An Open Model of Probabilistic Seismic Hazard Assessment for the Indian Subcontinent

N. Ackerley[1,2]

[1]Istituto Universitario di Studi Superiori, Pavia, Italy
[2]Université Joseph Fourier, Grenoble, France

February 2, 2016

Abstract

Open models encourage peer review and collaboration; open models can be built upon.

Contents

List of Figures

List of Tables

1 Introduction

In this work a seismic hazard model for peninsual India Nath and Thingbaijam (2012) is implemented within the OpenQuake (Weatherill, 2014; Crowley et al., 2015) platform.

This report is intended to be archived with the input and output files necessary to replicate the results at `https://hazardwiki.openquake.org/`. References to file names in the electronic data are shown in `typewriter` font, as are keywords specific to OpenQuake, such as `bGRRelative`.

1.1 Seismic hazard in the Indian subcontinent

Seismic hazard in India has of course been studied extensively, progressing from deterministic studies to probabilistic seismic hazard assessment (PSHA) and from site-specific to regional studies. Ashish et al. (2016) gives an up-to-date overview of the importance and history of this work. Of particular note is the fact that the Bureau of Indian Standards has not updated their seismic hazard zonation since 2002 (Bureau of Indian Standards, 2002). Nath and Thingbaijam (2012) summarizes concerns with the standard currently in force.

Some studies have focused on the extreme hazard of the Himalayas (Bilham et al., 2001) in the the northeast, including the Shillong plateau, (Das et al., 2006) and northwest (Mahajan et al., 2009). Other studies have focused on regions of lesser but nonetheless high hazard such as Gujarat (Yadav et al., 2008) or considered the whole of stable "peninsular India" (Jaiswal and Sinha, 2007; Ashish et al., 2016). Only Bhatia et al. (1999) considered the whole of India, but as Ashish et al. (2016) points out, since it was part of a global hazard mapping project (GSHAP) it only included "only a few sources for Peninsular India focusing on the inter-plate region along the Himalayan belt".

Nath and Thingbaijam (2012) is thus distinguished from previous work in providing a detailed probabilistic hazard assessment for the whole of India, including neighbouring states such as Bangladesh and Nepal. It is the culmination of several previous works, some unpublished, involving the same group of authors. These works include development of a uniform catalogue (Nath et al., 2010), development of GMPEs specific to the Shillong region (Nath et al., 2012), evaluation of a suite of GMPEs applicable to India (Nath and Thingbaijam, 2011) and development of smoothed-gridded and areal seismicity models (Thingbaijam and Nath, 2011). Although

there are inevitably some limitations, as we shall see later, this work represents the current state-of-the-art as far as PSHA in India.

(In the discussion of possible improvements to Nath and Thingbaijam (2012) it will be noted that while earlier investigations (Bhatia et al., 1999; Das et al., 2006; Yadav et al., 2008; Jaiswal and Sinha, 2007) relied on areal seismogenic source zonation, Nath and Thingbaijam (2012) adds smoothed-gridded point sources while Ashish et al. (2016) adds fault-modelling. Different types of seismogenic zones address different types of hazard and return periods, and can be effectively combined using logic trees to better encapsulate epistemic uncertainty.)

1.2 Open science and OpenQuake

The seismological research community is a collegial one and, as they must, researchers generally share data, models, software and results freely. However it is becoming a generally recognized problem that scientific computation, even as significant technological advances are made, is falling far short of expectations in terms of reproducibility.

Reproducibility is one of the fundamental tenets of science. In other disciplines, the components of a properly-documented experiment are well-known and widely practiced. Scientific computing is a relative newcomer, and presents new challenges, such as the constant evolution of programming languages.

Although Nath and Thingbaijam (2012) provides much of the information necessary to reproduce their results as an electronic supplement, it is incomplete, and worse, the software used to run the simulations is not freely available. The consequence is that results cannot be verified, errors cannot be corrected and improvements cannot be made.

The OpenQuake (Crowley et al., 2015) is a fully-featured suite of software for the modelling of seismic hazard and risk. The source code is open, freely distributable and modifiable. It is an ideal platform for development of PSHA model. In fact, there is an ongoing effort to build a Global Earthquake Model based on OpenQuake.

Essential components of reproducibility: model data (completely documented), result data (completely documented), freely available software (version numbers, repository commits), hardware capable of running the software.

1.3 Overview

In Section 2 the process followed to translate the model of Nath and Thingbaijam (2012) for OpenQuake is detailed. Specific issues around tectonic region assignments are addressed in Section 2.1.1. Difficulties encountered in interpreting and implementing the smoothed-gridded seismicity models are detailed in Section 2.1.2 while recommendations for an improved smoothed-gridded model are made in Section A.2.

In Section 2.2 issues encountered in implementing ground motion prediction equations are discussed.

Section 2.3 discusses the development of the logic trees. Some inconsistencies were encountered in the application of previous research regarding the usage of GMPEs in India. This is discussed briefly in Section 2.3.1 and recommendations for future work are made in Section A.1. The modelling of source frequency magnitude distribution uncertainty described in Nath and Thingbaijam (2012) turned out to be unimplementable in the strictest sense in OpenQuake and possibly on any platform, so compromises made are described in Section 2.3.2.

Section 3.1 verifies the current results against those of Nath and Thingbaijam (2012). In particular hazard curves and maps and tables of ground motion with various probabilities of

Table 1: Summary of layer characteristics used for source models. Completeness magnitudes and years used in generating original smoothed-gridded seismicity models are from Table 1 of Thingbaijam and Nath (2011). Layer identifiers used throughout this report are indicated. Tops and bottoms of layers have been taken as seismogenic depth limits. Hypocentral depths listed are at mid-layer.

				4		4.5		5.5	
minimum magnitude									
layer	depth (km)			start	end	start	end	start	end
	min.	max.	hypo.						
1	0	25	12.5	1994	2008	1964	2008	1903	2008
2	25	70	47.5	1990	2008	1964	2008	1902	2008
3	70	180	125	1996	2008	1964	2008	1914	2008
4	180	300	240	1970	2008	1984	2008	1912	2008

exceedence are presented and evaluated. Inconsistencies between the figures and electronic supplement of Nath and Thingbaijam (2012) are discussed. Section 3.2 investigates the importance of the use of region-specific ground motion prediction via their impact on hazard levels.

Section 3.3 briefly explores the possibilities for future work while reserving more detailed discussion reserved for the appendices.

2   Implementation

2.1   Seismogenic sources

The electronic supplement of Nath and Thingbaijam (2012) provides most but not all of the information required to generate a complete source model, even supplemented with the helpful additional information found in Thingbaijam and Nath (2011). This section thus focuses on bridging the gaps to construct a complete source model.

As shown in Figure 5, Nath and Thingbaijam (2012) actually proposed three source models, a single set of areal seismogenic source zones, and two smoothed-gridded point source models. Doing so retains benefits from both types of models without neglecting the related epistemic uncertainty model. Both models are derived from the catalogue of Nath et al. (2010) for sub-catalogues with different minimum magnitudes and depth ranges.

Thingbaijam and Nath (2011) divide their model into four layers as summarized in Table 1 and Figure 7. Crustal thicknesses vary significantly across the region of study, but the convenience of constant model layer thicknesses turns out to be not entirely unrealistic. The continental crust is 75-80 km thick beneath the Himalayas where the tectonics can be divided into shallow crust and interface (Thingbaijam and Nath, 2011). Similarly in the the Shillong plateau of Northeast India the crust is quite thick and significant variation of stress drop with depth has been noted, with devastating "pop-up" type events Bilham and England (2001) being generated in the lower crust (Nath et al., 2012). In stable continental regions the crustal thickness is a more usual 35-45 km, with seismicity concentrated in the uppermost 25 km. The preceding seismotectonic features can be represented reasonably well using two seismogenic layers: 0-25 km and 25-70 km.

Intra-slab subduction occurs in three broad zones: the Hindu-Kush and Pamir ranges in the north-west, the Indo-Myanmar subduction zone in the north-east and the Sumatra-Andaman subduction zone in the south east. Deep-seated seismicity only occurs in the first and last region. The tectonics of the Indo-Myanmar region are a combination of oblique subduction,

accretion and collision Wang et al. (2014). These tectonic zones are represented by two deeper seismogenic layers: 70-150 km and 150-300 km.

This stack of depth-limited seismogenic zones can crudely represent the fact that subduction events are generally spread over a dipping plane (see Figure 8). The four-layer structure furthermore captures the fact that there are 4 clear modes in the distribution of depths (see Figure 7).

2.1.1 Areal zones

Areal source models are appropriate when source mechanisms and seismicity rates are relatively constant across the area, and can provide a sound basis for regional variation of b-value, maximum magnitude and other key parameters of a frequency-magnitude distribution.

As we will see in Section 2.3.1 ground motion prediction logic trees depend on correct assignment of tectonic region types. The main difficulty in implementing the areal source model of Nath and Thingbaijam (2012) was that although their intentions were generally clear, these region assignments were not made explicit. Tectonic region type assignments were therefore made using a combination of the representative focal mechanisms reported by Nath and Thingbaijam (2012) and fault maps such as the HimaTibetMap database (Styron et al., 2010). Since the representative focal mechanism was computed as the average of the moment tensors reported in the GCMT database weighted by magnitude it is biased in favour of the larger earthquakes (Thingbaijam and Nath, 2011). The inferred tectonic region assignments assumed are shown in Figure 1.

Potentially problematic tectonic region type assignments include:

- zone 906 in the Great Himalayas just north of the Shillong plateau was assigned "active shallow crust strike-slip reverse" even though the main trace of the Himalayan subduction fault runs through it, because the representative focal mechanism is strike-slip.

- zones 71, 86 on layer 1 and zones 9031, 9081, 9131, 9151 and 9171 on layer 2 have have a values of zero and so were not included in the areal source model

Magnitude-scaling relations are used in PSHA to determine the actual rupture dimensions once a magnitude has been drawn from a frequency-magnitude distribution. These were relatively straightforward to select once the tectonic region assignments were made, since "Wells and Coppersmith (1994) for crustal events and those given by Strasser et al. (2010) for the subduction earthquakes" (Nath and Thingbaijam, 2012, p. 140). For interface and intraslab regions `StrasserInterface` and `StrasserIntraslab` were used, respectively. The comment that "the fault-rupture area estimated from the magnitude is constrained by a factor of 2" (Nath and Thingbaijam, 2012, p. 140) was similarly interpreted as a width/depth aspect ratio of 2.

Since it is not explicitly stated in Nath and Thingbaijam (2012) the seismogenic depth was assumed to be midway between the minimum and maximum for each layer. Potential refinements to this setup are discussed in Section A.3.

The supplementary information required to generate the fully specified areal source model from the electronic supplement files `polygonlay%d.txt` and `seismicitylay%d.txt` files in the is contained in `auxiliary data.csv`.

2.1.2 Smoothed-gridded

Smoothed-gridded seismicity models aim to replicate geographic variations of activity rates in a catalogue-driven way. Typically a smoothing kernel is used which enforces a correlation distance and limits the resolution.
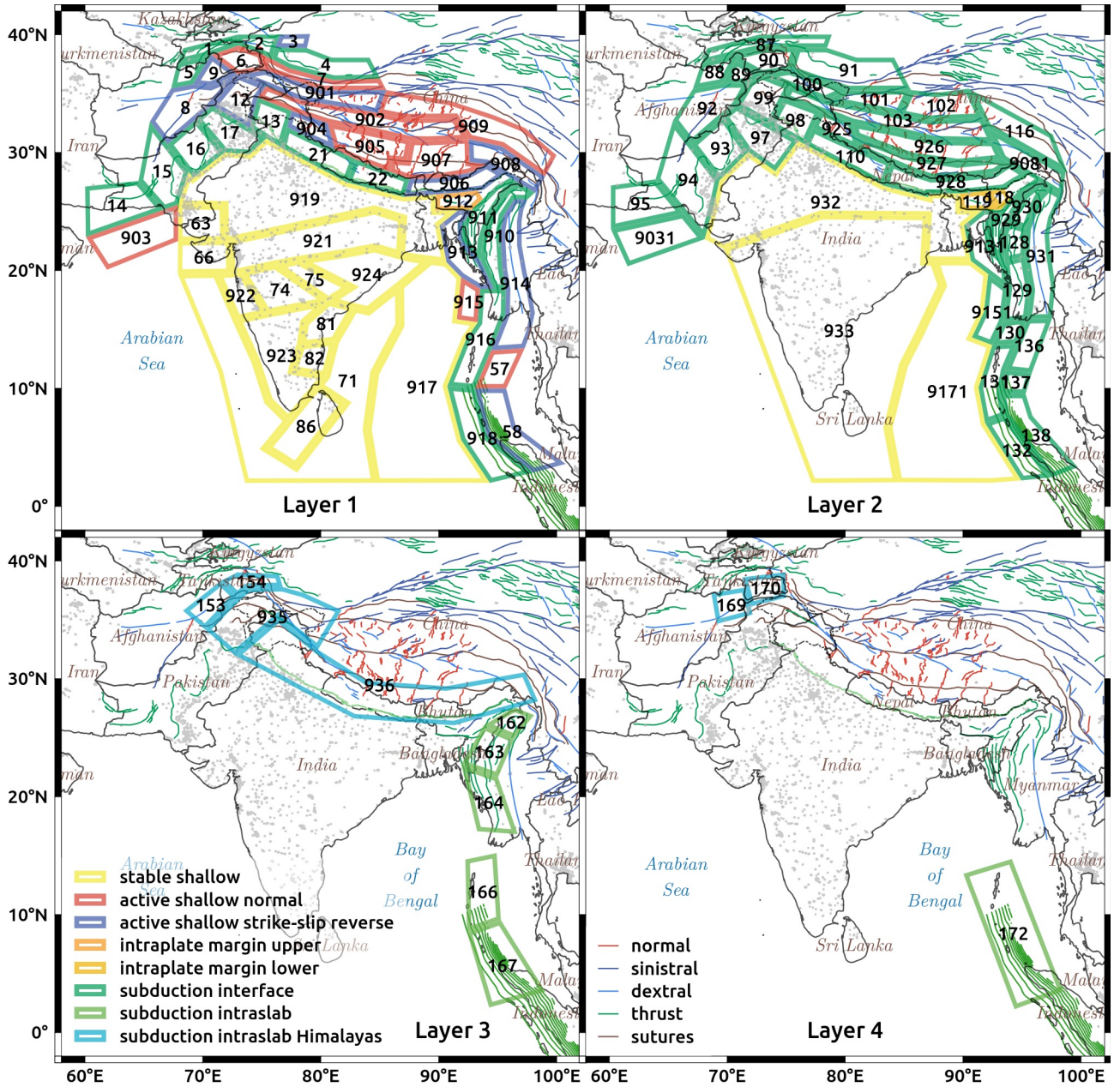
Figure 1: Areal source model tectonic region assignments used in GMPE logic tree. The areal source models is encoded in `areal_source_model.xml`. Zone identification numbers from Nath and Thingbaijam (2012) are indicated. Fault traces are from HimaTibetMap-1.0 (Styron et al., 2010) except the Sumatran subduction fault which is from SLAB 1.0 (Hayes et al., 2012). Fault data from the stable regions of India is lacking. Urban areas, "contiguous patches of built-up land greater than 1 km²" (Schneider et al., 2009), are indicated in darker grey.

After some discussion with the second author it was decided that although the models are described as "spatially varying annual activity rates" (Nath and Thingbaijam, 2012, p. 140) the electronic supplement actually contains spatially smoothed total seismicity, i.e. number of events (per cell). In order to convert this information to activity rates, i.e. number of events per year (per cell), it was necessary to obtain the duration of each sub-catalogue. Fortunately the missing ingredient is summarized in (Thingbaijam and Nath, 2011, Table 1) and reproduced in Table 1 below.

In OpenQuake each point in the smoothed seismicity model is treated as a point source with a specified frequency-magnitude distribution: at a minimum a, b and M_{max} must be spedified. Nath and Thingbaijam (2012) indicate that "b-value and m max remain fixed within the source zone". Thus for the smoothed seismicity model the parameters b and M_{max} of the truncated Gutenberg-Richter magnitude-frequency distributions are to be inferred from the areal source model zonation. For points inside zones with non-zero a values in the areal source model this is trivial; for points outside these zones the zone with the shortest perpendicular distance was chosen.

A point source model in OpenQuake also requires definition of the uncertainty of the probability distribution, as well as the tectonic region type and source mechanism for the selection and implementation of GMPEs, respectively. Thus the same procedure was used to assign σ_b, $\sigma_{m_{max}}$, tectonic subregion, rake, dip, strike and magnitude scaling relations were used.

The electronic supplement to Nath and Thingbaijam (2012) for the smoothed-gridded seismicity models simply gives values for `nu4_5` and `nu5_5` for each latitude and longitude, where "ν_i, is the annual activity rate for ith seismogenic source for a threshold magnitude" (Nath and Thingbaijam, 2012, p. 140).

The truncated Gutenberg-Richter magnitude-frequency distribution in OpenQuake implements
$$\lambda(M \geq m) = 10^{a-bm} = e^{\alpha - \beta m}$$

where, since λ is an annual rate, 10^a is too. If we ignore events below some threshold m_{min} then the annual rate becomes

$$\lambda(M \geq m_{min}) = e^{\alpha - \beta m_{min}} e^{-\beta(m-m_{min})} = \nu e^{-\beta(m-m_{min})}$$

Thus to compute the a value required by OpenQuake from the activity rate ν for a given magnitude threshold, we must also take into account the b value for the zone:

$$a = \log_{10}(\nu) + b m_{min}$$

Applying this formula gives rates for the areal model which agree well with the catalogue, but the rates in the smoothed-gridded seismicity model are on the order of 100x higher. Figure 3.6 of Thingbaijam (2011) shows rates in line with those of Thingbaijam and Nath (2011) and which better agree with the areal rates. In discussion with the author was concluded that the electronic supplement to Nath and Thingbaijam (2012) is in error.

Other assumptions:

- zones 9031, 9081, 9131, 9151 and 9171 on layer 2 have m_{max} values values of zero so the the smoothed seismicity points in or nearest to these zones on layer 2 were assigned the m_{max} values from the corresponding zones on layer 1, namely zones 903, 908, 913, 915 and 917.

Note differences in grids: although the hazard maps in the electronic supplement are at 0.2° and the paper says the smoothed-gridded models are also at 0.2° they are in fact at 0.1°. Figure shows the model at just 0.2°.

Table 2: Comparison of seismicity rates in areal and smoothed-gridded seismicity models to those obtained from the catalogue.

areal 4.5	areal 5.5	smoothed 4.5	smoothed 5.5	ratio 4.5	ratio 5.5	cat 4.5	cat 5.5
79.5	8.4	57.4	3.3	0.72	0.39	54	3
67.5	10.4	85.1	3.9	1.26	0.38	78	3
35.7	2.9	42.9	1.8	1.20	0.62	39	1
11.5	1.6	23.9	1.3	2.08	0.81	10	1
194.2	23.3	209.3	10.3	1.08	0.44	181	8

Note smoothing details. A Gaussian kernel is used, following the methodology of Frankel (1995) with correlation distances of 65 and 85 km for m_{min} of 4.5 and 5.5 respectively. The method involves smoothing event counts, thus events per year, for a given minimum magnitude.

Thingbaijam (2011)

Nath et al. (2010)

2.2 Ground-motion prediction

Importance of models implemented for this study.

The great Assam earthquake of 1897 destroyed buildings within several hundred km. The two main fault structures involved are capable of M > 8 plateau-building events with a recurrence interval of 3-8 kyr each (Bilham and England, 2001).

Nath et al. (2012) notes stress drop apparently increasing with depth and models κ using a database of recent well-recorded micro-earthquakes, and uses this information to develop stochastic models for events in the upper and lower crust. The simulations are of vertical motion at a hard-rock site and no site corrections are attempted.

Sharma et al. (2009) points out that the decay rate of PGA for shallow India-Bangladesh and deep India-Burma border events have different distance scaling. The former leads to the necessity of a GMPE specific to the Shillong plateau Nath et al. (2012) while the latter means interface subduction events need to be treated differently.

Issues encountered while implementing GMPE logic tree:

- layer 4 depth range of 180-300 km is significantly deeper than deepest events used in regression for ATBO03 (100 km), LILE08 (161 km), ZHAO06 (120 km) and GUPT10 (148 km) are specified for. YCSH97 only included events to 229 km. KANN06 is specified to 200 km depth, but is only used for interface events (layer 2).

- Assumed "and Andaman-Sumatra subduction" missing from Figure 3.

- Why is Youngs (1997) not used in the subduction interfaces?

- Should the Japan/Cascadia distinction not also be used for interface subduction with Atkinson & Boore (2003)?

- Nath and Thingbaijam (2012) doesn't seem to me to follow the recommendations of Nath and Thingbaijam (2011) as far as having two subduction intra-slab sub-regions: the former uses Indo-Myanmar and Himalayas while the latter recommends Indo-Myanmar and Hindukush. Nath and Thingbaijam (2012) is followed strictly for for phase 1.
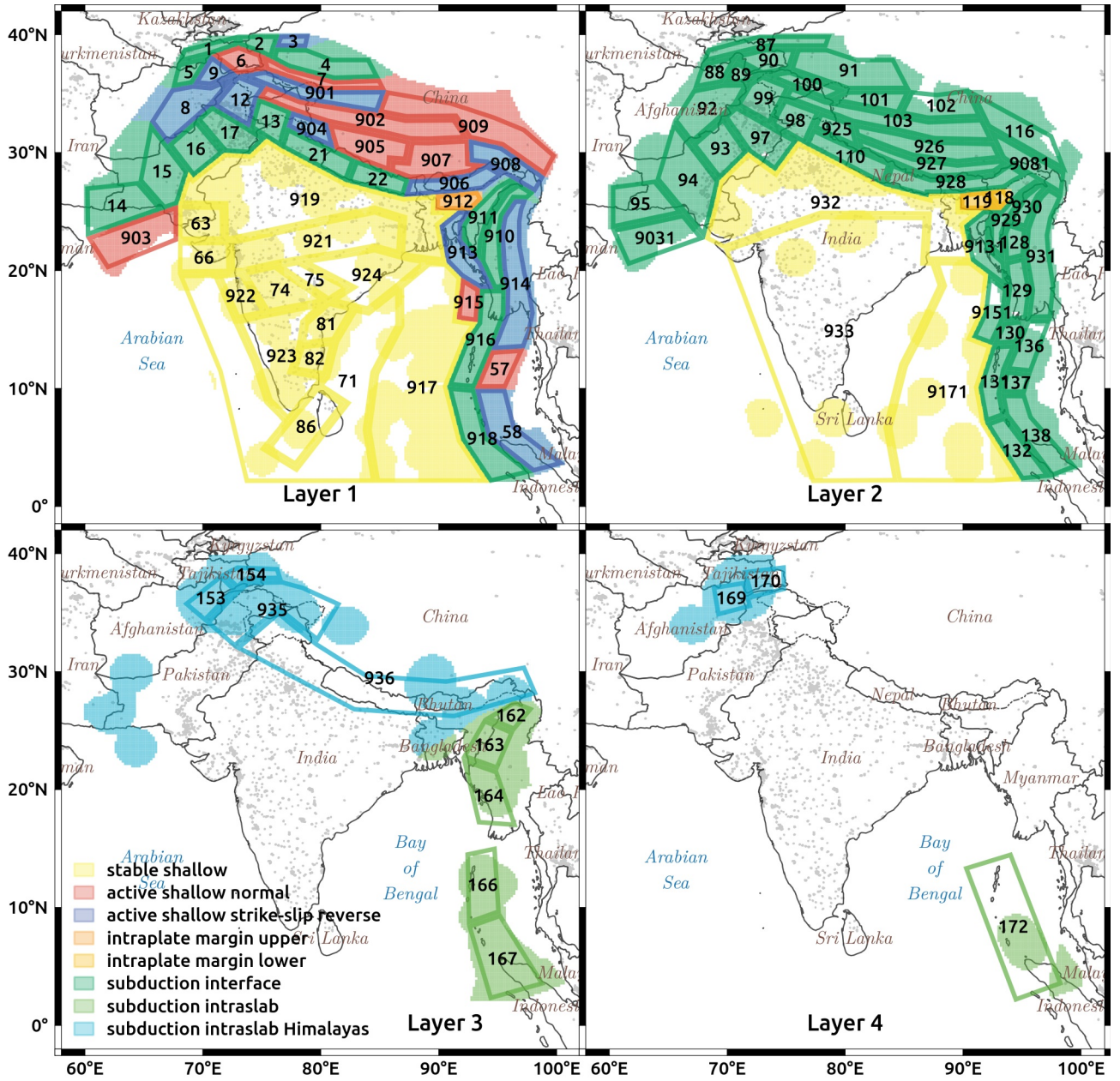
Figure 2: Tectonic region assignments and activity rates for smoothed seismicity point source model. The smoothed seismicity source models are encoded in `smoothed_source_model_mmin4.5.xml` and `smoothed_source_model_mmin5.5.xml`

- Assignment of source mechanism (normal or not, matters in shallowest layer only) is tricky. Dip cannot be used to distinguish normal and reverse subduction because the subduction interface angle is not known. different GMPEs use different rake thresholds; a threshold of 30° was chosen, consistent with Boore and Atkinson (2008); Campbell and Bozorgnia

9

Table 3: Ground motion prediction equations. Models newly implemented in OpenQuake as part of the current work are indicated.

Code	New	Reference
AKBO10		Akkar and Bommer (2010)
BOAT08		Boore and Atkinson (2008)
CABO08		Campbell and Bozorgnia (2008)
ZHAO06		Zhao et al. (2006)
ATBO03		Atkinson and Boore (2003)
ATMA09		Atkinson and Macias (2009)
LILE08		Lin and Lee (2008)
YCSH97		Youngs et al. (1997)
TORO02		Toro (2002)
ATBO06		Atkinson and Boore (2006)
CAMP03		Campbell (2003)
SDBK09	✓	Sharma et al. (2009)
RAIY07	✓	Raghukanth and Iyengar (2007)
NTMN12	✓	Nath et al. (2012)
GUPT10	✓	Gupta (2010)
KNMF06	✓	Kanno et al. (2006)

(2008) but not Zhao et al. (2006).

Issued encountered while implementing GMPEs:
Sharma et al. (2009)

- lacks a M^2 term Cotton et al. (2006)

- does not define rock vs. soil

Raghukanth and Iyengar (2007)

- typographical errors in coefficient tables: grossest error fixed, 3 other errors causing approximately 10% error not fixed

- actually defines 4 different models: must assume that for all of peninsular India was used by Nath and Thingbaijam (2012), not one of those for sub-regions.

Kanno et al. (2006) -> Douglas (2003)

2.3   Logic Trees

2.3.1   Ground-Motion Prediction

The GMPE logic tree implemented in Nath and Thingbaijam (2012) is shown in Figure 4. Since some of these GMPEs are new to OpenQuake (see Table 3) a comparison was done between that model and that obtained with only the standard GMPEs. For this purpose a "simplified" GMPE logic tree was constructed which simply omitted the newly-implemented GMPEs and retained equal weighting for the rest.

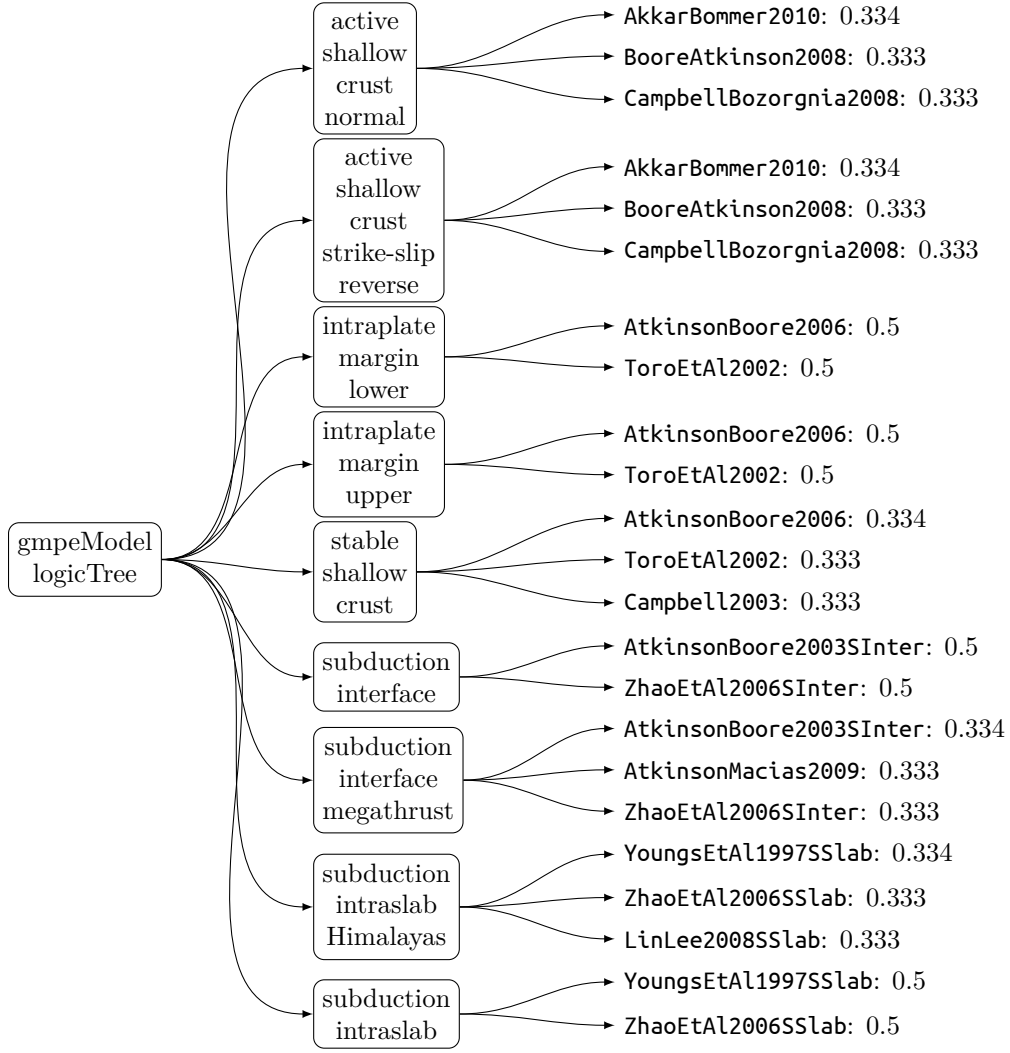Moving forward an obvious modification is to replace superseded NGA models with their more up-to-date versions.

Figure 3: Simplified GMPE logic tree employing only established OpenQuake models, as encoded in `gmpe_logic_tree_omit_new.xml`. Middle column selects tectonic region types as defined in Figure 1. OpenQuake model class names and assigned weights are given on the right side. New and established models are enumerated in Table 3

[It may also be desirable to rationalize the weighting of the GMPEs or include entirely new ones, such as the new BC Hydro subduction model (Abrahamson et al., 2012). The discussion continues below and is incomplete.]

Anbazhagan et al. (2015) seem to be proposing different weights for different regions based on single events in those regions. An extreme example is to define different weights for Anjar, 1956 and Bhuj , 2001 earthquakes even though the epicentres and depths were very close together. In contrast Nath and Thingbaijam (2011) compute LLH for 7 regions (using 38 events total) and state that, "individual events do not have significant number of observations to support a viable ranking basis."
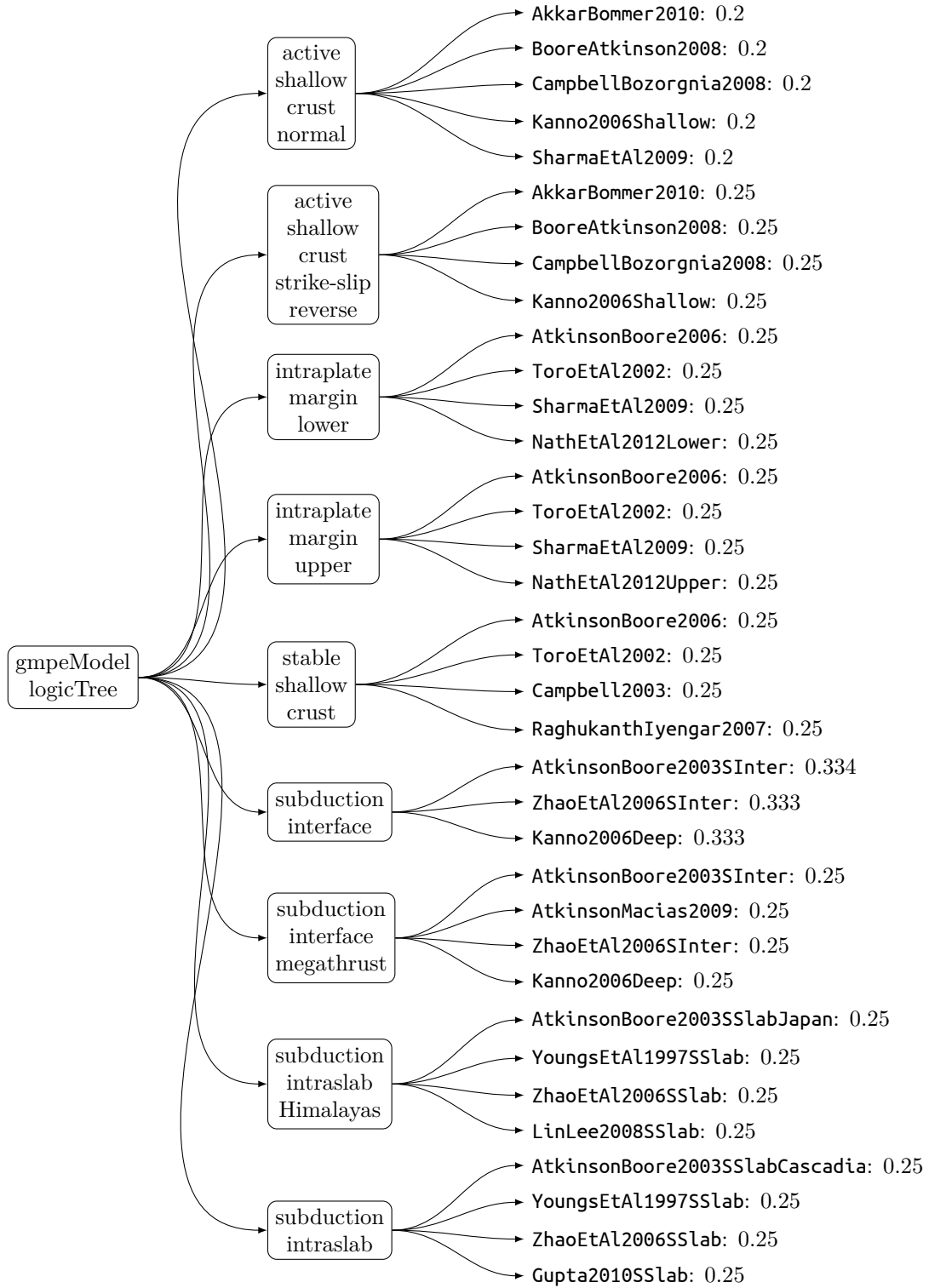
Figure 4: GMPE logic tree of Nath and Thingbaijam (2012), as encoded in `gmpe_logic_tree.xml`.
See Figure 3 for complete description

Anbazhagan et al. (2015) seem to misuse the concept of data support index (DSI) (Delavaud et al., 2012) by setting weights to zero when the DSI is negative. The threshold is arbitrary and is chosen without discussion. As Delavaud et al. (2012) point out "more important than the sign of the DSI is the difference of DSI between two models."

Both Anbazhagan et al. (2015) and Nath and Thingbaijam (2011) rely on estimating ground motions from macroseismic intensity. I'm sure it is a matter of low seismicity and lack of instrumentation, but I'm still surprised. I would expect the catalogue for peninsular India to be complete for 20 years to magnitude 5 so that one could thus get 10 well-recorded events, at least. There is significant additional (aleatory and epistemic) variability in mapping EMS to PGA which must obscure the true performance of the GMPEs. Perhaps this is part of why Anbazhagan et al. (2015) and Nath and Thingbaijam (2011) arrive at such different LLH scores and rankings for the same events (Anbazhagan et al., 2015, Table 5). It would be interesting to compare the results of LLHs computed using EMS inferred from digitized intensity maps to those computed using instrumental PGA for at least a few events since 1990. Nath and Thingbaijam (2011) take a step towards this by looking at the scatter in their mapping of PGA to EMS but it's not quite the same.

Many authors (Scherbaum et al., 2009; Nath and Thingbaijam, 2011; Delavaud et al., 2012; Anbazhagan et al., 2015) seem unduly interested in "ranking", i.e. constructing an ordered list of GMPEs. This is not a horse race. Scherbaum et al. (2009) suggests a way to turn an LLH score into a logic-tree weight and the formula does not require ranking. Furthermore, in constructing a logic tree one must include factors outside the performance-based scoring, for example an assessment of whether the set is "mutually exclusive and collectively exhaustive" (Bommer and Scherbaum, 2008). For me the question of ranking is just "noise" which obscures more important questions.

The mutual exclusivity requirement means, to me, that models should be omitted which are redundant in the sense of being too similar to other models in terms of the methodology of their construction, especially if that means they make similar predictions and have similar limitations as a result. For example the exclusion of models which have been superseded (Cotton et al., 2006) can be seen as an application of the requirement that models be mutually exclusive. Another example would be, for a GMPE logic tree intended for the Indian subcontinent, to omit a model such as Hwang and Huo (1997) in favour of Atkinson and Boore (2006) since both are based on stochastic simulation in Eastern North America.

The collective exhaustiveness requirement means, is trickier. It is this requirement which pushes hazard modellers to seek out and evaluate more and complementary types of models. Thus models with broad data support from other regions complement models with poor data support from the target region. Stochastic models supplement data-driven models. Models with different functional forms, distance or magnitude ranges can complement each other.

The process of developing a logic tree to assess epistemic uncertainty is thus a dialectical one. Mutual exclusivity and collective exhaustiveness comprise opposing forces which must be exerted alternately and in tandem.

[Now apply these principles to move forward from Nath and Thingbaijam (2012)!]

2.3.2 Source Models

The source model logic tree is shown in symbolic form in Figure 5.

Nath and Thingbaijam (2012) accounts for the epistemic uncertainty in seismicity model parameters by estimating the standard deviations of b and m_{max} in each source zone and assigning weights to ± 1 standard deviation for each source. This results in a source model logic tree too large to represent on a page; just a portion of it is shown in Figure 6.
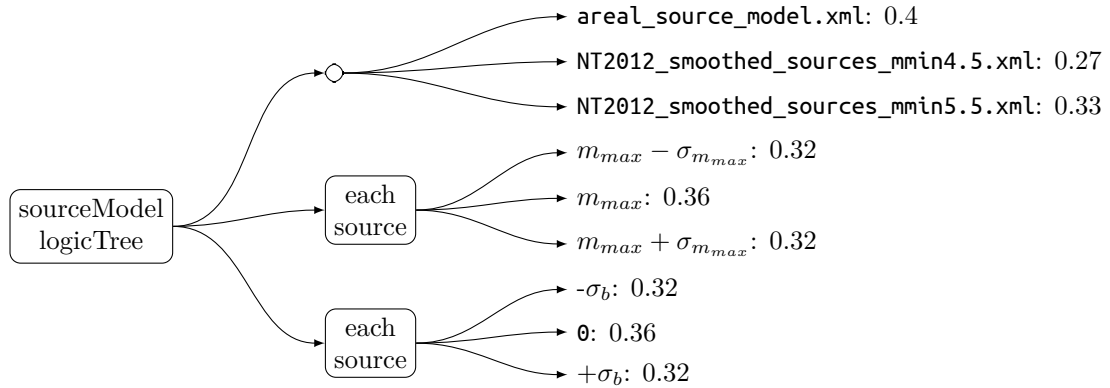
13

Figure 5: Symbolic source model logic tree of Nath and Thingbaijam (2012).

Note that although Figure 4 of Nath and Thingbaijam (2012) shows the activity rate ν (and by implication a) varying with b, no estimates of the standard deviation of a or nu. The in OpenQuake happens to recalculate a as b After modifying b using the uncertainty type bGRRelative the a value is automatically recalculated to maintain constant total moment rate. It has been assumed that this is the behaviour which Nath and Thingbaijam (2012) implemented.

The fact that Figure 6 has to be truncated is not simply a lack of page space. Despite the common rendering of them in parallel, full enumeration actually takes place in series, so rather than just 3×223 branches there are in fact $3^{223} \approx 10^1 06$, or just over a googol of branches. Not only was full-enumeration out of the question, even partial enumeration was problematic on the current version of OpenQuake code.

It is unlikely that Nath and Thingbaijam (2012) performed full-enumeration. Some practitioners

3 Hazard results

3.1 Verification

3.2 Sensitivity
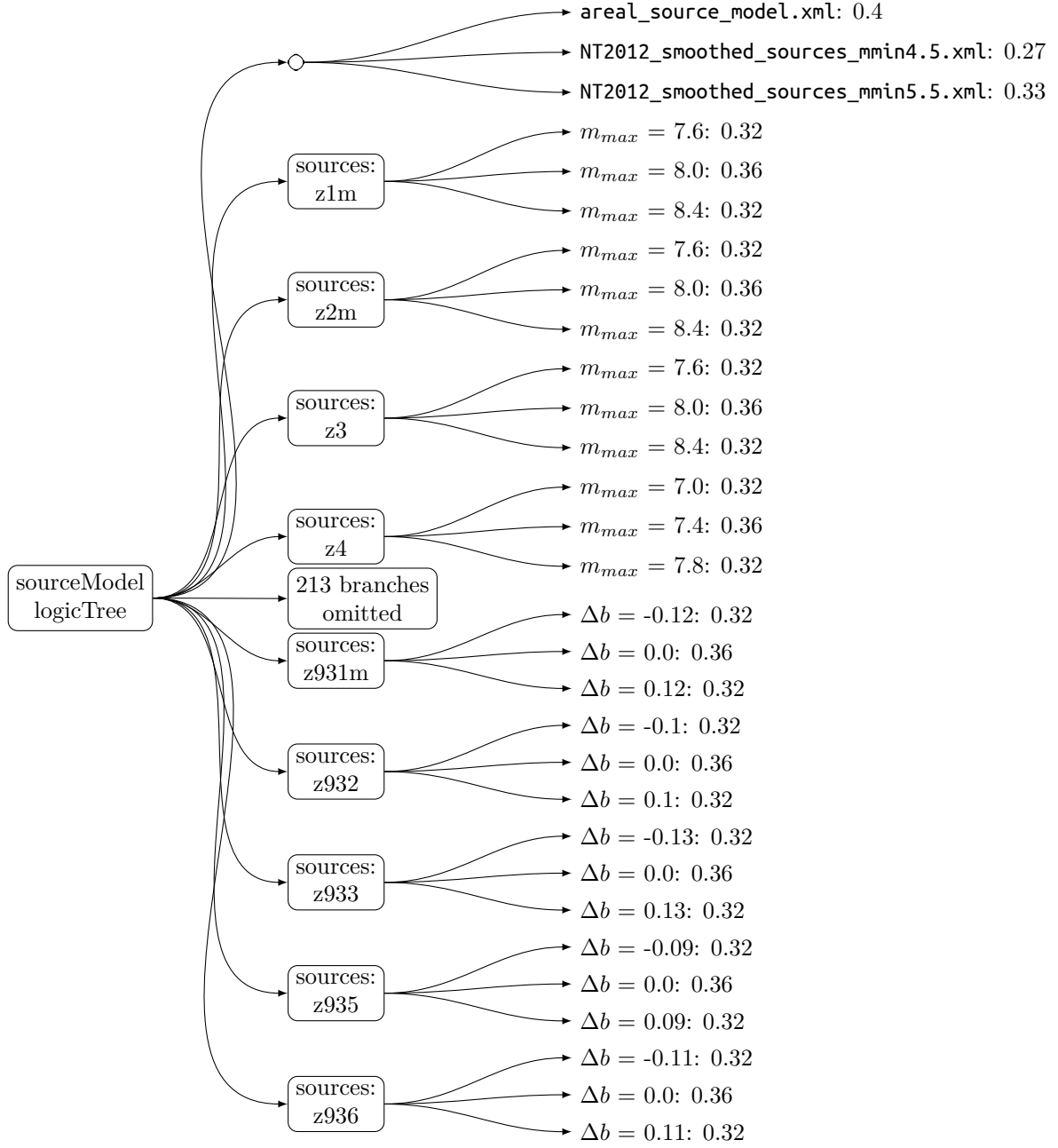
3.3 Discussion

4 Conclusions

Acknowledgement

Figure 6: Partial source model logic tree of Nath and Thingbaijam (2012). The full model is encoded in source_model_logic_tree.xml

References

Abrahamson, N., Gregor, N., and Addo, K. (2012). Bc hydro ground motion prediction equations for subduction earthquakes. *Earthquake Spectra*. (Cited on page 11.)

Akkar, S. and Bommer, J. J. (2010). Empirical equations for the prediction of PGA, PGV, and spectral accelerations in Europe, the Mediterranean region, and the Middle East. *Seismological Research Letters*, 81(2):195–206. (Cited on page 10.)

Anbazhagan, P., Sreenivas, M., Ketan, B., Moustafa, S. S., and Nassir, S. A.-A. (2015). Selection of ground motion prediction equations for seismic hazard analysis of peninsular India. *Journal of Earthquake Engineering*, (just-accepted). (Cited on pages 11 and 13.)

Ashish, Lindholm, C., Parvez, I. A., and Kühn, D. (2016). Probabilistic earthquake hazard assessment for peninsular India. *Journal of Seismology*, pages 1–25. (Cited on pages 2 and 3.)

Atkinson, G. M. and Boore, D. M. (2003). Empirical ground-motion relations for subduction-zone earthquakes and their application to Cascadia and other regions. *Bulletin of the Seismological Society of America*, 93(4):1703–1729. (Cited on page 10.)

Atkinson, G. M. and Boore, D. M. (2006). Earthquake ground-motion prediction equations for eastern North America. *Bulletin of the seismological society of America*, 96(6):2181–2205. (Cited on pages 10 and 13.)

Atkinson, G. M. and Macias, M. (2009). Predicted ground motions for great interface earthquakes in the Cascadia subduction zone. *Bulletin of the Seismological Society of America*, 99(3):1552–1578. (Cited on page 10.)

Bhatia, S. C., Kumar, M. R., and Gupta, H. K. (1999). A probabilistic seismic hazard map of india and adjoining regions. *Annals of Geophysics*, 42(6). (Cited on pages 2 and 3.)

Bilham, R. and England, P. (2001). Plateau 'pop-up'in the great 1897 Assam earthquake. *Nature*, 410(6830):806–809. (Cited on pages 4 and 8.)

Bilham, R., Gaur, V. K., and Molnar, P. (2001). Himalayan seismic hazard. *Science*, 293(5534):1442–4. (Cited on pages 2 and 20.)

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. (Cited on page 13.)

Boore, D. M. and Atkinson, G. M. (2008). Ground-motion prediction equations for the average horizontal component of PGA, PGV, and 5%-damped PSA at spectral periods between 0.01 s and 10.0 s. *Earthquake Spectra*, 24(1):99–138. (Cited on pages 9 and 10.)

Bureau of Indian Standards (2002). Criteria for earthquake resistant design of structures, part 1 - general provisions and buildings. Number 1893–2002 in IS. New Delhi. (Cited on page 2.)

Campbell, K. W. (2003). Prediction of strong ground motion using the hybrid empirical method and its use in the development of ground-motion (attenuation) relations in eastern North America. *Bulletin of the Seismological Society of America*, 93(3):1012–1033. (Cited on page 10.)

Campbell, K. W. and Bozorgnia, Y. (2008). Nga ground motion model for the geometric mean horizontal component of PGA, PGV, PGD and 5% damped linear elastic response spectra for periods ranging from 0.01 to 10 s. *Earthquake Spectra*, 24(1):139–171. (Cited on pages 9 and 10.)

Cotton, F., Scherbaum, F., Bommer, J. J., and Bungum, H. (2006). Criteria for selecting and adjusting ground-motion models for specific target regions: Application to central Europe and rock sites. *Journal of Seismology*, 10(2):137–156. (Cited on pages 10 and 13.)

Crowley, H., Monelli, D., Pagani, M., Silva, V., Weatherill, G., and Rao, A. (2015). *OpenQuake Engine User Instruction Manual v1.5*. Global Earthquake Model (GEM). `http://www.globalquakemodel.org/media/cms_page_media/432/oq-manual-15.pdf` Accessed 2015-10-15. (Cited on pages 2 and 3.)

Das, S., Gupta, I. D., and Gupta, V. K. (2006). A Probabilistic Seismic Hazard Analysis of Northeast India. *Earthquake Spectra*, 22(1):1–27. (Cited on pages 2 and 3.)

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. (Cited on page 13.)

Douglas, J. (2003). Earthquake ground motion estimation using strong-motion records: a review of equations for the estimation of peak ground acceleration and response spectral ordinates. *Earth-Science Reviews*, 61(1):43–104. (Cited on page 10.)

Frankel, A. (1995). Mapping seismic hazard in the central and eastern united states. *Seismological Research Letters*, 66(4):8–21. (Cited on page 8.)

Gupta, I. (2010). Response spectral attenuation relations for in-slab earthquakes in Indo-Burmese subduction zone. *Soil Dynamics and Earthquake Engineering*, 30(5):368–377. (Cited on page 10.)

Hayes, G. P., Wald, D. J., and Johnson, R. L. (2012). Slab1.0: A three-dimensional model of global subduction zone geometries. *Journal of Geophysical Research: Solid Earth (1978–2012)*, 117(B1). (Cited on page 6.)

Hwang, H. and Huo, J.-R. (1997). Attenuation relations of ground motion for rock and soil sites in eastern United States. *Soil Dynamics and Earthquake Engineering*, 16(6):363–372. (Cited on page 13.)

Jaiswal, K. and Sinha, R. (2007). Probabilistic Seismic-Hazard Estimation for Peninsular India. *Bulletin of the Seismological Society of America*, 97(1B):318–330. (Cited on pages 2 and 3.)

K, B., W, R., and N, L. (2014). The Himalayan frontal thrust: attributes for seismic hazard, version 1.0. Technical report. `http://www.nexus.globalquakemodel.org/gem-faulted-earth/posts/the-himalayan-frontal-thrust-attributes-for-seismic-hazard` Accessed 2014-12-04. (Cited on page 19.)

Kanno, T., Narita, A., Morikawa, N., Fujiwara, H., and Fukushima, Y. (2006). A new attenuation relation for strong ground motion in Japan based on recorded data. *Bulletin of the Seismological Society of America*, 96(3):879–897. (Cited on page 10.)

Lin, P.-S. and Lee, C.-T. (2008). Ground-motion attenuation relationships for subduction-zone earthquakes in northeastern Taiwan. *Bulletin of the Seismological Society of America*, 98(1):220–240. (Cited on page 10.)

Mahajan, A. K., Thakur, V. C., Sharma, M. L., and Chauhan, M. (2009). Probabilistic seismic hazard map of NW Himalaya and its adjoining area, India. *Natural Hazards*, 53(3):443–457. (Cited on page 2.)

Nath, S. K. and Thingbaijam, K. K. S. (2011). Peak ground motion predictions in India: an appraisal for rock sites. *Journal of Seismology*, 15(2):295–315. (Cited on pages 2, 8, 11, and 13.)

Nath, S. K. and Thingbaijam, K. K. S. (2012). Probabilistic seismic hazard assessment of India. *Seismological Research Letters*, 83(1):135–149. (Cited on pages 2, 3, 4, 5, 6, 7, 8, 10, 12, 13, 14, and 15.)

Nath, S. K., Thingbaijam, K. K. S., and Ghosh, S. K. (2010). Earthquake catalogue of South Asia – a generic m_w scale framework. `http://www.earthqhaz.net/sacat/` Accessed 2015-12-01. v2. (Cited on pages 2, 4, 8, and 19.)

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. (Cited on pages 2, 4, 8, and 10.)

Raghukanth, S. and Iyengar, R. (2007). Estimation of seismic spectral acceleration in peninsular India. *Journal of Earth System Science*, 116(3):199–214. (Cited on page 10.)

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. (Cited on page 13.)

Schneider, A., Friedl, M. A., and Potere, D. (2009). A new map of global urban extent from MODIS satellite data. *Environmental Research Letters*, 4(4):044003. (Cited on page 6.)

Sharma, M. L., Douglas, J., Bungum, H., and Kotadia, J. (2009). Ground-motion prediction equations based on data from the Himalayan and Zagros regions. *Journal of Earthquake Engineering*, 13(8):1191–1210. (Cited on pages 8 and 10.)

Styron, R., Taylor, M., and Okoronkwo, K. (2010). Database of active structures from the Indo-Asian collision. *Eos, Transactions American Geophysical Union*, 91(20):181–182. (Cited on pages 5 and 6.)

Thingbaijam, K. (2011). *Synoptic modeling for probabilistic seismic hazard analysis of India*. PhD thesis, Indian Institute of Technology Kharagpur, India. (Cited on pages 7 and 8.)

Thingbaijam, K. K. S. and Nath, S. K. (2011). A seismogenic source framework for the Indian subcontinent. Unpublished manuscript. (Cited on pages 2, 4, 5, and 7.)

Toro, G. R. (2002). Modification of the Toro et al.(1997) attenuation equations for large magnitudes and short distances. *Risk Engineering, Boulder, Colorado*. (Cited on page 10.)

Wang, Y., Sieh, K., Tun, S. T., Lai, K.-Y., and Myint, T. (2014). Active tectonics and earthquake potential of the Myanmar region. *Journal of Geophysical Research: Solid Earth*, 119(4):3767–3822. (Cited on page 5.)

Weatherill, G. A. (2014). *OpenQuake Hazard Modeller's Toolkit - User Guide*. Global Earthquake Model (GEM). (Cited on page 2.)

Yadav, R. B. S., Tripathi, J. N., Rastogi, B. K., and Chopra, S. (2008). Probabilistic Assessment of Earthquake Hazard in Gujarat and Adjoining Region of India. *Pure and Applied Geophysics*, 165(9-10):1813–1833. (Cited on pages 2 and 3.)

Youngs, R., Chiou, S.-J., Silva, W., and Humphrey, J. (1997). Strong ground motion attenuation relationships for subduction zone earthquakes. *Seismological Research Letters*, 68(1):58–73. (Cited on page 10.)

Zhao, J. X., Zhang, J., Asano, A., Ohno, Y., Oouchi, T., Takahashi, T., Ogawa, H., Irikura, K., Thio, H. K., Somerville, P. G., et al. (2006). Attenuation relations of strong ground motion in Japan using site classification based on predominant period. *Bulletin of the Seismological Society of America*, 96(3):898–913. (Cited on page 10.)

Appendix A

A.1 Alternative GMPEs

A.2 Catalogue Evaluation
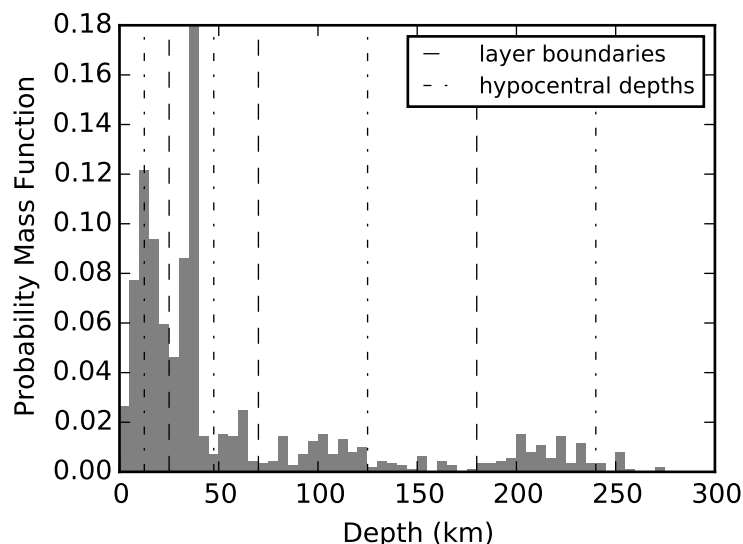
A.3 Potential Source Modelling Improvements

Figure 7: Depth histogram for mainshocks over magnitude 5.5. Mainshock identification is that of Nath et al. (2010). Seismogenic layer boundaries and hypocentral depths used in the current implementation are indicated as dashed and dash-dotted lines respectively.

- Base hypocentral depths on actual seismogenic depth distribution as shown in Figure 7. Placing the hypocentral depth at the modes of the overall catalogue would be a minor improvement. Better still would be to capture the mode or to construct an approximate distribution for each areal zone.

- Model the main Himalayan thrust as a simple fault. In particular K et al. (2014) breaks the fault into three segments and provides necessary details such as dip and depth limits. No variation of dip with depth is given, which is perhaps unrealistic, but at least the result
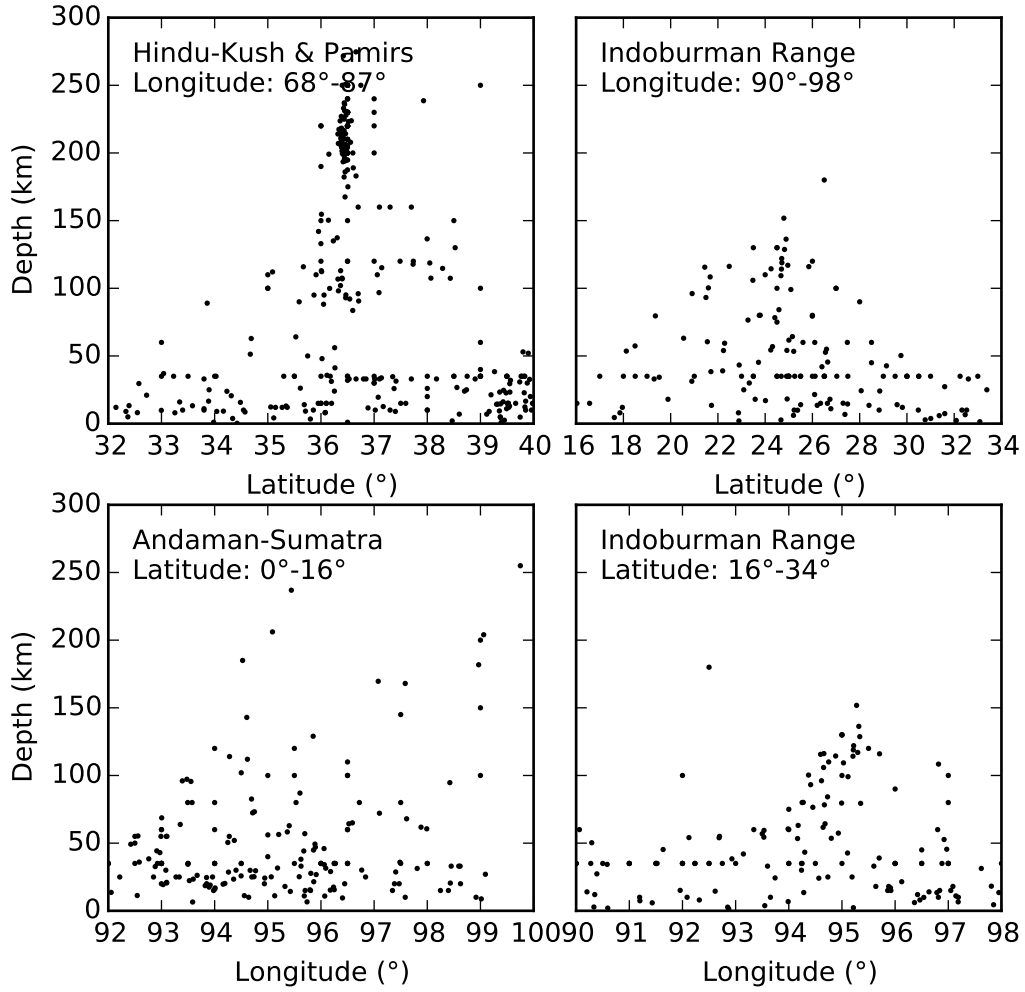
19

Figure 8: Depth vs. distance for mainshocks in regions with deep events. Subregions are indicated on each map; top left is the Hindu-Kush and Pamir ranges in the northwest of India viewed from the east, top right is the Andoman-Sumatran subduction zone viewed from the south while bottom left and right are beneath the Indoburman range and viewed from the east and south respectively. Sub-catalogues were selected for events over magnitude 5.5 within a rectangular box of latitude and longitude as indicated on each individual plot. Horizontal and vertical axes are plotted at different scales.

is simpler to model. Not all of the slip is being taken up on the frontal thrust; there are second and third folds which take up significant amount of slip, but the first is the most important from the standpoint of risk.

- Model the Oldham and Dauki faults under the Shillong plateau as simple faults (Bilham et al., 2001).

A.4 Summary of Electronic Data

This is an appendix because if you're reading this then you should already have the zip file with all of this data.

    auxiliary data.csv

Listing 1: phase1–job.ini

```ini
[general]

description = Open PSHA for India - Phase 1
calculation_mode = classical
random_seed = 42

[geometry]

sites_csv = NT2012_Table_3_lon_lat.csv

[logic_tree]

number_of_logic_tree_samples = 1000

[erf]

rupture_mesh_spacing = 5.0
width_of_mfd_bin = 0.1
area_source_discretization = 10.0

[site_params]

reference_vs30_type = measured
reference_vs30_value = 800.0
reference_depth_to_2pt5km_per_sec = 5.0
reference_depth_to_1pt0km_per_sec = 100.0

[calculation]

source_model_logic_tree_file = source_model_logic_tree.xml
gsim_logic_tree_file = gmpe_logic_tree_omit_new.xml
investigation_time = 1.0
intensity_measure_types_and_levels = {"PGA": [0.0047, 0.0068, 0.01, 0.015, 0.022, 0.033, 0.047, 0.068,
    0.1, 0.15, 0.22, 0.33, 0.47, 0.68, 1.0, 1.5, 2.2, 3.3, 4.7], "SA(0.2)": [0.0047, 0.0068, 0.01, 0.015,
    0.022, 0.033, 0.047, 0.068, 0.1, 0.15, 0.22, 0.33, 0.47, 0.68, 1.0, 1.5, 2.2, 3.3, 4.7], "SA(1.0)":
    [0.0047, 0.0068, 0.01, 0.015, 0.022, 0.033, 0.047, 0.068, 0.1, 0.15, 0.22, 0.33, 0.47, 0.68, 1.0,
    1.5, 2.2, 3.3, 4.7]}
truncation_level = 3
maximum_distance = 200.0

[output]

export_dir = ./export
mean_hazard_curves = true
hazard_maps = true
uniform_hazard_spectra = true
poes = 0.002105 0.00040397
individual_curves = false
```