

# Collapsing epistemic uncertainty in frequency-magnitude distributions of regional seismic hazard models

Nick Ackerley <sup>1, 2</sup>      Graeme Weatherill <sup>3</sup>

December 31, 2016

## Abstract

Logic trees are an accepted way of modelling epistemic uncertainty in seismic hazard. Source model logic trees pose no particular computational difficulty in site-specific studies involving on the the order of ten source zones and a small number of sites. Regional or national hazard models can require on the order of a hundred zones, however, and in these cases, with a relatively simple source model logic tree with on the order of ten branches, the hazard integral can turn out to be too computationally intensive to compute.

We show that in a classical probabilistic seismic hazard assessment, the complexity of a mean hazard calculation can be drastically reduced by “collapsing” the epistemic uncertainty for each source zone, prior to computation of the hazard integral. Our result is exact but is available only for the mean hazard. Full enumeration of logic trees is still required to obtain an exact result for medians or other quantiles, however we propose procedures for pruning of trees which can greatly reduce the complexity of such computations while introducing negligible error.

## 1 Introduction

The established procedure for incorporating epistemic uncertainties in probabilistic seismic hazard models is to construct logic trees to weight different possible models (Bommer and Scherbaum, 2008). Ground-motion prediction equations (GMPEs), for example, are assigned weights according to the likelihood that each is representative for a given source zone (Delavaud et al., 2012; CEUSSSCn, 2012). It is routine to further weight different models of seismicity (Field et al., 2014; Woessner et al., 2015), specifically frequency-magnitude distributions (FMDs), which could give rise to the actual instrumental, historic and paleoseismic records. Assignment of logic tree weights to maximum magnitude is common, but different b-values are also sometimes weighted independently. When applying these methods to a site-specific study, no particular difficulty arises.

Regional models can have hundreds of source zones, each with its own logic trees for GMPEs and FMDs. For full enumeration of the logic tree for such a model the number of source zones must be raised to the power of the total number of logic tree branches. Because of this, methods for the reduction of logic tree branches are crucial. Procedures for assessing the appropriateness of GMPEs include data-driven (Scherbaum et al., 2009; Kale and Akkar, 2013) and non-data-driven (Bommer et al., 2010) approaches. This means there are numerous strategies for reducing GMPE

---

<sup>1</sup>Istituto Universitario di Studi Superiori, Pavia, Italy

<sup>2</sup>Geological Survey of Canada, Ottawa, Canada

<sup>3</sup>Global Earthquake Model (GEM), Pavia, Italy

logic trees to the minimum necessary complexity. Aside from noting that it is inappropriate to treat a- and b-values as independent variables in FMD logic trees (Bommer and Scherbaum, 2008), little guidance exists for the reduction of FMD logic tree complexity. Partial enumeration of the logic tree is one means to reduce computational complexity, but it is not clear how one can ensure that the subset chosen properly represents the spread of the distribution.

In the next section it is shown that the mean hazard can be computed exactly by “collapsing” the FMDs prior to evaluation of the hazard integral. A simple case study for which the mean hazard can be computed by full enumeration follows. It is shown that the mean hazard is exact as long as the mean annual rate of exceedance is sufficiently low. These results are discussed and their ramifications for regional seismic hazard models are explored. Finally, methods for pruning logic trees of complex regional models are discussed.

## 2 Computation of mean hazard given epistemic uncertainty

In classical PSHA the hazard integral gives an estimate of rate at which the ground motion  $Y$  at a site of interest will exceed some value  $y$ . For a set of  $N_S$  sources generating ruptures of  $N_M$  magnitudes  $m_j$  at  $N_R$  distances  $r_k$  using discrete summations, (Baker, 2008):

$$\lambda(Y > y) = \sum_{i=1}^{N_S} \lambda(M_i > m_{\min}) \sum_{j=1}^{N_M} \sum_{k=1}^{N_R} P_i(Y > y|m_j, r_k) P_i(M_i = m_j) P_i(R_i = r_k) \quad (1)$$

Where for source  $i$ :

- $\lambda_i(M_i > m_{\min})$  is the rate of earthquakes greater than  $m_{\min}$
- $P_i(Y > y|m_j, r_k)$  is the ground motion prediction equation (GMPE) and incorporates aleatory uncertainty in the ground motion
- $P_i(M_i = m_j)$  is the frequency-magnitude distribution (FMD)
- $P_i(R_i = r_k)$  is the distribution of distance measures from points within the source to the site of interest.

For a single point source, neglecting finite source effects, this simplifies to:

$$\lambda(Y > y) = \lambda(M > m_{\min}) \sum_{j=1}^{N_M} P(Y > y|m_j, R) P(M = m_j)$$

But now suppose that certain epistemic uncertainties have been identified. In order to estimate the effect of this lack of knowledge on the mean and spread of ground motions suppose  $N_G$  GMPEs  $P_\ell(Y > y|m_j, R)$  have been assigned weights  $w_\ell$  and  $N_F$  FMDs  $P_m(M = m_j)$  have been assigned weights  $w_m$ . Now

$$\lambda(Y > y) = \lambda(M > m_{\min}) \sum_{m=1}^{N_F} w_m \sum_{\ell=1}^{N_G} w_\ell \sum_{j=1}^{N_M} P_\ell(Y > y|m_j, R) P_m(M = m_j)$$

Since we can reorder the summations to obtain

$$\lambda(Y > y) = \lambda(M > m_{\min}) \sum_{\ell=1}^{N_G} w_\ell \sum_{j=1}^{N_M} P_\ell(Y > y|m_j, R) \sum_{m=1}^{N_F} w_m P_m(M = m_j)$$

We can pre-compute the “collapsed” FMD for the source:

$$P_C(M = m) \equiv \sum_{m=1}^{N_F} w_m P_m(M = m)$$

And this simplifies the hazard summation to

$$\lambda(Y > y) = \lambda(M > m_{\min}) \sum_{\ell=1}^{N_G} w_{\ell} \sum_{j=1}^{N_M} P_{\ell}(Y > y | m_j, R) P_C(M = m_j)$$

Note that a similar “collapse” of GMPE uncertainty is not possible because the ground motion is conditional upon the magnitude. [check again and clarify!]

Returning to the case of multiple sources, we can include ground-motion and frequency-magnitude epistemic uncertainty by writing:

$$\lambda(Y > y) = \sum_{i=1}^{N_S} \lambda_i(M_i > m_{\min}) \sum_{m=1}^{N_F} w_{m,i} \sum_{\ell=1}^{N_G} w_{\ell,i} \sum_{j=1}^{N_M} \sum_{k=1}^{N_R} P_{\ell,i}(Y > y | m_j, r_k) P_{m,i}(M_i = m_j) P_i(R_i = r_k)$$

We can still reorder the summations, as long as we recognize that the collapsed FMD is different for each source

$$\lambda(Y > y) = \sum_{i=1}^{N_S} \lambda(M_i > m_{\min}) \sum_{\ell=1}^{N_G} w_{\ell,i} \sum_{j=1}^{N_M} \sum_{k=1}^{N_R} P_{\ell,i}(Y > y | m_j, r_k) P_i(R_i = r_k) \sum_{m=1}^{N_F} w_{m,i} P_{m,i}(M_i = m_j)$$

So that for source  $i$  the collapsed FMD is

$$P_{C,i}(M_i = m) \equiv \sum_{m=1}^{N_F} w_{m,i} P_{m,i}(M_i = m) \quad (2)$$

The important point here is that the “collapsed” FMD can be pre-computed for each source. The number of logic tree branches in full enumeration after collapsing FMD uncertainty, and thus the computation complexity and time, are thus greatly reduced.

### 3 Case Study: Guwahati

The effectiveness of this method of “collapsing” FMD logic trees is demonstrated using a case study for the city of Guwahati in India. The source regions, GMPE and FMD logic trees used are part of an implementation (Ackerley, 2017) using the OpenQuake platform (Pagani et al., 2014) of a PSHA model for the Indian subcontinent (Nath and Thingbaijam, 2012). Figure 1 shows an example of the pre-computation of collapsed FMDs as described in equation (2) for a single zone.

In order to show the effectiveness of “collapsing” FMD logic trees a greatly simplified subset of the full regional model was used. In particular, a single site (Guwahati), and just two areal source zones (one with the FMD shown in Figure 1 plus another zone the same FMD logic tree structure) were considered. This makes it possible to rapidly compute the hazard by fully enumerating the logic trees.

With this extremely simplified model the computation time dropped by a factor of 40 when FMD uncertainty was collapsed. In realistic regional models, with gridded sites and hundreds

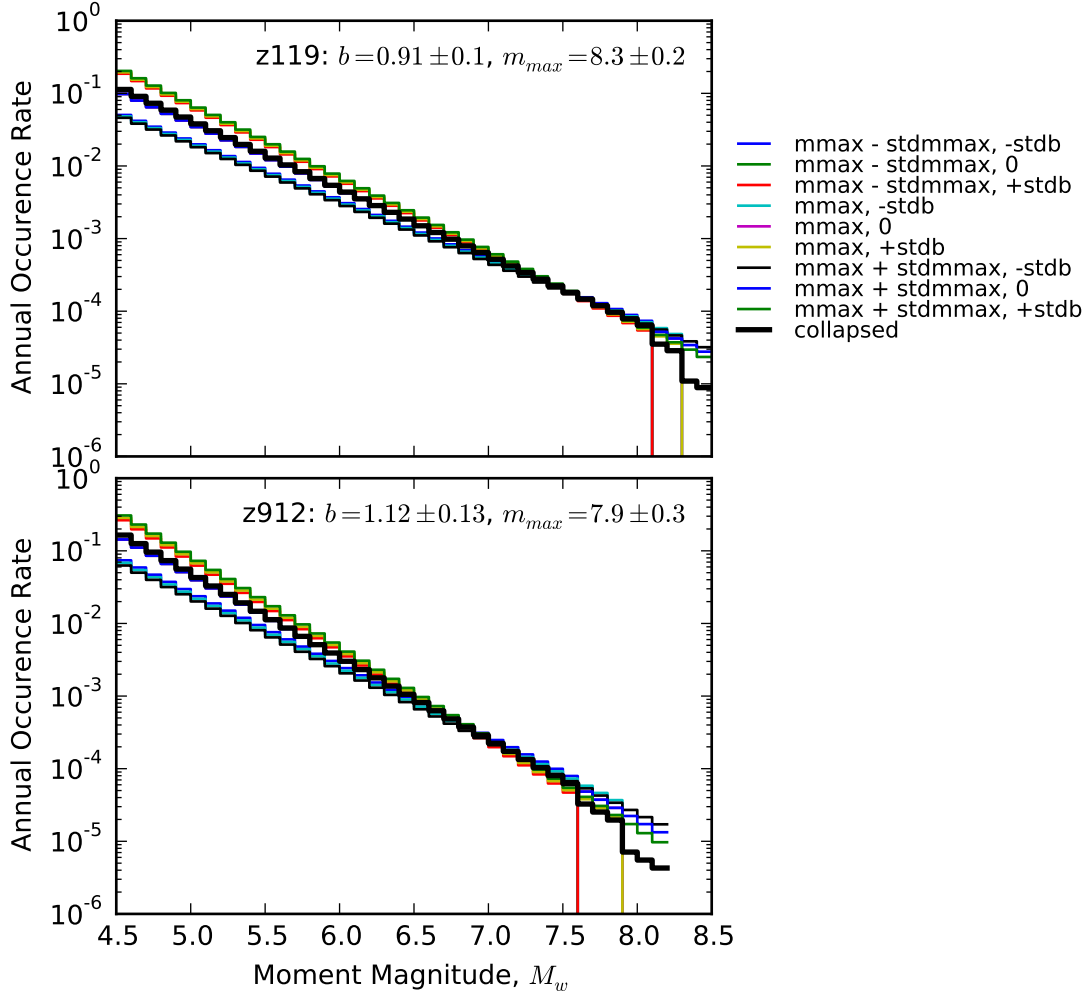


Figure 1: Frequency-magnitude distributions corresponding to various branches of the logic tree of Nath and Thingbaijam (2012). The zones chosen are the two closest to Guwahati, India. Zone 119 at 25-50 km depth is the zone capable of producing a devastating “pop-up” type event (Bilham et al., 2001; Nath et al., 2012). The “collapsed” FMD uncertainty computed using (2) is shown as a heavy black line.

of areal sources, each with FMD uncertainty, the computational savings are far greater, and can make an intractable problem tractable.

The analysis of the previous section predicts that collapsed FMDs should give an exact result for the mean exceedance rate. Figure 2 in fact shows that this procedure gives a very good approximation of the mean hazard for low but not for high probabilities of exceedance.

[Some further investigation is needed to find out what determines this threshold; if it is dependent on the return period chosen alone then inaccuracy at low-probability of exceedance is effectively a computational artefact, a curiosity of no engineering significance. Another possibility is that it is a result of the discretisation of the FMD used; in this case it would be possible to

reduce the error by discretise the FMDs using bin widths of say 0.01 magnitude units instead of 0.1, but there would be a corresponding computational penalty.

Perhaps it would be better to attempt a theoretical treatment of the errors. Conceivably it is the fully-enumerated model which is in error! A reference as a starting point would be good.]

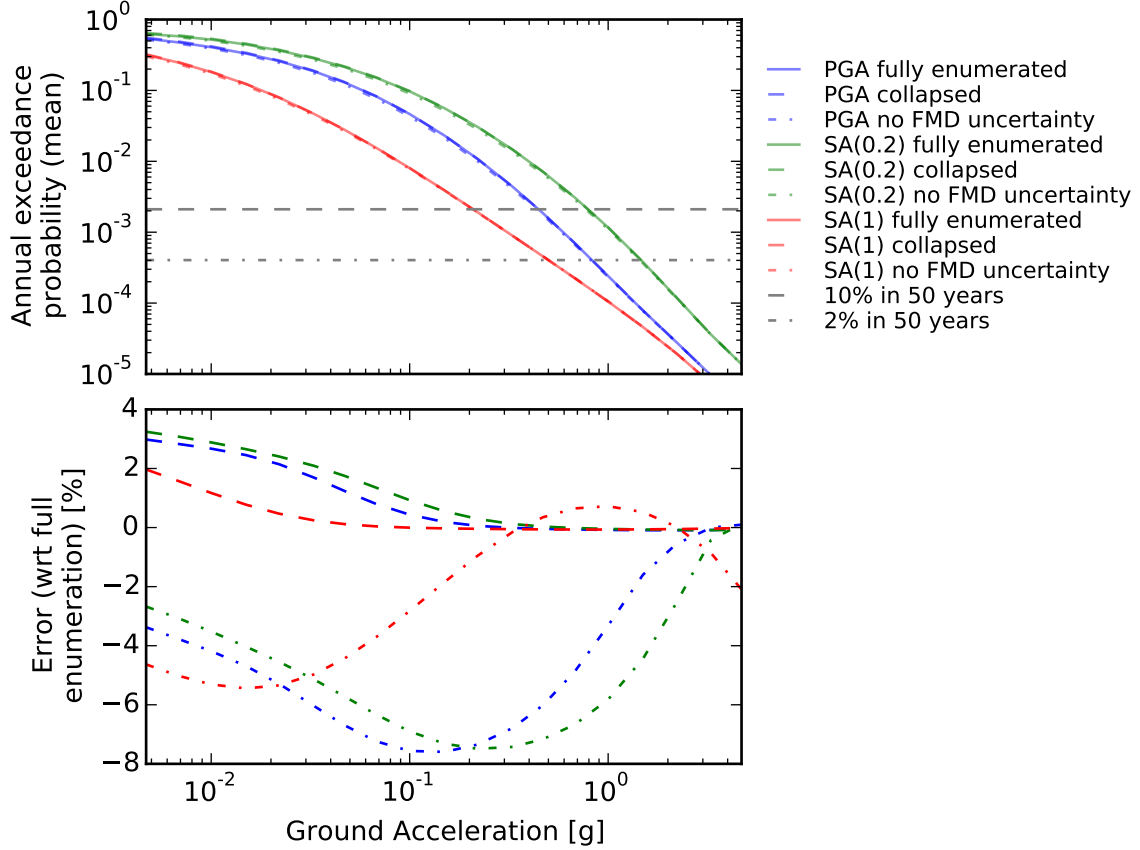


Figure 2: Mean hazard curves computed using various levels of FMD uncertainty. The site of interest is the city of Guwahati. The source model consists only of zones 119 and 912 of Nath and Thingbaijam (2012). The full GMPE logic tree of Nath and Thingbaijam (2012) is used. The “fully enumerated” result implements the FMD logic tree described in Nath and Thingbaijam (2012) while “collapsed” implements (2) and “no FMD uncertainty” models only the FMD logic tree branches with the largest weights.

Unfortunately the “collapse” of logic trees in source zones is of no use in computing the median or any other quantile of hazard. Figure 3 demonstrates, for the same simplistic example, that the error in the calculated median hazard is much too large, as large as if epistemic uncertainty in source seismicity was neglected entirely. The problem is that the most basic procedure for calculating a quantile is to compute the hazard for all (every combination of all branches of all) realizations of the logic trees and then sort them. The hazard at a given quantile is unaffected by the values of hazard at other quantiles except inasmuch as it might affect sort order near the specified quantile.

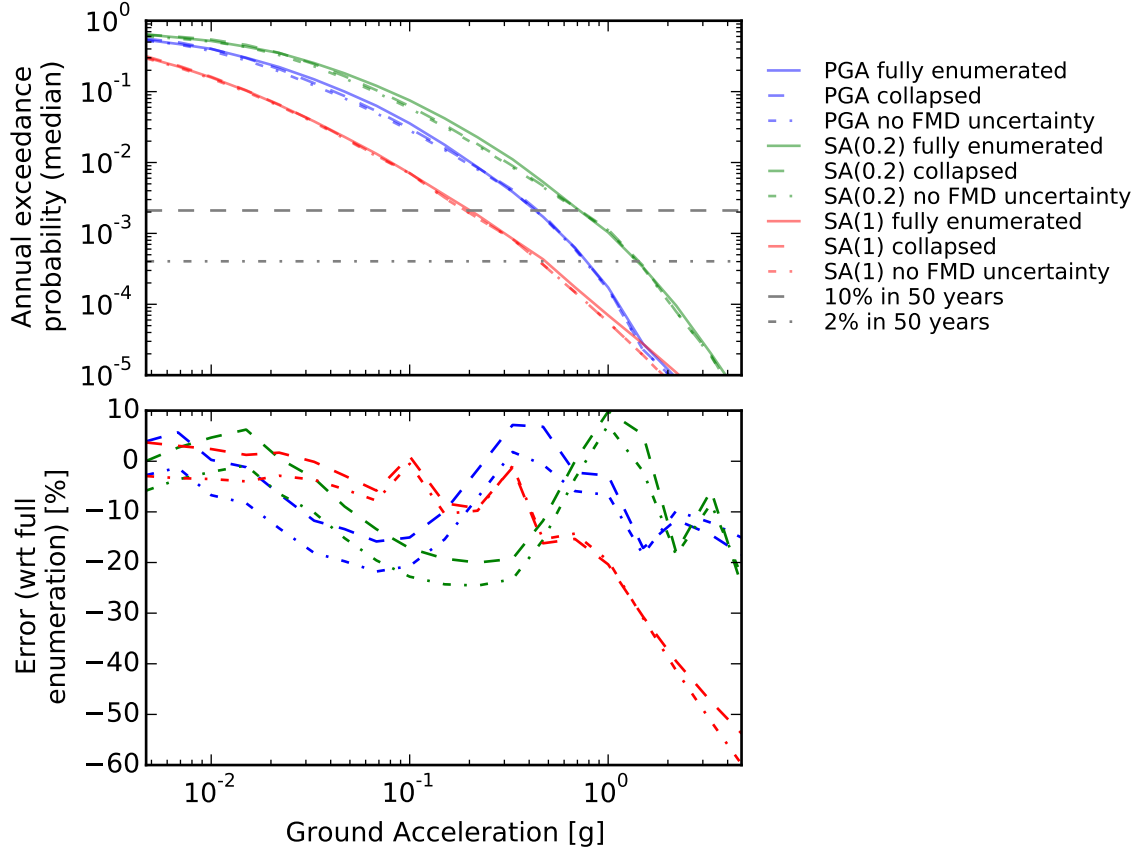


Figure 3: Median hazard curves computed using various levels of FMD uncertainty. For detailed description see Figure 2

## 4 Discussion

The “collapse” of FMD logic trees is practicable in using any software package capable of implementing arbitrary FMDs. However this method’s utility would be much greater if it were implemented directly in the software. Seismic hazard models are typically specified as an inter-related set of files describing geometry and other characteristics of the source zones, GMPE logic trees and FMD logic trees. If the user has to recompute the source zone characteristics in order to eliminate the FMD logic trees and calculate mean hazard, then one effectively cannot use that same model to compute the median hazard. If on the other hand the software can be relied upon to take the full FMD logic trees as input and “collapse” them when appropriate (i.e. when only the mean hazard is requested), then the same model can then be used to compute medians and other quantiles. This is an important benefit, because mean hazard is an important indicator of quality assurance, and the ability to compute it quickly permits more rapid development of complex models as the regional scale and beyond.

Finally a few words about possible efficiencies in the computation of hazard quantiles. Aleatory variability is modelled in most if not all GMPEs using a log-normal distribution so the ability to compute an arbitrary quantile of ground-motion exceedance is built-in. For epistemic variability

the first step in evaluating quantiles is simply the enumeration of the possible realizations, but even this can pose a problem for regional models. For a region with  $N_S = 64$  sources, each with  $N_B = 8$  frequency-magnitude distribution logic-tree branches and  $N_G = 4$  GMPE logic-tree branches the total number of logic tree realizations would be  $N = (N_B N_G)^{N_S} = (2^5)^{2^6} = 2^{320}$ . This is much larger than the largest integer representable on a 64-bit computer and thus the realizations are thus effectively uncountable without extra precautions. Fortunately, depending on the expected level of hazard in a sub-region, and the relative levels of seismicity in nearby sources, it is generally possible to set a cutoff radius (e.g. 200 km), beyond which source zones need not be considered. For the same regional model with 64 sources, suppose only  $N'_S = 8$  source zones are within the cutoff radius. Suddenly the total number of logic tree realizations is reduced to  $N' = (N_B N_G)^{N'_S} = (2^5)^{2^3} = 2^{40}$ , a much more manageable number. Effectively most of the logic tree branches can be “pruned” away by this process. It should furthermore be possible to come up with an efficient procedure to find subsets of gridded sites within a region which have the same set of source zones within the cutoff radius, allowing significant reuse of subsets of the hazard calculation. Implementation of logic tree pruning in a hazard calculation platform such as OpenQuake would enable the calculation of hazard quantiles for regional models.

## 5 Conclusions

The “collapse” of FMD logic trees by weighted summation of branches yields identical results to full enumeration as long as the annual probability of exceedance is low.

## References

- Ackerley, N. (2017). An open model for probabilistic seismic hazard assessment on the Indian subcontinent (unpublished). Technical report, Global Earthquake Model (GEM).
- Baker, J. W. (2008). An introduction to probabilistic seismic hazard analysis. *Report for the US Nuclear Regulatory Commission, page Version, 1*.
- Bilham, R., Gaur, V. K., and Molnar, P. (2001). Himalayan seismic hazard. *Science*, 293(5534):1442–4.
- Bommer, J. J., Douglas, J., Scherbaum, F., Cotton, F., Bungum, H., and Fäh, D. (2010). On the selection of ground-motion prediction equations for seismic hazard analysis. *Seismological Research Letters*, 81(5):783–793.
- Bommer, J. J. and Scherbaum, F. (2008). The use and misuse of logic trees in probabilistic seismic hazard analysis. *Earthquake Spectra*, 24(4):997–1009.
- CEUSSSCn (2012). Central and Eastern United States seismic source characterization for nuclear facilities. Technical report, EPRI, U.S. DOE, and U.S. NRC. <http://www.ceus-ssc.com/Report/Downloads.html> Accessed 2016-03-01.
- Delavaud, E., Cotton, F., Akkar, S., Scherbaum, F., Danciu, L., Beauval, C., Drouet, S., Douglas, J., Basili, R., Sandikkaya, M. A., et al. (2012). Toward a ground-motion logic tree for probabilistic seismic hazard assessment in Europe. *Journal of Seismology*, 16(3):451–473.
- Field, E. H., Arrowsmith, R. J., Biasi, G. P., Bird, P., Dawson, T. E., Felzer, K. R., Jackson, D. D., Johnson, K. M., Jordan, T. H., Madden, C., et al. (2014). Uniform california

- earthquake rupture forecast, version 3 (UCERF3) the time-independent model. *Bulletin of the Seismological Society of America*, 104(3):1122–1180.
- Kale, Ö. and Akkar, S. (2013). A new procedure for selecting and ranking ground-motion prediction equations (gmpes): The euclidean distance-based ranking (edr) method. *Bulletin of the Seismological Society of America*, 103(2A):1069–1084.
- Nath, S. K. and Thingbaijam, K. K. S. (2012). Probabilistic seismic hazard assessment of India. *Seismological Research Letters*, 83(1):135–149.
- Nath, S. K., Thingbaijam, K. K. S., Maiti, S. K., and Nayak, A. (2012). Ground-motion predictions in Shillong region, northeast India. *Journal of Seismology*, 16(3):475–488.
- Pagani, M., Monelli, D., Weatherill, G., Danciu, L., Crowley, H., Silva, V., Henshaw, P., Butler, L., Nastasi, M., Panzeri, L., et al. (2014). Openquake engine: an open hazard (and risk) software for the global earthquake model. *Seismological Research Letters*, 85(3):692–702.
- Scherbaum, F., Delavaud, E., and Riggelsen, C. (2009). Model selection in seismic hazard analysis: An information-theoretic perspective. *Bulletin of the Seismological Society of America*, 99(6):3234–3247.
- Woessner, J., Laurentiu, D., Giardini, D., Crowley, H., Cotton, F., Grünthal, G., Valensise, G., Arvidsson, R., Basili, R., Demircioglu, M. B., et al. (2015). The 2013 european seismic hazard model: key components and results. *Bulletin of Earthquake Engineering*, 13(12):3553–3596.