Manuscript

Accounting for Epistemic Uncertainty in PSHA: Logic Tree and Ensemble Modeling

Warner Marzocchi[a], Matteo Taroni[a], Jacopo Selva[b]

[a] Istituto Nazionale di Geofisica e Vulcanologia, Via di Vigna Murata 605, 00143 Rome, Italy

[b] Istituto Nazionale di Geofisica e Vulcanologia, Via Donato Creti 12, 40128 Bologna, Italy

Submitted to: *Bulletin of Seismological Society of America*

Abstract

Any trustworthy probabilistic seismic hazard analysis (PSHA) has to account for the intrinsic variability of the system (aleatory variability) and the limited knowledge of the system itself (epistemic uncertainty). The most popular framework for this purpose is the logic tree. Notwithstanding its vast popularity, the logic tree outcomes are still interpreted in two different and irreconcilable ways. In one case, practitioners claim that the mean hazard of the logic tree is *the* hazard and the distribution of all outcomes does not have any probabilistic meaning. On the other hand, other practitioners describe the seismic hazard using the distribution of all logic tree outcomes. In this paper, we explore in detail the reasons of this controversy about the interpretation of logic tree, showing that the distribution of all outcomes is more appropriate to provide a joined full description of aleatory variability and epistemic uncertainty. Then, we provide a more general framework – that we name *ensemble* modeling – in which the logic tree outcomes can be embedded. In this framework, the logic tree is not a classical probability tree, but it is just a technical tool that samples epistemic uncertainty. Ensemble modeling consists of inferring the parent distribution of the epistemic uncertainty from which this sample is drawn. Ensemble modeling offers some remarkable additional features. First, it allows a rigorous and meaningful validation of any PSHA; this is essential if we want to keep PSHA into a scientific domain. Second, it provides a proper and clear description of the aleatory variability and epistemic uncertainty that can help stakeholders to appreciate the whole range of uncertainties in PSHA. Third, it may help to reduce the computational time when the logic tree becomes computationally intractable because of the too many branches.

Introduction

Seismic hazard practitioners distinguish uncertainties of different nature in a convenient way, adopting the term aleatory variability to describe the intrinsic irreducible variability of the process generating ground shaking intensity, and epistemic uncertainty to characterize all reducible uncertainties due to our limited knowledge about the true model describing the aleatory variability. Notwithstanding the popularity of this distinction (e.g. SSHAC, 1997), many authors and philosophers take the view that this separation is ambiguous, and it does not have a theoretical significance, because, as far as our knowledge of the system may increase, all uncertainties become necessarily epistemic (e.g., NRC, 1997; Bedford and Cooke, 1991; Lindley, 2000; Jaynes, 2003).

The discussion about the distinction between aleatory variability and epistemic uncertainty is far to be purely academic. Indeed, this discussion is deeply rooted on the intrinsic meaning of probability (frequency versus degree of belief) and, more important, on the possibility to validate a probabilistic assessment like the outcome of probabilistic seismic hazard analysis (PSHA). Recently, Marzocchi and Jordan (2014) suggest that a clear and univocal taxonomy of uncertainties is not only of practical convenience, but it is of primary importance to validate meaningfully any probabilistic assessment, and, consequently, to keep PSHA into a scientific domain (see Marzocchi and Jordan, 2014 for a discussion on commonalities and differences with the traditional view of PSHA practitioners; e.g., SSHAC, 1997). In particular, Marzocchi and Jordan (2014) show that aleatory variability and epistemic uncertainty can be separated only in the framework of a well-defined *experimental concept*. The experimental concept defines collections of data, observed and not yet observed, that are judged to be exchangeable when conditioned on a set of explanatory variables (Draper et al., 1993). When we define the set of data that we aim to describe and that will be used to test the model, we are implicitly defining an experimental concept. In this framework, the aleatory variability is not associated to the true physical process, but it is described by the event frequency of the exchangeable dataset, and the epistemic uncertainty is represented by the lack of knowledge of what the true frequency is. In a time-independent PSHA context (Marzocchi and Jordan, 2014), an experimental concept can be defined by a sequence of ground shaking exceedances in one specific site that are assumed to be exchangeable in time. In this case, the aleatory variability is the long-term frequency of exceedances for that specific site, i.e., the true hazard, and the epistemic uncertainty is the lack of knowledge of what the true hazard is. (Note that in this paper we quantify the seismic hazard in terms of exceedance probability). Worthy of note, despite providing an unambiguous distinction between aleatory variability and epistemic uncertainty, this taxonomy is consistent to the SSHAC's view (1997) of uncertainties (as a footnote of section 2.2.3 of the main SSHAC report the authors write: "*The distinction between aleatory and epistemic uncertainty may at first appear inconsistent with the Bayesian view of probability, but, in fact, it is entirely consistent with this view. Aleatory uncertainties may be thought of as frequencies of a set of exchangeable events or as frequency distributions of an exchangeable set of continuous random variables. If the frequencies or frequency distributions are uncertain, it makes perfect sense to assess probability distributions over the unknown frequencies or parameters of the unknown frequency distributions.*").

According to this view, any trustworthy PSHA must provide a reliable estimate of the aleatory variability incorporating in a proper way the epistemic uncertainty. While each single PSHA model aims to describe the aleatory variability, the inclusion of epistemic uncertainty is usually tackled by analyzing the results of alternative and scientifically acceptable PSHA models (SSHAC 1997). This is usually made using the logic tree structure as originally suggested by Kulkarni et al. (1984). In essence, the logic tree dissects the PSHA problem into basic components embedded in a hierarchical framework. The nodes represent a logical progression of potential sources of epistemic uncertainty and the branches depict the possible alternative describing the uncertainty at each node. The final branches are meant to represent the complete epistemic uncertainty in PSHA and they are combined using the probabilistic structure of classical probability tree. Despite the use of the logic tree scheme has become *de* *rigueur*, it is well known that there are conceptual pitfalls that should be taken into account (Bommer and Scherbaum, 2008). The most important controversy regards the interpretation of the logic tree output. As a matter of fact, in PSHA practice there are two different and irreconcilable attitudes: some scientists describe the hazard using the percentiles of the logic tree outcomes (Abrahamson and Bommer, 2005; Stucchi et al., 2011; Field et al., 2014), while others firmly claim that the use of percentiles throws away probabilism and the mean hazard is *the* hazard (e.g., McGuire et al., 2005; Musson, 2005; 2012).

Understanding the reasons and consequences of these apparently irreconcilable views is essential for a proper description of the epistemic uncertainty in PSHA. In this paper we explore in detail this issue and we provide a framework to interpret the variability of logic tree outcomes. This general framework – that we name *ensemble modeling* – offers also further opportunities. It does not require necessarily a logic tree, but it may apply also to independent hazard models. It provides a formal framework to validate and test meaningfully PSHA models and to fully characterize aleatory variability and epistemic uncertainty. It may help to reduce significantly the computational time when moving from hazard to risk.

Two views of the logic tree outcomes in PSHA

The logic tree (Kulkarni et al., 1984) incorporates the epistemic uncertainty borrowing the same probabilistic structure of classical probability trees. Probability trees, like event trees and fault trees (e.g. Kumamoto and Henkley, 1996), are very useful tool to facilitate the treatment of probabilistic problems that may be described through a hierarchical structure with a discrete number of possibilities. One of the most remarkable common features of all flavors of probability tree is that they are structured to fully represent all possible outcomes. In other words, all branches emerging from a node of the tree must represent a mutually exclusive and collectively exhaustive (MECE) set of events, and, as a consequence, one path of the tree must represent the true outcome. The MECE postulation implies that the probabilities of the logic tree can be combined using the law of total probability that reads

$$\Pr(E) = \sum_{i=1}^{N} \Pr(E \cap H_i) = \sum_{i=1}^{N} \Pr(E|H_i)\Pr(H_i) \qquad (1)$$

where Pr(E) is the probability of the event of interest, Pr(E | H_i) is the conditional probability of the event E given the terminal branch of the probability tree H_i , and Pr(H_i) is the probability that the terminal branch H_i (i=1,...,N) is the true one; the latter is also known as the weight of the i-th terminal branch and it has a clear and univocal probabilistic interpretation (Scherbaum and Kuhen, 2011). Some relaxations of the MECE assumption have been made, mostly motivated by practical aspects (see, e.g., Newhall and Hoblitt, 2002; Marzocchi et al., 2010), but these changes involve much more cumbersome calculations, and, in any case, these generalizations of the probability trees still consider all possible outcomes.

We argue that applying this probabilistic structure – which is based on equation 1 and the MECE postulation – to the logic tree in PSHA raises several problems. The most important one stems from the fact that the structure of the probability tree has been designed to describe the aleatory variability, not the epistemic uncertainty. This important feature of the probability tree can be grasped through a simple example that does not pretend to be exhaustive of the functioning of any possible probability tree, but it underlies its basic features. In Figure 1 we plot a probability tree to calculate the probability to get *head*, Pr(E), from coin tosses. In particular, there are two boys (Tim and Tom) having two and three coins each. Tim's coins are biased having Pr(E | H_i) (i.e., the probability of getting head by the i-th coin) equal to 0.4 and 0.3. Tom's three coins are biased as well, with Pr(E | H_i) equal to 0.7, 0.7, and 0.8. If we do not know who will toss the next coin (Tim and Tom have the same probability to be selected) and the coin that will be used (each coin has the same probability to be thrown), the tree has five terminal branches with different weights, i.e., each one of Tim's branches has Pr(H_i) =0.25, while Tom's branches have weight Pr(H_i)=0.16. The probability of getting head when we don't know who is going to toss the coin and the coin that will be tossed is given by equation 1, i.e., Pr(*E*) =0.54. This value has a frequentist interpretation because it describes the aleatory variability of the experimental concept; if we run a simulation in which, for each run, we select randomly the boy who will toss the coin and the coin to be tossed, the expected long-term frequency of head is 0.54. In this example, the branches distribution describes exhaustively all possible cases, mimicking the lack of knowledge of which path (which boy and which coin) will be followed in each run; this uncertainty is taken into account by the averaging of equation 1.

If this probabilistic scheme is directly applied to PSHA, it follows that i) the mean hazard is the true hazard (McGuire et al., 2005); ii) Pr(H_i) represents the probability of the model H_i to be the true hazard model (since no practitioner believes that one of the paths of the logic tree represents the true hazard, the MECE assumption is pragmatically resumed replacing the term true with the *one that should be used*; Scherbaum and Kuhen, 2011); iii) the use of percentiles does not make sense in this framework (Musson, 2012). However, the logic tree applications in PSHA are meant to do something different. In fact, the branches of the logic tree represent different alternatives, not different possibilities as in the probability tree of Figure 1. Within the logic tree in PSHA we expect that *the branch that should be used* is always the same. Applying the logic tree concept to the example of Figure 1, we would have the same (unknown) coin tossed by the same (unknown) boy. In this case, the mean value (0.54) no longer has a frequentist interpretation, and it does not represent the true Pr(*E*) (the aleatory variability) that is given by the outcome of one (unknown) of the final branches. Coming back to a PSHA context, this implies that the mean hazard is not the true hazard, because the mean will almost never coincide with *the branch that should be used*.

This problem is not properly acknowledged in scientific literature, but probably it unconsciously motivates the peculiar use of the logic tree made by some practitioners. Instead of using only the mean, they give more emphasis to the full discrete distribution of the final branches outcome using percentiles (e.g., Abrahamson and Bommer, 2005; Stucchi et al., 2011; Field et al., 2014). Although this approach has an intuitive appealing because, using the

SSHAC words (SSHAC, 1997), it represents "*the center, the body, and the range of technical interpretations that the larger technical community would have if they were to conduct the study*", conversely it violates the probabilistic framework of the probability trees described by equation 1 and the MECE postulation, causing part of the controversies at the base of the use and misuse of the logic tree (e.g. Bommer and Scherbaum, 2008), and in particular the controversy related to the interpretation of the logic tree outcomes.

If the goal of the quantification of epistemic uncertainty is to establish where the true hazard (the true aleatory variability) should be (SSHAC 1997; Marzocchi and Jordan, 2014), we may use the variability of the outcomes generated by a set of reasonable models to bound where the true hazard is expected to be. This view is coherent with the use of percentiles in the context of a logic tree. Conversely, it is not coherent with the probabilistic structure of a classical probability tree, which must honor equation 1 and MECE postulation. So, when using the full distribution of the logic tree outcomes, the logic tree is not anymore a probability tree, but it is only a technical tool that facilitates the production of a range of models sampling the epistemic uncertainty. Moreover, the weight of each model no longer has a specific probabilistic meaning, and there is no need to keep limited the number of branches as advocated by Scherbaum and Bommer (2008).

To summarize, if we aim to estimate the true hazard, we should abandon the approach that considers the mean hazard as the true hazard (McGuire et al., 2005; Musson, 2005, 2012), that is, we should abandon the probabilistic interpretation of logic trees based on MECE assumption and equation 1. Of course this does not mean that the mean hazard of a logic tree should not be used, but we have to be aware that, alone, it does not represent a long-term frequency of exceedances (i.e., the aleatory variability). This aspect is of paramount importance when testing hazard models; indeed, Marzocchi and Jordan (2014) show that the practice of using only the mean to test hazard models (e.g. McGuire and Barnhard, 1981; Stirling and Petersen, 2006; Albarello and D'Amico, 2008; Stirling and Gerstenberger, 2010) may lead to reject reliable models. On the other hand, the collection of all logic tree outcomes is intuitively more informative as any single value (SSHAC, 1997), and in the next section we show that these outcomes may be embedded into a quantitative framework – named *ensemble modeling* – which provides a coherent description of the aleatory variability and epistemic uncertainty of the seismic hazard. This description is essential to carry out any robust statistical testing of PSHA, and to deliver a more complete description of the seismic hazard to any interested stakeholder.

Ensemble modeling

We have just showed that the use of percentiles of logic tree aims to sample the epistemic uncertainty, not to fully describe it. This difference is not only semantic, but it has important consequences. Having a sample of the epistemic uncertainty implicitly means that the epistemic uncertainty can be fully characterized by a parent distribution from which the sample has been drawn. In a PSHA context – where this sample consists of a set of exceedance probabilities – Marzocchi and Jordan (2014) call this parent distribution *extended experts' distribution*. Ensemble modeling consists essentially of inferring this extended experts' distribution from the sample provided by the logic tree, or by any set of models that sample the epistemic uncertainty. The terms ensemble modeling and ensemble forecasts are used in many disciplines in different ways since early seventies (e.g. Leith, 1974). The recent Nate Silver's book (Silver, 2012) gives a wide range of successful applications and uses of ensemble modeling. The common feature across all these different flavors of ensemble modeling/forecasts is the attempt to account for epistemic uncertainty merging models/forecasts in a proper way.

In PSHA the logic tree outcome can be described by a vector y_i, ω_i , where y_i is the hazard curve of the *i*-th branch in a set of *N* branches, and ω_i is its weight. The epistemic uncertainty is visually portrayed by a family of hazard curves for each site. The bundle of curves can be dissected horizontally or vertically. In the first case, we get the distribution of the ground motion parameter for a specific exceedance probability. The second case is of particular interest for two main reasons. First, it is easier to conceive an experimental concept for testing; for instance, collecting exceedance events of a reference ground shaking intensity in a set of exchangeable time intervals for one specific site (see, e.g., Marzocchi and Jordan, 2014). Second, we get a distribution of exceedance probability for one specific value of the ground shaking intensity. The use of a probability distribution of probability has been matter of discussion and controversies in statistical literature and among practitioners (e.g. Bedford and Cooke, 1991; Lindley, 2000; Vick, 2002; Jaynes, 2003; Cox et al., 2008). These controversies have been addressed by Marzocchi and Jordan (2014) who provide a formal and consistent probabilistic framework in which probability is described through a distribution. The central value of this distribution is the best guess of the frequency of an exchangeable dataset (i.e., the aleatory variability), and the dispersion around the central value mimics the epistemic uncertainty (see also SSHAC, 1997; Marzocchi et al., 2008). This distribution has an intuitive interpretation, because it bounds where the true aleatory variability is expected to be. In case a single value is required to characterize the distribution, we emphasize that the mean value has the same legitimacy as any other single statistics, like the mode and the median to represent the distribution.

When dissecting the bundle of hazard curves vertically, y_i is replaced by $\theta_i^{(z)}$ that represents the exceedance probability of the *i*-th model/branch for the *z*-th ground shaking threshold. Here ensemble modeling considers $\theta_i^{(z)}, \omega_i$ as a sample of an unknown parent distribution $f\ \theta^{(z)}$ that describes the random variable $\theta^{(z)}$ taking into account the aleatory variability and epistemic uncertainty. The sample $\theta_i^{(z)}, \omega_i$ can either stem from one or more logic trees, or from a collection of models (hereafter with the term 'model' we mean either an independent model or a final branch of a logic tree); the only requirement is that $\theta_i^{(z)}, \omega_i$ represents an unbiased sample of the epistemic uncertainty. Models' output may be correlated and the weight attached to each model should properly take into account not only the confidence on each model (based on expert opinion and/or on quantitative evaluation of the forecasting performances), but also the possible strong correlation with other models (Marzocchi et al., 2012; Rhoades et al, 2014).

Notwithstanding inferring a distribution from a sample of data always introduces a potential source of ontologic error, we argue that this step is not more subjective than considering the percentiles as describing the real distribution of epistemic uncertainty. The inference of a parent distribution from a sample is one of the cornerstones of statistics, because it allows more meaningful tests and comparisons of models/hypotheses. In the following examples we show some of the additional features provided by the ensemble modeling framework.

Some realistic examples

In this section we show how ensemble modeling applies to two different realistic cases with few (a synthetic case for Italy) and many (UCERF3; Field et al., 2014) logic tree outcomes that describe well a wide range of possible scenarios for PSHA calculations. We underline again that the same example could have been made using independent hazard models without using any logic tree.

In the first example, we consider the seismic hazard for two cities in Italy, Cosenza and Bologna; Cosenza is located in a region with the highest seismic hazard in Italy, while Bologna is located in a medium seismic hazard area. The seismic hazard is obtained by a simple logic tree (Figure 2) composed by 5 different seismicity rate models, and three GMPEs. We arbitrarily select from the Italian CSEP experiment (Schorlemmer et al., 2010) five seismicity rate models: Hazgridx (Akinci, 2010), PHMzone (Faenza and Marzocchi, 2010), ALM (Gulia et al., 2010), MPS04 (MPS Working Group, 2004) and TripleS (Zechar and Jordan, 2010). The GMPEs are the ones proposed by Cauzzi and Faccioli (2008), Akkar and Bommer (2010), and Bindi et al. (2011). The weight of each model is assigned arbitrarily (Figure 2). This example does not aim at providing the true hazard in these sites, but it has been set up in order to show the functioning of the ensemble modeling in a realistic situation made by few branches of a logic tree.

In Figures 3 and 4 we show the seismic hazard for Cosenza and Bologna, respectively. In the upper panels, we show the mean and percentiles of the hazard curve for the peak ground acceleration (PGA). Although ensemble modeling does not impose any specific parametric distribution for $f\left(\theta^{(z)}\right)$, the Beta distribution is commonly used to describe a unimodal random variable bounded between 0 and 1 (Gelman et al., 2003). In this case, we assume that $\theta^{(z)} \sim \text{Beta}(\alpha, \beta)$, where the parameters α and β are related to the average and variance of $\theta^{(z)}$ that are provided by the set of hazard models/branches. In particular,

$$E\left(\theta^{(z)}\right) = \frac{\alpha}{\alpha+\beta} \quad (2)$$

and

$$\text{var}\left(\theta^{(z)}\right) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \quad (3)$$

where $E\left(\theta^{(z)}\right)$ and $\text{var}\left(\theta^{(z)}\right)$ are the weighted average and variance of the exceedance probabilities of the *z*-th ground shaking threshold. Inverting equations 2 and 3 we can get the parameters of the Beta distribution. Calculating the Beta parameters of the exceedance probability for a set of ground shaking thresholds, we can plot the uncertainty over the full hazard curve.

In particular, the percentiles of Figures 3 and 4 are obtained plotting the percentiles of the Beta distribution applied to the exceedance probability associated to a set of ground shaking thresholds. The area bounded by the 10-th and 90-th percentiles shows where the true hazard curve is expected to be with 80% of probability. In the lower panels we show the distribution of the exceedance probability for one specific ground shaking intensity (marked by a vertical line in the upper panel). The Beta distribution fits well the outcomes of the logic tree in both cases (we verify this hypothesis using the Kolmogorov-Smirnov one-sample test modified by

Lilliefors (1967) applied to the cumulative distributions in the lower right corner of Figures 3 and 4).

In this application the use of ensemble modeling offers to PSHA practitioners some additional features. The most important is that replacing few probability outcomes with a continuous distribution describing the aleatory variability and epistemic uncertainty is crucial for a meaningful test of any PSHA model (Marzocchi and Jordan, 2014). For example, having 15 branches the confidence interval defined by the lowest and highest percentiles is about 87%, implying that the true value has a probability of 0.13 to be outside from this interval. Having a continuous distribution allows practitioners to define more appropriate confidence intervals for testing and validation. Moreover, describing the epistemic uncertainty with a continuous distribution allows more meaningful comparisons and quantitative tests between hazards in different sites, and facilitates the identification of the sites where the true hazard may be more distant from the mean hazard, i.e., where we expect the largest variations of the mean hazard in future hazard evaluations (e.g. Paté-Cornell, 1996). For example, two sites with the same mean hazard may have a quite different dispersion of the exceedance probability distribution; this means that, although the mean hazard is the same, the site with the largest dispersion may have the true hazard much lower (or much higher) than the other site, and future analysis may provide mean significantly different for that site.

When the logic tree is composed by many branches like in UCERF3 (Field et al., 2014), the use of a continuous distribution may become superfluous, because the difference between adjacent percentiles becomes more and more negligible. Anyway, also in this case the ensemble modeling view offers some additional features. In Figures 5 and 6 we show the 7200 exceedance probabilities relative to the average PGA for 2% in 50 years for two different sites, Los Angeles and Redding. All these values come from one logic tree developed in the framework of UCERF3 (Field et al., 2014). The Beta distribution (equations 2 and 3) fits very well for Los Angeles, while for Redding the Beta distribution does not fit well the data because the outcomes of the logic tree are markedly bimodal. Adopting an ensemble modeling strategy, here practitioners have two options: if they think that their models are a representative sample of the epistemic uncertainty (i.e., they are assuming that additional model are not expected to fill that gap), they may use a different parametric distribution or a nonparametric fitting. For example, in Figure 6 we use the MATLAB function ksdenstity(x) (Bowman and Azzalini, 1997) that computes a probability density estimate from the set of weighted exceedance probabilities (we use 50 equally spaced points that cover the range of the exceedance probability). This option is quite similar to the direct use of percentiles to estimate $f\left(\theta^{(z)}\right)$. Otherwise, if they think that the bimodality is only due to the fact that the models used are just exploring only two extreme scenarios, they may still use a Beta distribution that fills the gap between the two modes. Of course, the choice of the most proper option introduces further subjectivity in PSHA, but we argue that this choice is certainly less subjective than describing the hazard using the mean alone, or using the percentiles of the distribution that implicitly means to impose a nonparametric distribution.

In this case, ensemble modeling framework offers also a further practical advantage. A seismic hazard logic tree with many branches can be hardly used for risk calculations if we still want to honor the logic tree structure, because it may require a prohibitive computational time (Field et al., 2005; Selva et al., 2013). In an ensemble modeling perspective, there is no need anymore to preserve the logic tree structure (intimately related to the MECE assumption) for further analysis. In practice, we may randomly sample (taking into account the relative weight of each model/branch) a convenient number, L, of hazard curves from the outcome of the seismic hazard logic tree and to combine each one of them with a correspondent randomly sampled vulnerability function.

The L combinations will yield a set of L risk curves that can be eventually used to build a parent distribution using the same ensemble modeling strategy. For example, while the use of a logic tree structure imposes the number of combinations, say L^*, between the hazard and vulnerability branches, the ensemble modeling approach allows practitioners to select a manageable number of combinations L, reducing the computational time of about a factor L^*/L.

Discussion and conclusions

In this paper we have explored the rationales behind some apparently irreconcilable interpretations of the logic tree outcomes in PSHA. In particular we have showed that a proper interpretation of the logic tree outcomes requires considering all final branches. In this case, the logic tree does not have to conform to the probabilistic scheme of classical probability trees, but it is just a technical tool that facilitates the construction of multiple models that sample the epistemic uncertainty.

We have also showed that the interpretation of the logic tree outcomes can benefit if we embed these outcomes into a more general probabilistic framework that we name *ensemble modeling*. Ensemble modeling allows scientists to define a parent distribution (called extended experts' distribution by Marzocchi and Jordan, 2014, when the sample is composed by exceedance probabilities) from a discrete set of values that can be obtained either from the branches of a logic tree or from the collection of different hazard models. The central value of this extended experts' distribution represents the best guess of the aleatory variability (the true hazard), and the dispersion around the central value mimics the epistemic uncertainty that bounds where the true hazard is expected to be (see also SSHAC, 1997). Ensemble modeling assumes that models are independent or that the weights associated to each model account for possible correlation between models (e.g. Bommer and Scherbaum, 2008; Marzocchi et al., 2012). If possible dependences among models are not properly accounted for, the parent distribution turns to be biased and it can be rejected through a formal test, or, using the Marzocchi and Jordan (2014) terminology, it exposes the model to an ontologic error. Noteworthy, this more general approach makes no longer the use of logic tree *de rigueur*, because the epistemic uncertainty can be either sampled by a logic tree, or by a set of different models.

Finally, we have showed that the use of the ensemble modeling view has some remarkable additional features. First, it may serve to design a rigorous testing phase of PSHA models, and to properly compare seismic hazards in different sites. Second, it provides a proper description and distinction of the aleatory variability and epistemic uncertainty; this can be helpful to show to the stakeholders the sites with the highest epistemic uncertainty, i.e., the sites where future large variations of the mean hazard are more likely. Third, it may drastically reduce the computational time, because we can combine different levels of information without preserving necessarily the logic tree structure.

Data and resources

The earthquake rate models used for Figures 3 and 4 are taken from the CSEP Italian experiment and they are described in the quoted references. The hazard data for Los Angeles and Redding (Figures 5 and 6) have been provided by Peter M. powers on March 20, 2014.

Acknowledgments

This work was supported by the Seismic Hazard Center of INGV (Centro per la Pericolosità Sismica, CPS), and it has been partially carried out in the framework of the REAKT (Strategies and tools for Real-Time EArthquake RisK ReducTion) and STREST (Harmonized approach to stress tests for critical infrastructures against natural hazards) projects funded by the European Union's Seventh Programme for research, technological development and demonstration under grant agreement No 282862 and No 603389, respectively, and the project "Quantificazione del Multi-Rischio con approccio Bayesiano: un caso studio per i rischi naturali della città di Napoli", funded by the Italian Ministry of Education, Universities and Research (Ministero dell'Istruzione, dell'Università e della Ricerca). We thank Peter Marion Powers for sharing with us the data used in Figures 5 and 6. TO BE COMPLETED

REFERENCES

Abrahamson, N. A., and J. J. Bommer (2005). Probability and uncertainty in seismic hazard analysis. *Earthquake Spectra* **21**, 603-607.

Akkar, S., and J. Bommer (2010). Empirical equations for the prediction of PGA, PGV, and spectral accelerations in Europe, the Mediterranean region, and the middle east, *Seismol. Res. Lett.* **81**, 195-206.

Akinci, A. (2010). HAZGRIDX: earthquake forecasting model for $M_L \geq 5.0$ earthquakes in Italy based on spatially smoothed seismicity. *Ann. Geophys*. **53**, 51-61.

Albarello, D., and V. D'Amico (2008). Testing probabilistic seismic hazard estimates by comparison with observations: An example in Italy. *Geophys. J. Int*. **175**, 1088-1094.

Bedford, T., and R. Cooke (2001). *Probabilistic risk analysis: foundations and methods*, Cambridge University Press.

Bindi, D., F. Pacor, L. Luzi, R. Puglia, M. Massa, G. Ameri, and R. Paolucci (2011). Ground motion prediction equations derived from Italian strong motion data-base, *Bull. Earthquake Eng.* **9**, 1899-1920.

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

Bowman, A. W., and A. Azzalini (1997). *Applied Smoothing Techniques for Data Analysis*, New York, Oxford University Press.

Cauzzi, C., and E. Faccioli (2008). Broadband (0.05 to 20 s) prediction of displacement response spectra based on worldwide digital records, *J. Seismol.* **12**, 453-475.

Cox, L.A., G.C. Brown, and S.M. Pollock (2008). When is uncertainty about uncertainty worth characterizing? *Interfaces* **38**, 465-468.

Draper D., J. Hodges, C. Mallows and D. Pregibon (1993). Exchangeability and data analysis (with discussion). *J. Roy. Statist. Soc, Ser. A* **156**, 9-37.

Faenza, L., W. Marzocchi (2010). The Proportional Hazard Model as applied to the CSEP forcasting area in Italy. *Ann. Geophys*. **53**, 77-84.

Field, E.H., N. Gupta, V. Gupta, M. Blanpied, P. Maechling, and T. H. Jordan (2005). Hazard calculations for the WGCEP-2002 earthquake forecast using openSHA and distributed object technologies. *Seismol. Res. Lett*. **76**, 161-167.

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

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2003). *Bayesian Data Analysis*, second edition. Chapman and Hall, London.

Gulia, L., S. Wiemer, D. Schorlemmer (2010). Asperity-based earthquake likelihood models for Italy. *Ann. Geophys*. **53**, 63-75.

Jaynes, E.T. (2003). *Probability theory: the logic of Science*, Cambridge Univ. Press, NewYork.

Krinitzsky, E. L. (1995). Problems with logic trees in earthquake hazard evaluation, *Eng. Geol*. **39**, 1-3.

Kulkarni, R. B., R. R. Youngs, and K. J. Coppersmith (1984). Assessment of confidence intervals for results of seismic hazard analysis, Proceedings of the Eighth World Conference on Earthquake Engineering, San Francisco, 1.

Kumamoto, H., and E. Henkley (1996). *Probabilistic risk assessment and management for engineers and scientists*. IEEE Press, Piscataway, NJ.

Leith, C.E. (1974) Theoretical Skill of Monte Carlo Forecasts. *Monthly Weather Rev*. **102**, 409-418.

Lilliefors, H.W. (1967). On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *J. Am. Stat. Assoc*. **62**, 399-402.

Lindley, D.V. (2000). The philosophy of statistics. *Statistician* **49**, 293-337.

Marzocchi W., and T. H. Jordan (2014). Testing for Ontological Errors in Probabilistic Forecasting Models of Natural Systems. *Proc. Nat. Acad. Sci*. **85**, 955-959.

Marzocchi, W., L. Sandri, and J. Selva (2008). BET_EF: a probabilistic tool for long- and short-term eruption forecasting. *Bull. Volcanol*. **70**, 623-632.

Marzocchi W., L. Sandri, J. Selva (2010). BET_VH: a probabilistic tool for long-term volcanic hazard assessment. *Bull. Volcanol*. **72**, 705-716.

Marzocchi, W., J. D. Zechar, and T.H. Jordan (2012). Bayesian forecast evaluation and ensemble earthquake forecasting, *Bull. Seismol. Soc. Am*. **102**, 2574-2584.

Marzocchi, W., A.M. Lombardi, and E. Casarotti (2014). The establishment of an operational earthquake forecasting system in Italy. *Seismol. Res. Lett*. **85**, 961-969.

McGuire, R.K., and T.P. Barnhard (1981). Effects of temporal variations in seismicity on seismic hazard. *Bull. Seismol. Soc. Am*. **71**, 321-334.

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

MPS Working Group (2004). Redazione della mappa di pericolosita sismica prevista dall’Ordinanza PCM del 20 marzo 2003, rapporto conclusivo per il Dipartimento della Protezione Civile (Istituto Nazionale di Geofisica e Vulcanologia, Milano-Roma, http:// zonesismiche.mi.ingv.it).

Musson, R. M. W. (2005). Against fractiles, *Earthquake Spectra* **21**, 887-891.

Musson, R.M.W. (2012). On the nature of logic trees in probabilistic seismic hazard assessment. *Earthq. Spectra* **28**, 1291-1296.

Newhall, C., and R. Hoblitt (2002). Constructing event trees for volcanic crises. *Bull. Volcanol*. **64**, 3-20.

NRC, National Research Council Panel on Seismic Hazard Evaluation (1997). Review of Recommendations for Probabilistic Seismic Hazard Analysis: Guidance on Uncertainty and Use of Experts, National Academy of Sciences, Washington, D.C., 84 pp., ISBN:0-309-56207-4.

Paté-Cornell, M.E. (1996). Uncertainties in risk analysis: Six levels of treatment. *Reliability Engineering and System Safety* **54**, 95-111.

Rhoades, D.A., M. C. Gerstenberger, A. Christophersen, J. D. Zechar, D. Schorlemmer, M. J. Werner, and T. H. Jordan (2014). Regional earthquake likelihood models II: information gains of multiplicative hybrids. *Bull. Seismol. Soc. Am*. **104**, 3072-3083.

Scherbaum, F., and N. M. Kuehn (2011). Logic tree branch weights and probabilities: Summing up to one is not enough, *Earthquake Spectra* **27**, 1237-1251.

Schorlemmer, D., A. Christophersen, A. Rovida, F. Mele, M. Stucchi, W. Marzocchi (2010). Setting up an earthquake forecast experiment in Italy. *Ann. Geophys*. **53**, 1-9.

Selva, J., and L. Sandri (2013). Probabilistic Seismic Hazard Assessment: Combining Cornell-like approaches and data at sites through Bayesian inference, *Bull. Seismol. Soc. Am*. **103**, 1709-1722.

Selva J., S. Argyroudis, and K. Pitilakis (2013). Impact on loss/risk assessments of inter-model variability in vulnerability analysis, *Natural Hazards* **67**, 723-746.

Silver, N. (2012). *The Signal and the Noise: Why So Many Predictions Fail-but Some Don't,* The Penguin Press, New York, New York.

SSHAC, Senior Seismic Hazard Analysis Committee (1997). Recommendations for Probabilistic Seismic Hazard Analysis: Guidance on Uncertainty and Use of Experts, U.S. Nuclear Regulatory Commission, U.S. Dept. of Energy, Electric Power Research Institute; NUREG/CR-6372, UCRL-ID-122160, vol. 1-2.

Stirling M., and M. Petersen (2006). Comparison of the historical record of earthquake hazard with seismic hazard models for New Zealand and the continental United States. *Bull. Seismol. Soc. Am*. **96**, 1978-1994.

Stirling M., and M. Gerstenberger (2010). Ground motion-based testing of seismic hazard models in New Zealand. *Bull. Seismol. Soc. Am*. **100**, 1407-1414.

Stucchi, M., C. Meletti, V. Montaldo, H. Crowley, G.M. Calvi, and E. Boschi (2011). Seismic Hazard Assessment (2003–2009) for the Italian Building Code. *Bull. Seismol. Soc. Am.* **101**, 1885-1911.

Vick, S.G. (2002). *Degrees of belief: subjective probability and engineering judgment*, ASCE Press, Reston, VA.

Zechar, J.D., T.H. Jordan (2010). Simple smoothed seismicity earthquake forecasts for Italy. *Ann. Geophys*. **53**, 99-105.

Figure captions

Figure 1. Probability tree of coin toss (see text for more details). On the right end side of the tree, the weight of the path Pr(H_i) (in blue) and the branch value Pr(E| H_i) (in black) are reported.

Figure 2. Logic tree for the seismic hazard analysis in Cosenza and Bologna. The first five branches on the left represent the earthquake rate models; the second three branches are the GMPEs used. The description of the models is reported in the text and in the cited references. On the right end side of the tree, the (arbitrary) weight of each branch is reported.

Figure 3. a) Mean and 10-th, 50-th and 90-th percentiles of the PGA seismic hazard curve for the city of Cosenza; The vertical line marks one specific ground shaking value that is used for the other panels of the figure. b) 50-years exceedance probability distribution for a PGA of 0.15 g. The vertical gray lines show the outcomes of the logic tree, and the height is the weight of each datum; the black line is the Beta PDF estimated by the data using equations 2 and 3. c) The empirical cumulative distribution of the logic tree outcomes (in gray), and the cumulative distribution of the Beta distribution (in black).

Figure 4. As for Figure 3, but relative to the city of Bologna.

Figure 5. PDF (left *y*-axis) and hystogram (right *y*-axis) of the UCERF3 logic tree exceedance probabilities relative to the reference PGA (2% in 50 years) for the site of Los Angeles.

Figure 6. As for Figure 5, but relative to the site of Redding.

Figure 1 with caption

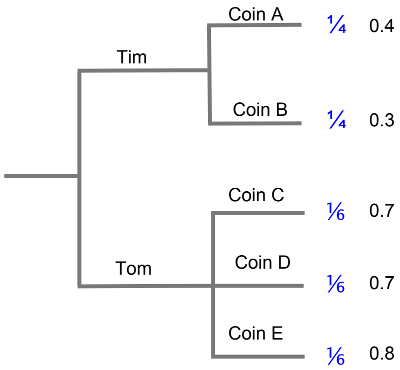


Figure 1. Probability tree of coin toss (see text for more details). On the right end side of the tree, the weight of the path $\Pr(H_i)$ (in blue) and the branch value $\Pr(E|H_i)$ (in black) are reported.

Figure 2 with caption

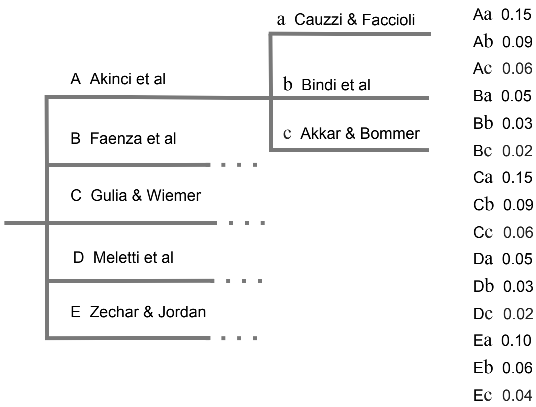


Figure 2. Logic tree for the seismic hazard analysis in Cosenza and Bologna. The first five branches on the left represent the earthquake rate models; the second three branches are the GMPEs used. The description of the models is reported in the text and in the cited references. On the right end side of the tree, the (arbitrary) weight of each branch is reported

Figure 3 with caption

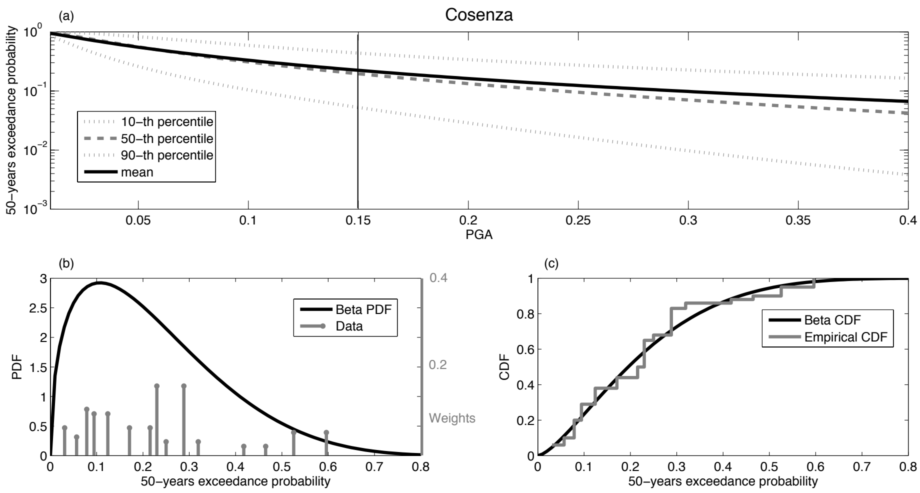


Figure 3. a) Mean and 10-th, 50-th and 90-th percentiles of the PGA seismic hazard curve for the city of Cosenza; The vertical line marks one specific ground shaking value that is used for the other panels of the figure. b) 50-years exceedance probability distribution for a PGA of 0.15 g. The vertical gray lines show the outcomes of the logic tree, and the height is the weight of each datum; the black line is the Beta PDF estimated by the data using equations 2 and 3. c) The empirical cumulative distribution of the logic tree outcomes (in gray), and the cumulative distribution of the Beta distribution (in black).

Figure 4 with caption

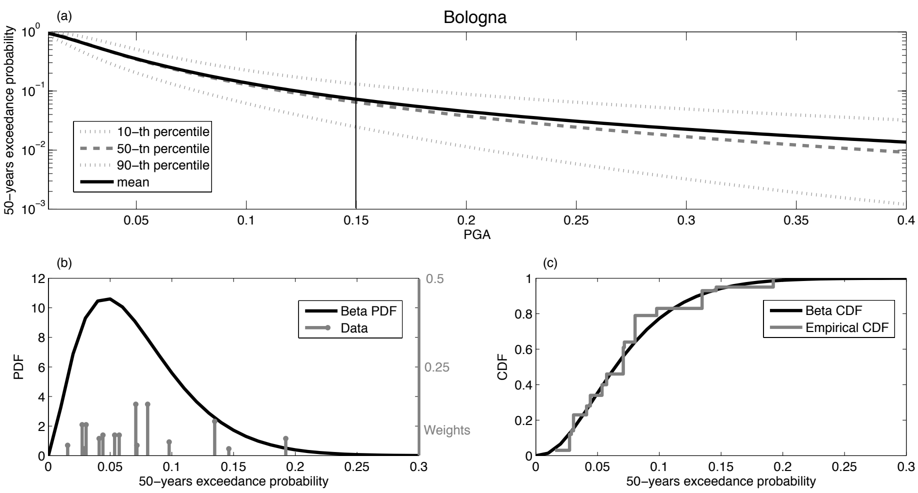


Figure 4. As for Figure 3, but relative to the city of Bologna.

Figure 5 with caption

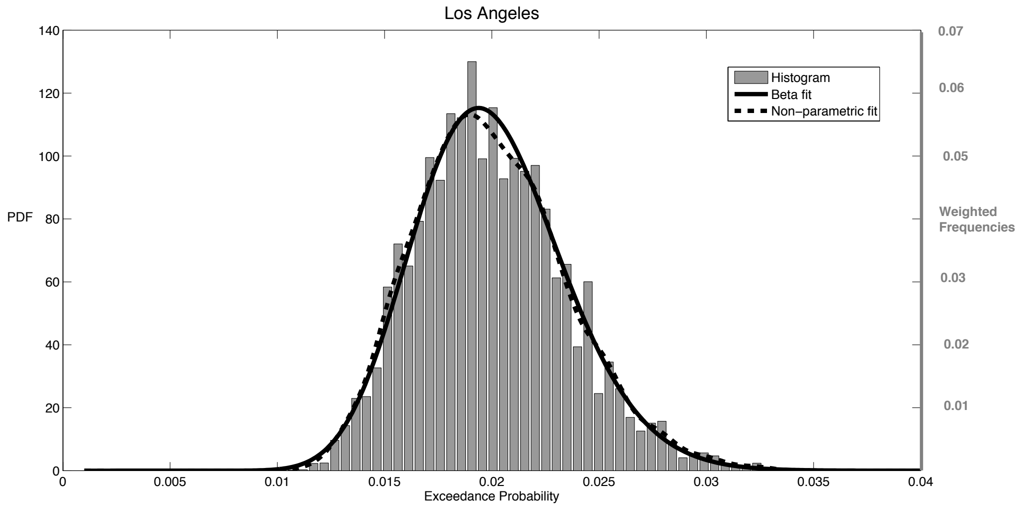


Figure 5. PDF (left *y*-axis) and hystogram (right *y*-axis) of the UCERF3 logic tree exceedance probabilities relative to the reference PGA (2% in 50 years) for the site of Los Angeles.

Figure 6 with caption

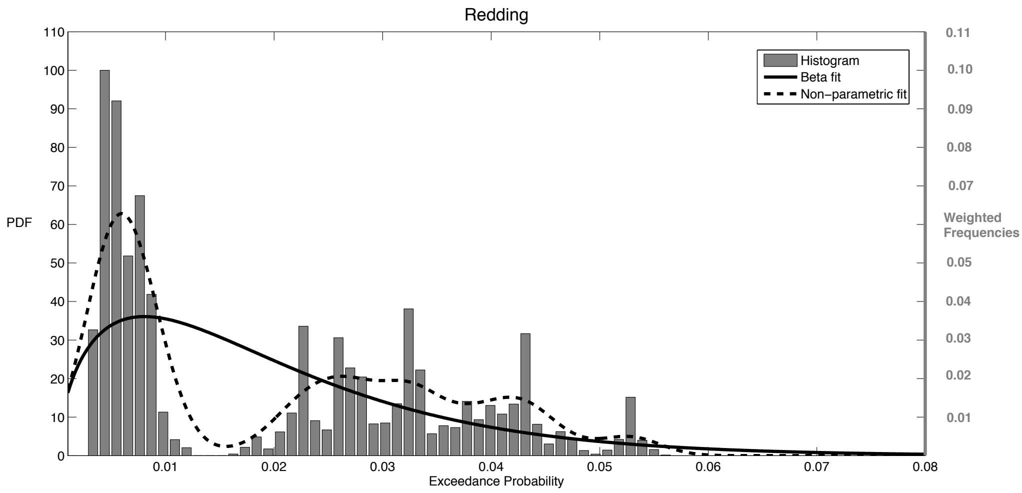


Figure 6. As for Figure 5, but relative to the site of Redding

Figure 1

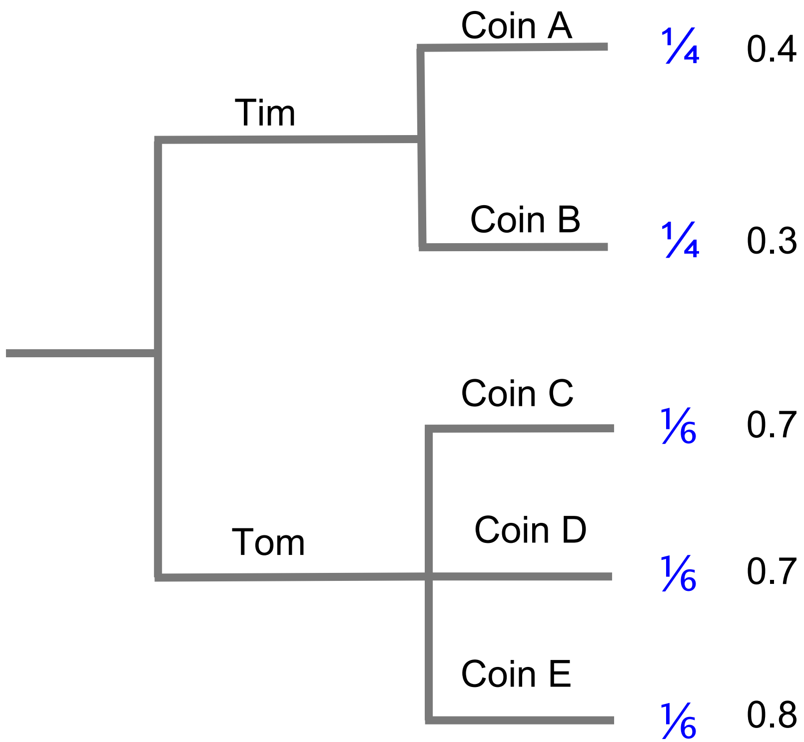


Figure 2

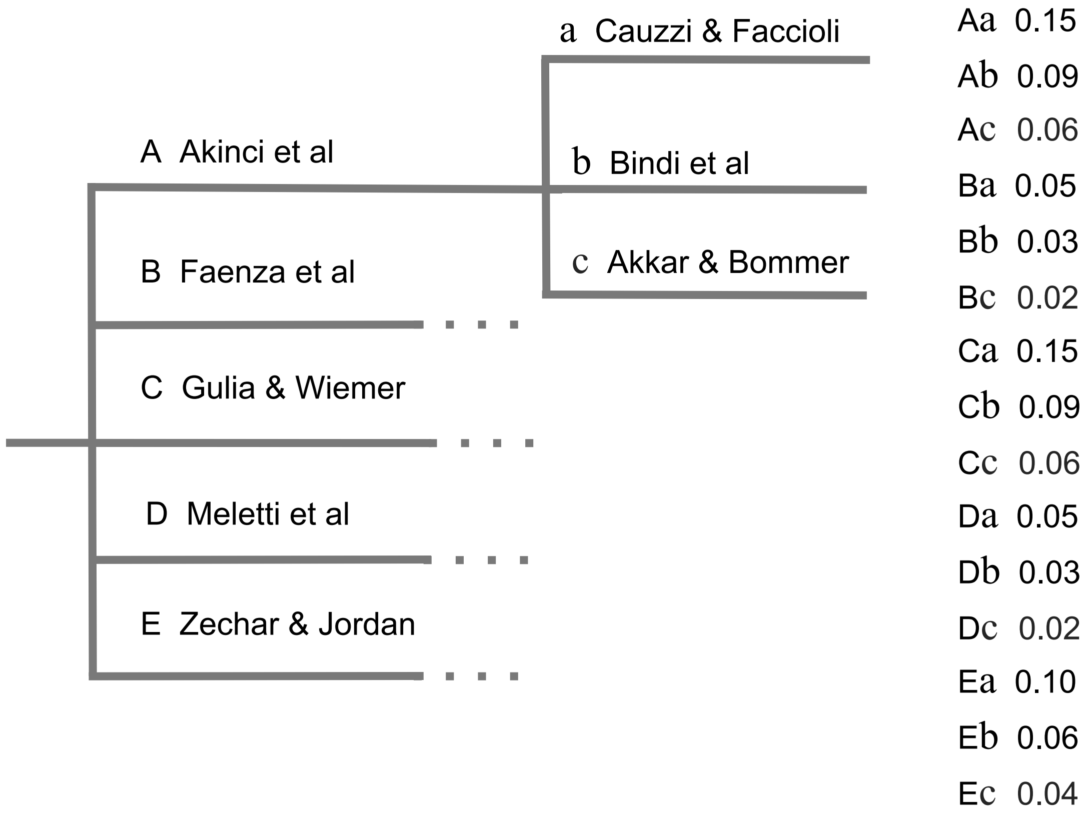

a Cauzzi & Faccioli
A Akinci et al
b Bindi et al
c Akkar & Bommer
B Faenza et al
C Gulia & Wiemer
D Meletti et al
E Zechar & Jordan
Aa 0.15
Ab 0.09
Ac 0.06
Ba 0.05
Bb 0.03
Bc 0.02
Ca 0.15
Cb 0.09
Cc 0.06
Da 0.05
Db 0.03
Dc 0.02
Ea 0.10
Eb 0.06
Ec 0.04

Figure 3

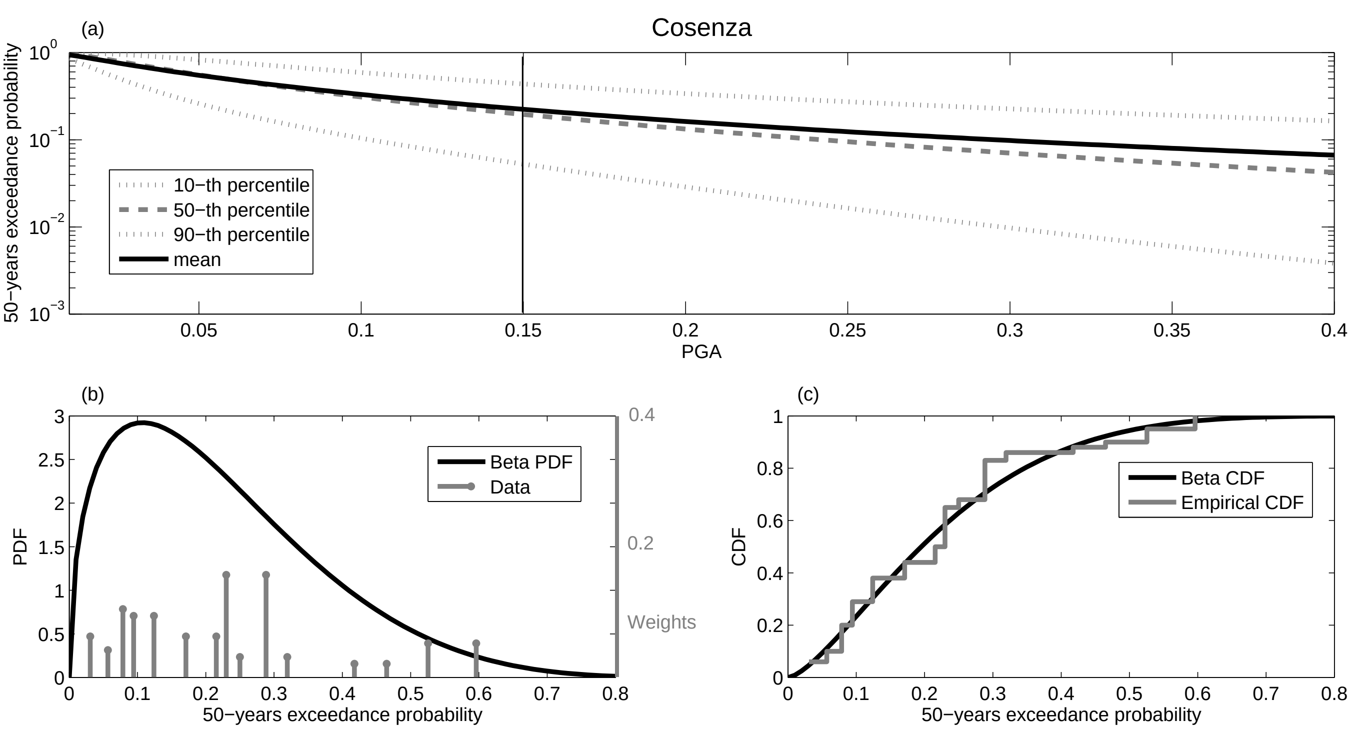
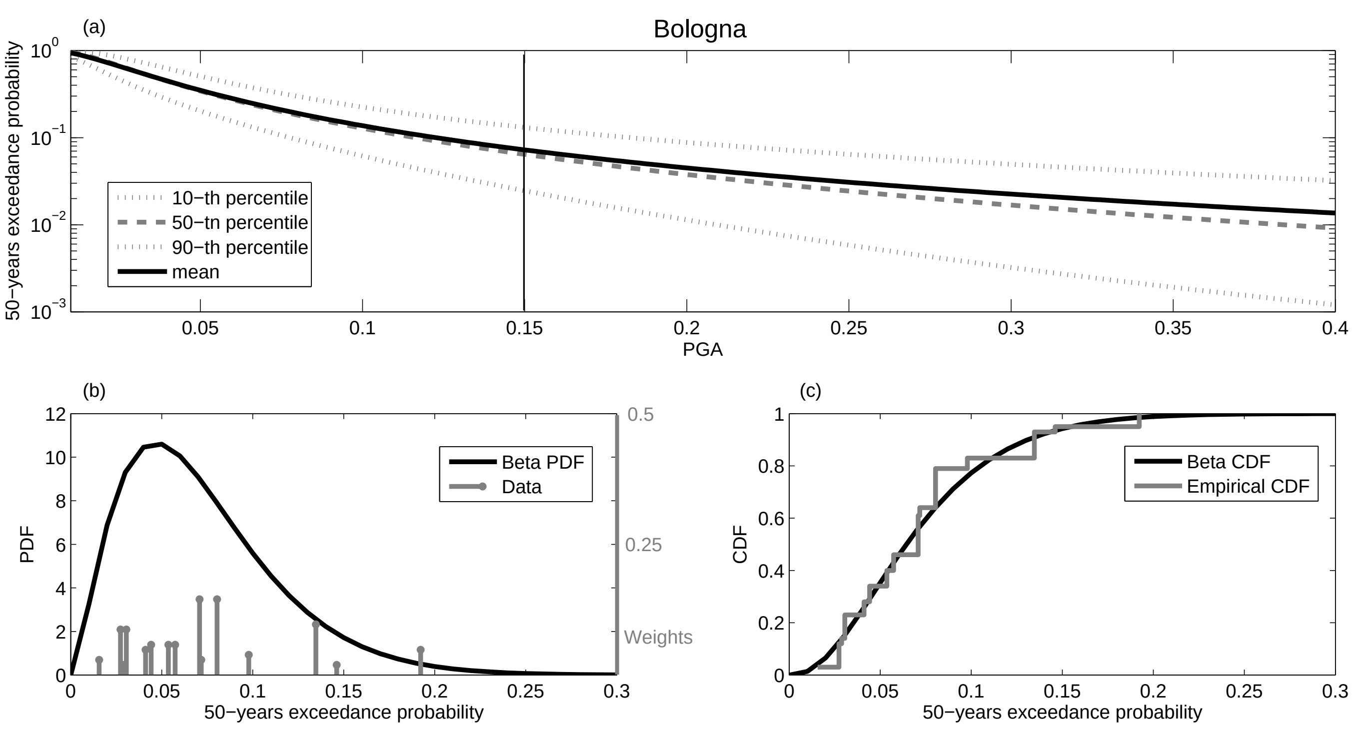
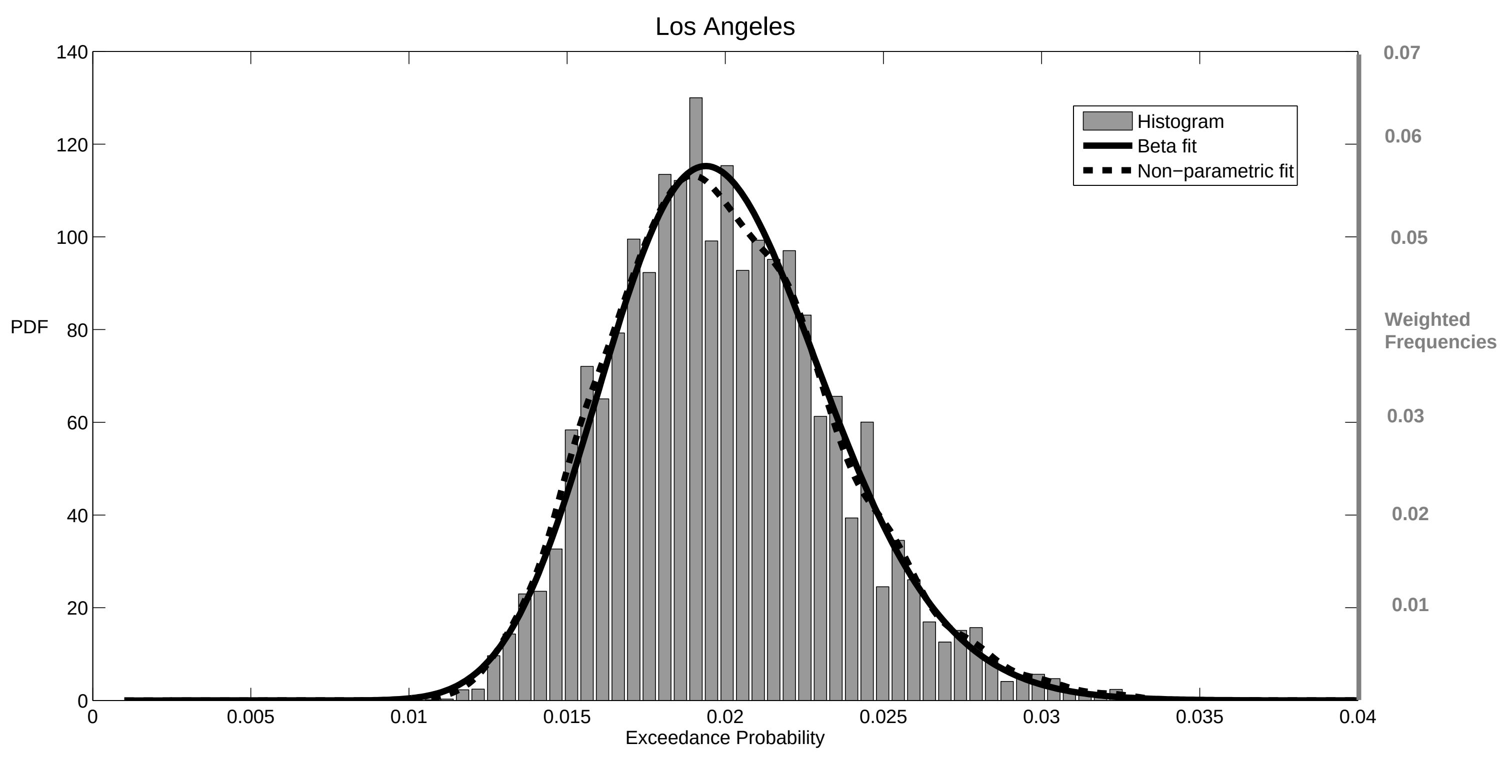

(a)
Cosenza
50−years exceedance probability
10−th percentile
50−th percentile
90−th percentile
mean
PGA
(b)
PDF
Beta PDF
Data
Weights
50−years exceedance probability
(c)
CDF
Beta CDF
Empirical CDF
50−years exceedance probability

Figure 4



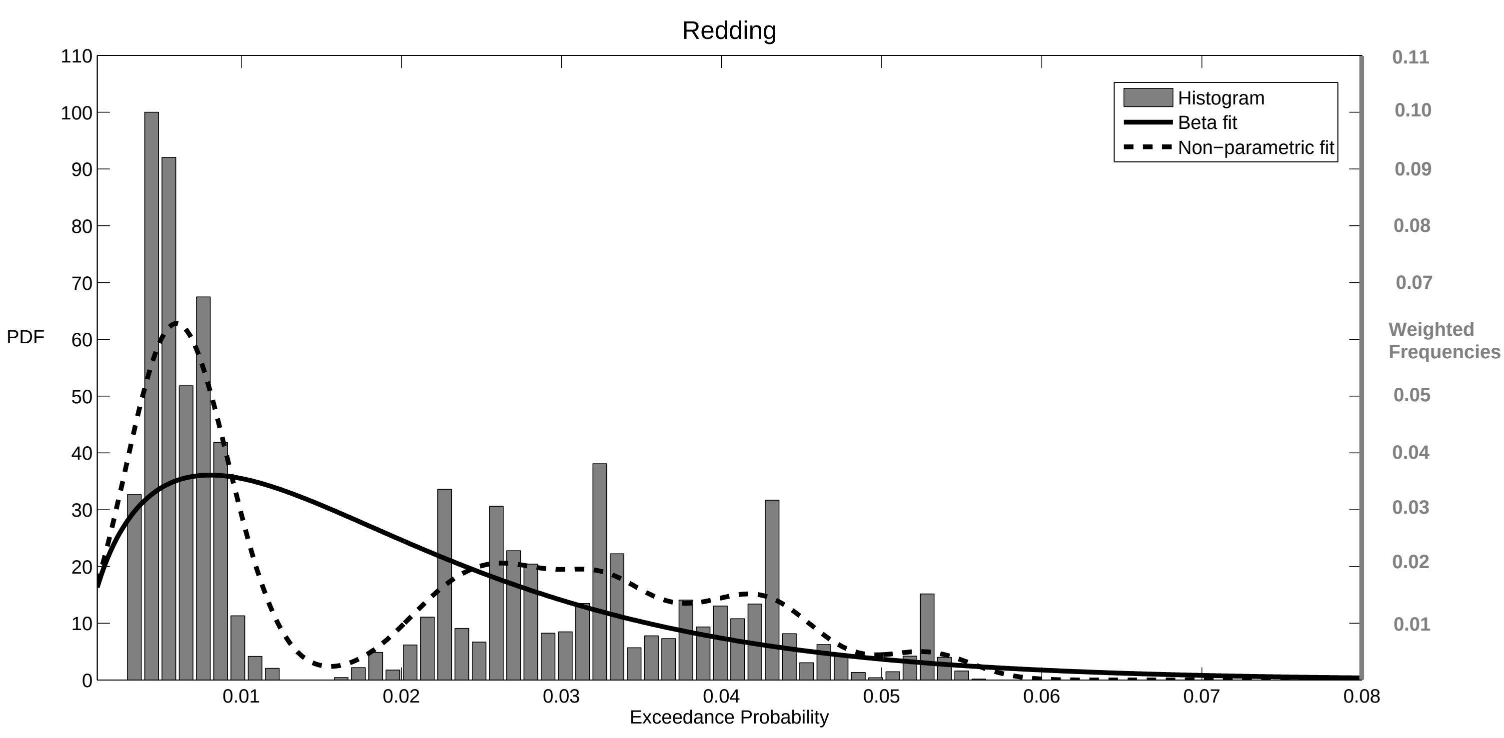
(a)
Bologna
50−years exceedance probability
PGA
10−th percentile
50−tn percentile
90−th percentile
mean
(b)
PDF
Weights
Beta PDF
Data
50−years exceedance probability
(c)
CDF
Beta CDF
Empirical CDF
50−years exceedance probability

Figure 5



Los Angeles
Histogram
Beta fit
Non−parametric fit
PDF
Weighted
Frequencies
Exceedance Probability
140
120
100
80
60
40
20
0
0.07
0.06
0.05
0.03
0.02
0.01
0
0.005
0.01
0.015
0.02
0.025
0.03
0.035
0.04

Figure 6


Redding
Histogram
Beta fit
Non−parametric fit
PDF
Weighted Frequencies
Exceedance Probability
0
10
20
30
40
50
60
70
80
90
100
110
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.10
0.11