainty that closely match those derived from the full logic tree.

It is important to note that our analyses do not consider contributions to hazard from subduction interface and intraslab events associated with the Cascadia subduction zone. Our analysis also assumes that each logic tree branch is mutually exclusive, which it may not be. For example, the selection of a particular deformation or slip model implies that all other earthquake sources are slipping at a rate derived from the chosen model. We are currently developing the capability to perform Monte Carlo sampling of the California NSHMP logic tree to account for inter-branch correlations.

6. References

[1] Field EH, Arrowsmith RJ, Biasi GP, Bird P, Dawson TE, Felzer KR, Jackson DD, Johnson KM, Jordan TH, Madden C, Michael AJ, Milner KR, Page MT, Parsons T, Powers PM, Shaw BE, Thatcher WR, Weldon RJ II, and Zeng Y (2014): Uniform California Earthquake Rupture Forecast, version 3 (UCERF3) — The time-independent model. *Bulletin of the Seismological Society of America*, **104**, 1122–1180.

[2] Working Group on California Earthquake Probabilities (WGCEP) (2015): Official website for the working group, available at http://www.wgcep.org (last accessed 31 May 2016).

[3] Bozorgnia, Y, Abrahamson, NA, Atik LA, Ancheta TD, Atkinson GM, Baker JW, Baltay A, Boore DM, Campbell KW, Chiou BS-J, Darragh R, Day S, Donahue J, Graves RW, Gregor N, Hanks T, Idriss IM, Kamai R, Kishida T, Kottke A, Mahin SA, Rezaeian S, Rowshandel B, Seyhan E, Shahi S, Shantz T, Silva W, Spudich P, Stewart JP, Watson-Lamprey J, Wooddell K, and Youngs R (2014): NGA-West2 research project. *Earthquake Spectra*, **30**, 973–987.

[4] Field EH, Dawson TE, Felzer KR, Frankel AD, Gupta V, Jordan TH, Parsons T, Petersen MD, Stein RS, Weldon RJ II, and Wills CJ (2009): Uniform California Earthquake Rupture Forecast, version 2 (UCERF 2). *Bulletin of the Seismological Society of America*, **99**, 2053–2107.

[5] Petersen MD, Moschetti MP, Powers PM, Mueller CS, Haller KM, Frankel AD, Zeng Y, Rezaeian S, Harmsen SC, Boyd OS, Field N, Chen R, Rukstales KS, Luco N, Wheeler RL, Williams RA, and Olsen AH (2014): Documentation for the 2014 update of the United States National Seismic Hazard Maps. U.S. Geological Survey, *Open-File Report* 2014-1091, 243 pp.

[6] Plesch A, Shaw JH, Benson C, Bryant WA, Carena S, Cooke M, Dolan J, Fuis G, Gath E, Grant L, Hauksson E, Jordan T, Kamerling M, Legg M, Lindvall S, Magistrale H, Nicholson C, Niemi N, Oskin M, Perry S, Planansky G, Rockwell T, Shearer P, Sorlien C, Süss MP, Suppe J, Treiman J, and Yeats R (2007): Community fault model (CFM) for southern California. *Bulletin of the Seismological Society of America*, **97**, 1793–1802.

[7] Parsons T, Johnson KM, Bird P, Bormann JM, Dawson TE, Field EH, Hammond WC, Herring TA, McCaffrey R, Shen Z-K, Thatcher WR, Weldon RJ II, and Zeng Y (2013): Appendix C: Deformation models for UCERF3.3. U.S. Geological Survey, *Open-File Report* 2013-1165-C, 66 pp.

[8] Dawson TE, and Weldon RJ II (2013): Appendix B: Geologic slip-rate data and geologic deformation model. U.S. Geological Survey, *Open-File Report* 2013-1165-B, 29 pp.

[9] Hanks TC, and Bakun WH (2008): M- log A observations of recent large earthquakes. *Bulletin of the Seismological Society of America*, **98**, 490–494.

[10] Working Group on California Earthquake Probabilities (WGCEP) (2003): Earthquake probabilities in the San Francisco Bay region: 2002–2031. U.S. Geological Survey, *Open-File Report* 2003-214, 235 pp.

[11] Shaw BE (2013): Appendix E: Evaluation of magnitude-scaling relationships and depth of rupture: Recommendation for UCERF3. U.S. Geological Survey, *Open-File Report* 2013-1165-E, 23 pp.

[12] Felzer KR (2013): Appendix L: Estimate of the seismicity rate and magnitude–frequency distribution in California from 1850 to 2011. U.S. Geological Survey, *Open-File Report* 2013-1165-L, 13 pp.

[13] Frankel A (1995): Mapping seismic hazard in the central and eastern United States. *Seismological Research Letters*, **66**, 8–21.

[14] Werner MJ, Helmstetter A, Jackson DD, and Kagan YY (2011): High-resolution long- term and short-term earthquake forecasts for California. *Bulletin of the Seismological Society of America*, **101**, 1630–1648.

[15] Felzer KR (2013): Appendix M: Adaptive smoothed seismicity model. U.S. Geological Survey, *Open-File Report* 2013-1165-M, 12 pp.

[16] Zechar JD, Schorlemmer D, Werner MJ, Gerstenberger MC, Rhoades DA, and Jordan TH (2013): Regional earthquake likelihood models I: First-order results. *Bulletin of the Seismological Society of America*, **103**, 787–798.

[17] Idriss IM (2014): An NGA-West2 empirical model for estimating the horizontal spectral values generated by shallow crustal earthquakes. *Earthquake Spectra*, 30, 1155-1177.