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

March 11, 2016

Abstract

Open models enable 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 peninsular 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

The study of seismic hazard in India has been progressing steadily, from deterministic studies (Bureau of Indian Standards, 2002) to probabilistic seismic hazard assessment (PSHA) and from site-specific towards larger 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) summarize concerns with this standard (currently in force), including underestimation of hazard, application of single zone factor to regions with very different hazard, and lack of treatment of uncertainty.

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 ground-motion prediction equations (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 the Indian subcontinent.

(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: researchers generally share data, models, software and results freely. However it is becoming generally recognized that scientific computation is falling short of expectations in terms of reproducibility (Fomel and Claerbout, 2009; Donoho et al., 2009). As computing power grows, so do models and their complexity; it seems that our ability to describe these models is not keeping pace. In other disciplines, the components of a properly-documented experiment are well-known and widely practised. Scientific computing is a relative newcomer, and presents new challenges, such as the constant evolution of programming languages.

Reproducibility is one of the fundamental tenets of science. In the context of scientific computing reproducibility requires, at a minimum, a complete description of model, software versioning and results, open source code, and access to sufficient computing power (Hinsen, 2011).

Nath and Thingbaijam (2012) provides the majority of the model description and results as an electronic supplement. Unfortunately this description 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, and version-controlled at `https://github.com/gem/`. It is an ideal platform for development of PSHA models. In fact, there is an ongoing effort to build a Global Earthquake Model based on OpenQuake.

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 Appendix B.

In Section 2.2 issues encountered in implementing ground motion prediction equations (GMPEs) 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 Appendix A. The modelling of source frequency-magnitude distribution (FMD) 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 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 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 when supplemented by the earlier unpublished work of Thingbaijam and Nath (2011) which focuses specifically on source modelling. This section thus focuses on bridging the gaps to construct a complete source model.

Nath and Thingbaijam (2012) proposed three source models: a single set of areal seismogenic source zones, and two smoothed-gridded point source models. Combining these models using a logic tree (as discussed in Section 2.3.2 and diagrammed in Figure 5) allows the benefits of each model to be combined. All 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.

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.

minimum magnitude				4		4.5		5.5	
layer	depth (km) min.	max.	hypo.	start	end	start	end	start	end
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

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 uniform across a given area. They can provide a sound basis for regional assessment of b-value, maximum magnitude and other key parameters of a frequency-magnitude distribution, as shown in Thingbaijam and Nath (2011).

Selection of GMPEs (and thus the implementation of GMPE logic trees, see Section 2.3.1) depends on correct assignment of tectonic region types. The main difficulty in implementing the areal source model of Nath and Thingbaijam (2012) was that although the authors' intentions were generally clear, tectonic region assignments were not made explicit. Region assignments were made for this study 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 17 at the edge of the Pamir ranges is arguably "stable shallow crust" but was assigned "subduction interface".
- 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 903 and 915-918 are predominantly oceanic crust but have been assigned "active shallow crust" or "subduction interface" according to the dominant focal mechanism and fault types. For example zone 903 includes the Murray Ridge and so exhibits predominantly normal faulting as expected for a spreading ridge. It is classified for the purpose of GMPE selection as "active shallow crust normal", but is likely in reality to produce ground motions distinct from an active continental crust.
- 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 assigned a "no seismicity" type and omitted from the areal source model.
- Zones 169, 170 and 172 on layer 4 capture seismicity at 180-300 km depth, but only Youngs et al. (1997) and Kanno et al. (2006) support depths below 180 km (see Table 3).

Zones numbered as "9xx" in Nath and Thingbaijam (2012) represent the amalgamation of several zones from Thingbaijam and Nath (2011). In some cases this was done because of similarity of source mechanisms and statistics while in others it was necessary because the amount of seismicity in one of the zones was insufficient for FMD characterisation. Furthermore, zones

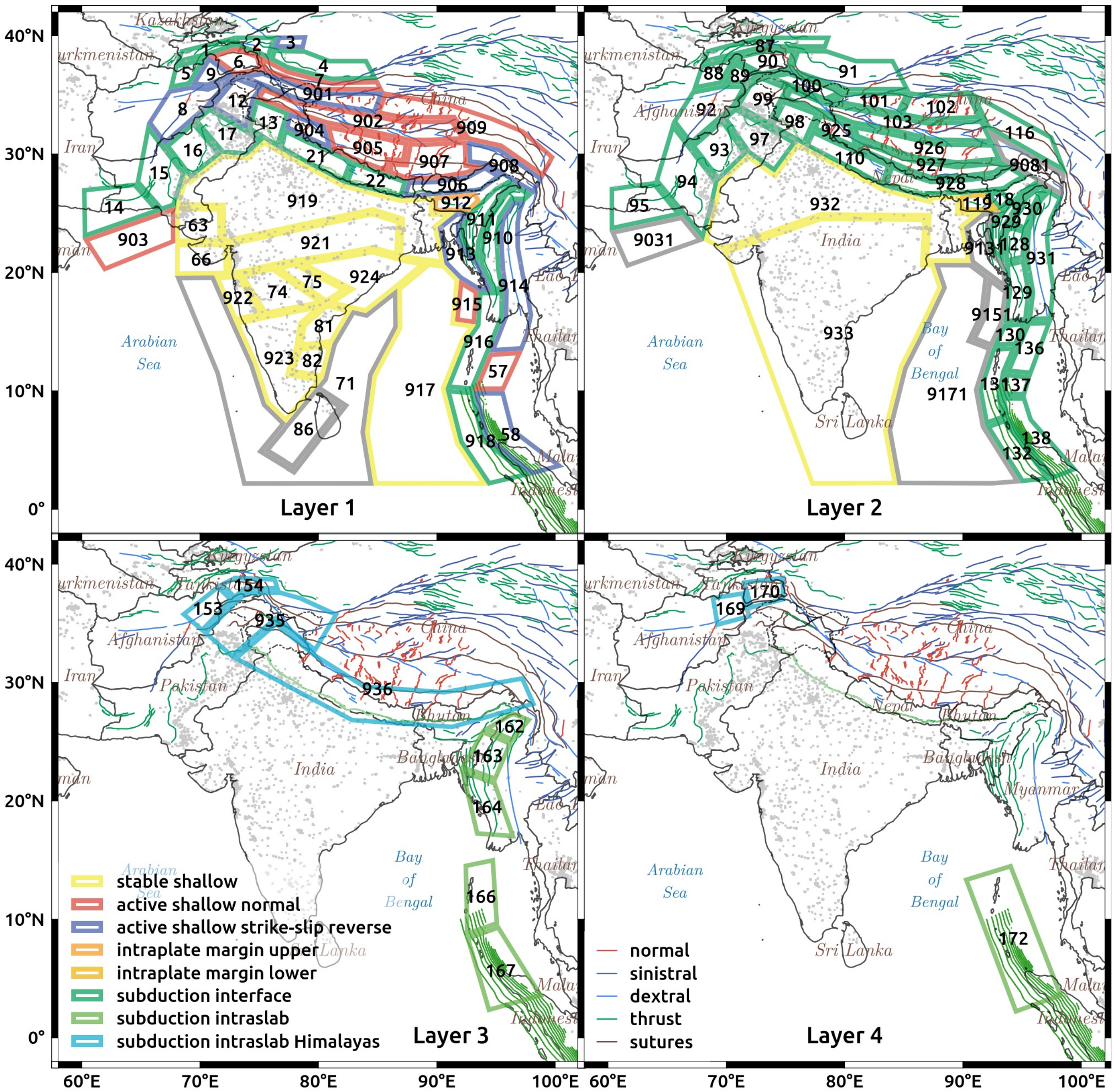

Figure 1: Areal source model tectonic region assignments used in GMPE logic tree. The areal source models is encoded in `areal_source_model.xml`. Zone identification numbers from Nath and Thingbaijam (2012) are indicated. Fault traces are from HimaTibetMap-1.0 (Styron et al., 2010) except the Sumatran 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 km^2" (Schneider et al., 2009), are indicated in darker grey.

numbered “9xx1” in layer 2 have effectively had their seismicity transferred to the corresponding zones 9XX in layer 1. For example zones 32 and 115 in Thingbaijam and Nath (2011) become zones 908 and 9081 in Nath and Thingbaijam (2012), where zone 9081 has no seismicity.

Note that on layers 3 and 4 two distinct tectonic region types are defined for intraslab subduction (Nath and Thingbaijam, 2012, p. 137). Specifically, the “Indo-Myanmar and Andaman-Sumatra subduction zones” are assigned “intraslab” while the “Himalayas and northwest India-Eurasia convergence” are assigned “intraslab Himalayas”. Different GMPEs are applied in these regions, as described in Section 2.3, in particular the Japan and Cascadia adjustments of Atkinson and Boore (2003) are applied, respectively.

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). It was inferred that for interface and intraslab regions `StrasserInterface` and `StrasserIntraslab` should be 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 Appendix C.

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.

Some details of the smoothing are contained in the unpublished Thingbaijam and Nath (2011): a Gaussian kernel was 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.

After some discussion with K. Thingbaijam 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 this missing ingredient is summarized in (Thingbaijam and Nath, 2011, Table 1) and reproduced in Table 1.

Given the total seismicity N and the length in years of the relevant catalogue T (see Table 1) the annual rate ν for a given model is obtained using:

$$\nu = N/T \tag{1}$$

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 in the present study for the smoothed seismicity model the parameters b and m_{max} of the truncated Gutenberg-Richter magnitude-frequency distributions are 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 to the point was chosen.

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

The truncated Gutenberg-Richter magnitude-frequency distribution in OpenQuake implements

$$\lambda(M \geq m) = 10^{a-bm} = e^{\alpha-\beta m}$$

Ignoring events below some threshold m_{min}, 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 for a point source from the activity rate ν for a given magnitude threshold, we take into account the b value for the zone as follows:

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

Similarly to compute the activity rate for an areal source we can use

$$\nu = 10^{a} - 10^{bm_{min}} \tag{2}$$

In order to verify that the smoothed and areal models are approximately equivalent to each other and the catalogue, annual activity rates were computed for each. Areal activity rates were computed using (2). Smoothed model activity rates were computed by summing the seismicity for all points and then applying (1). Catalogue activity rates were computed by querying the catalogue of Nath et al. (2010) with appropriate minimum magnitudes and within the bounds of the areal source model. Note also that events are only counted if the epicentre is within one of the zones of the areal model. This was done on a layer-by-layer basis as well as over the whole model.

The results are tabulated in Table 2. Both the areal and smoothed models tend to overestimate the seismicity in the catalogue. Discrepancies between the areal model and the catalogue are likely an artefact of taking the total seismicity for a given zone, computing a frequency-magnitude distribution, and applying that FMD uniformly over the zone. Discrepancies between the smoothed model and the catalogue cannot be explained by the smearing effect of the smoothing kernel, because this should result in smoothed seismicity rates lower than the catalogue rates when computed over the same area, whereas we observe smoothed seismicity rates which are higher. Improvements to the smoothed seismicity model are proposed in Appendix C.

Other issues of note:

- Zones 9031, 9081, 9131, 9151 and 9171 on layer 2 have m_{max} values values of zero. These zones all 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.
- Given that the Japan/Cascadia regional adjustments are used for intraslab subduction, it is not clear why they are not also applicable for interface subduction.
- 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 2 shows the model at just 0.2° for convenience.

Table 2: Comparison of annual seismicity rates for areal model, smoothed-gridded seismicity model and catalogue. In each case the value shown is the average or expected number of events per year ν above the given minimum magnitude. Catalogue events and smoothed-gridded point sources are only counted if the epicentre is within one of the zones of the areal model.

m_{min}		4.5			5.5	
source	areal	smoothed	catalogue	areal	smoothed	catalogue
layer						
1	80	130	54	8.4	4.1	3.1
2	68	174	78	10.4	3.6	3.8
3	36	89	40	2.9	1.7	1.6
4	12	43	10	1.6	1.2	1.2
total	194	435	182	23.3	10.6	9.7

2.2 Ground-motion prediction

In order to evaluate seismic hazard across the Indian subcontinent it is necessary to consider

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?
- 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.
- 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 (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)

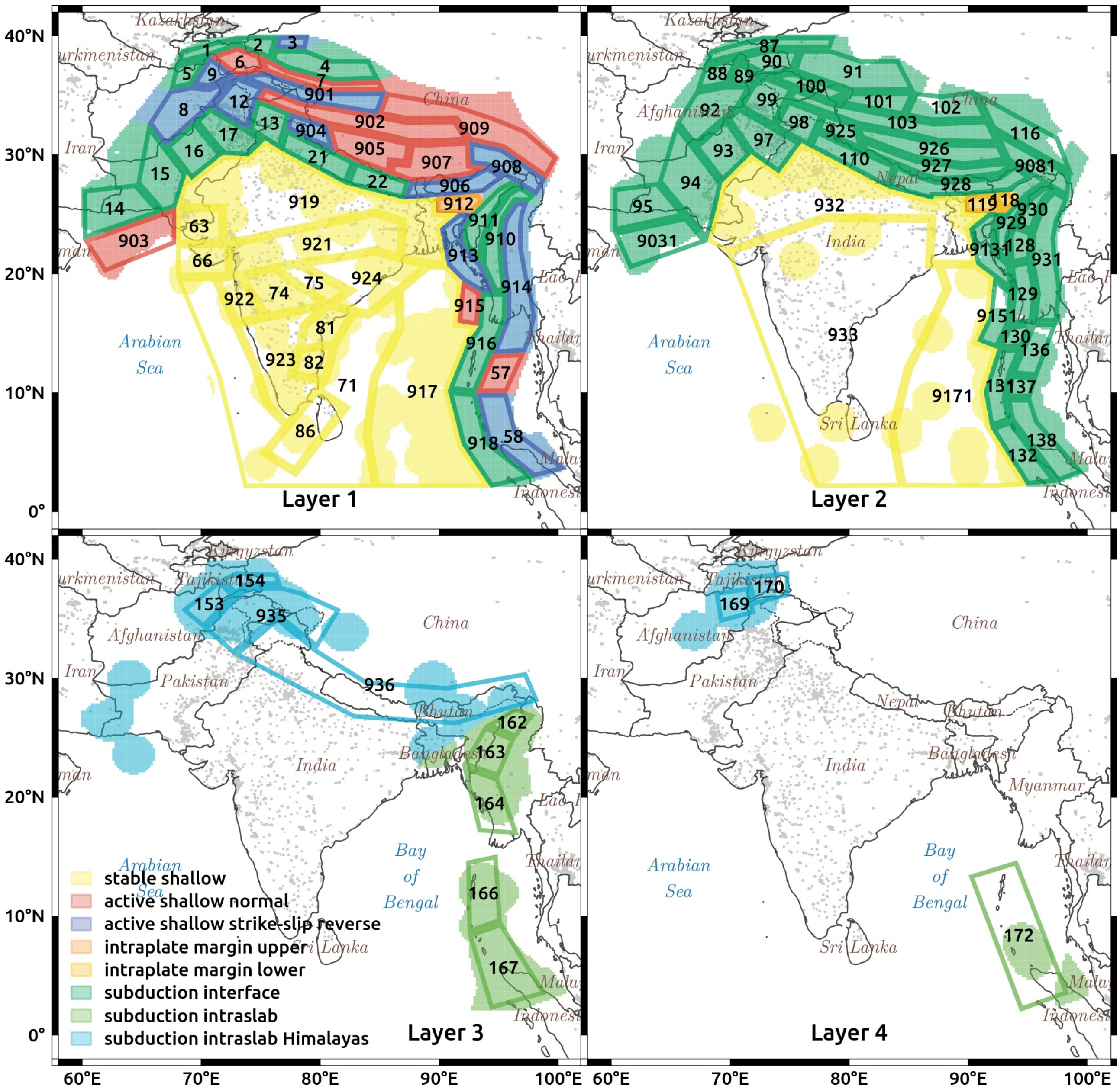

Figure 2: Tectonic region assignments and activity rates for smoothed seismicity point source model. The smoothed seismicity source models are encoded in `smoothed_source_model_mmin4.5.xml` and `smoothed_source_model_mmin5.5.xml`

- does not define PGA (inferred from 0.04 Hz)

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.

[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.”

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.

Table 3: Ground motion prediction equations used in this study. Citations can be inferred from OpenQuake class names. "N" indicates that models were newly implemented in OpenQuake for the current study. "S" indicates that the model has since been superseded by an equivalent model from the same authors. Among the databases used, "ENA" stands for eastern North America, and "NGA" stands for next generation attenuation. The tectonic region "Type" uses the following abbreviations: "active" shallow crust, "intraplate" margin, "stable" continental crust, "interface" subduction and "intraslab" subduction. N_E and N_R are the number of earthquakes and records in the database, respectively. H, M and R are the ranges of depth, magnitude and distance over which the GMPE is considered by the authors to be valid.

OpenQuake class	N	S	Database	Type	N_E	N_R	H [km]		M		R [km]	
`AkkarBommer2010`		✓	Europe & Middle East	active	131	532			5.0	7.6	0	100
`BooreAtkinson2008`		✓	NGA-West	active	58	1574			5.0	8.0	0	200
`CampbellBozorgnia2008`		✓	NGA-West	active	72	942			4.0	8.0	0	200
`Kanno2006Shallow`	✓		Japan	active	83	3769	0	30	5.5	8.2		450
`SharmaEtAl2009`	✓		Himalayas & Zagros	active	16	201			5.0	7.0		100
`NathEtAl2012Upper`	✓		Shillong	intraplate	simulation		0	25	4.8	7.6	10	100
`NathEtAl2012Lower`	✓		Shillong	intraplate	simulation		25	40	4.8	8.1	10	100
`AtkinsonBoore2006`			ENA	stable	10	34800	2	30	5.0	8.3		1000
`Campbell2003`			ENA	stable	hybrid				5.0	8.2	0	1000
`RaghukanthIyengar2007`	✓		Peninsular India	stable	simulation		5	15	4.0	8.0		300
`ToroEtAl2002`			ENA	stable	simulation				5.0	8.0		1000
`Kanno2006Deep`	✓		Japan	subduction	111	8150	30	200	5.5	8.2		450
`AtkinsonBoore2003SInter`			global	interface	80	1155	20	50	5.0	8.3	10	550
`ZhaoEtAl2006SInter`		✓	Japan	interface	269	1520	25	50	5.0	8.3		300
`AtkinsonMacias2009`			Cascadia	interface	simulation				7.5	9.0		400
`AtkinsonBoore2003SSlabJapan`	✓		global	intraslab	80	1155	50	100	5.0	8.3	30	550
`AtkinsonBoore2003SSlabCascadia`	✓		global	intraslab	80	1155	50	100	5.0	8.3	30	550
`LinLee2008SSlab`			NE Taiwan	intraslab	54	4823	39	161	4.1	6.7	40	600
`Gupta2010SSlab`	✓		Indo-Myanmar Arc	intraslab	3	56	91	148	6.3	7.2		375
`YoungsEtAl1997SSlab`			global	intraslab	164	480	50	229	5.0	7.8	10	500
`ZhaoEtAl2006SSlab`		✓	Japan	intraslab	269	1725	50	120	5.0	8.3		300

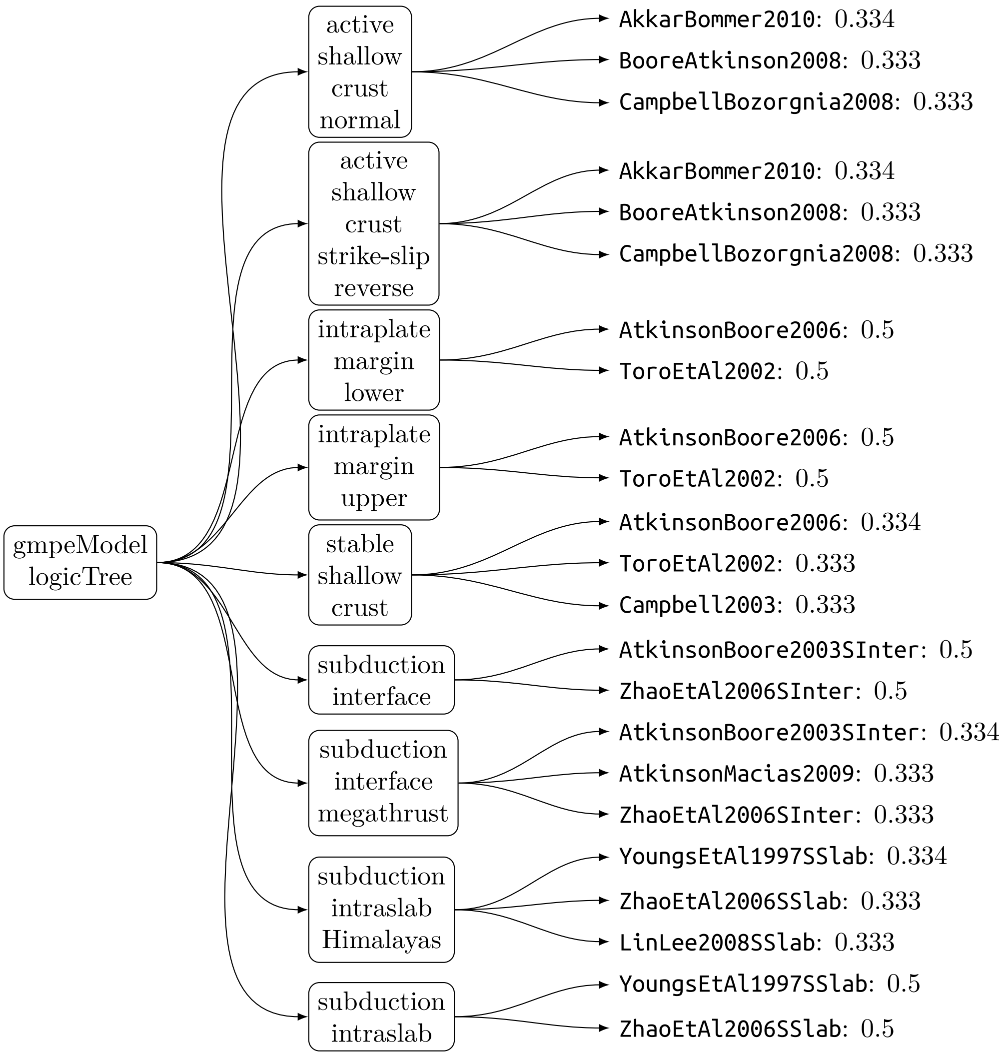

Figure 3: Simplified GMPE logic tree employing only established OpenQuake models, as encoded in `gmpe_logic_tree_omit_new.xml`. Middle column selects tectonic region types as defined in Figure 1. OpenQuake model class names and assigned weights are given on the right side. New and established classes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

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

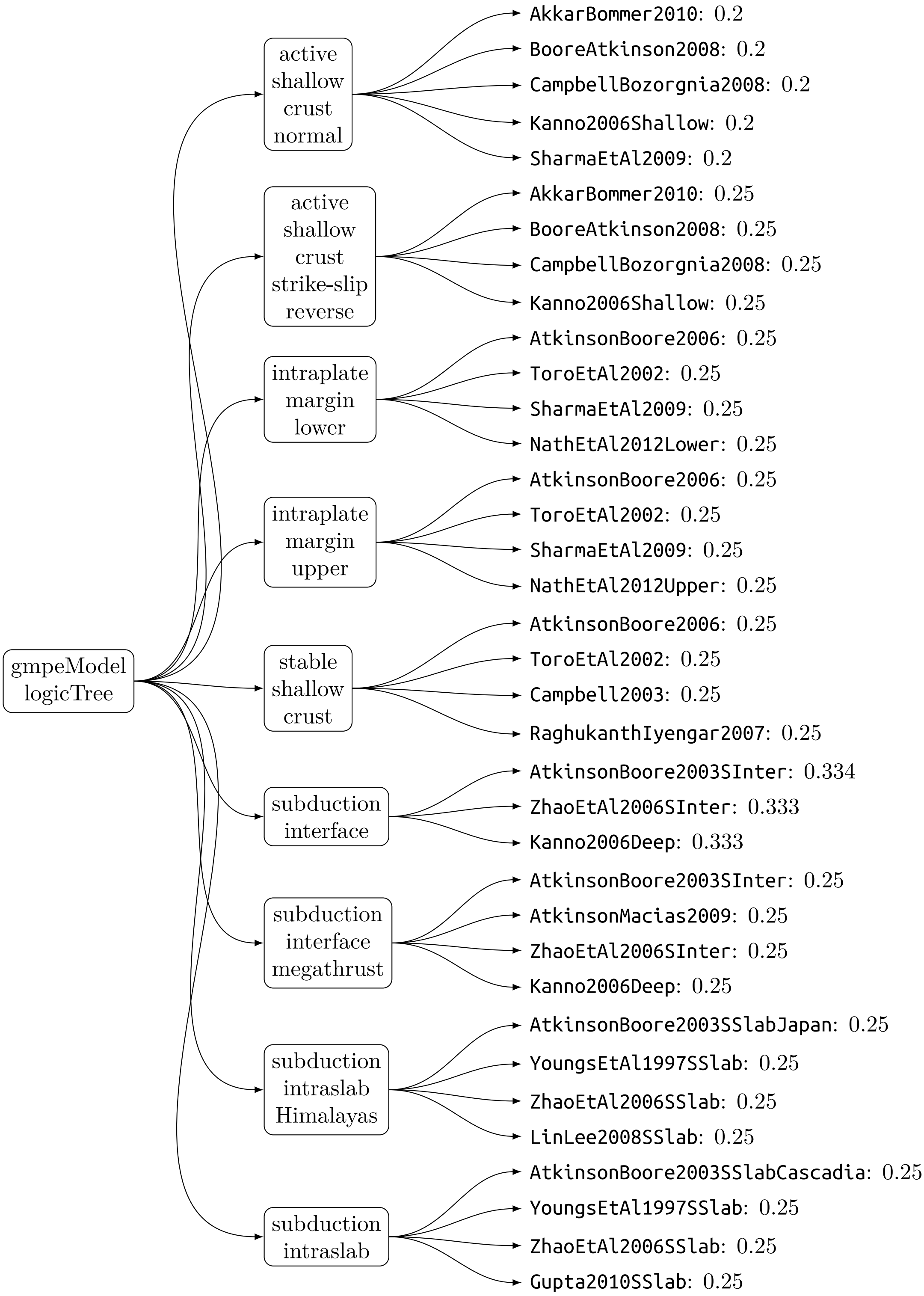

Figure 4: GMPE logic tree of Nath and Thingbaijam (2012), as encoded in `gmpe_logic_tree.xml`. See Figure 3 for complete description

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.

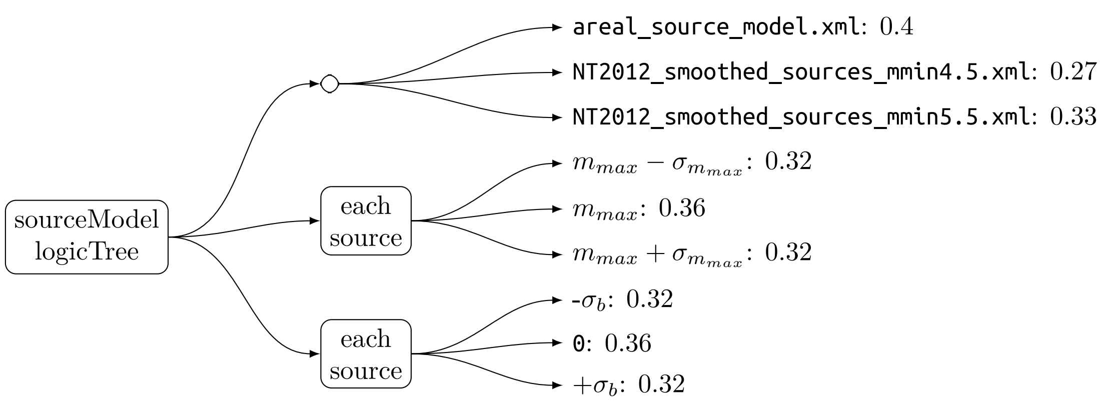

Figure 5: Symbolic source model logic tree of Nath and Thingbaijam (2012).

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.

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

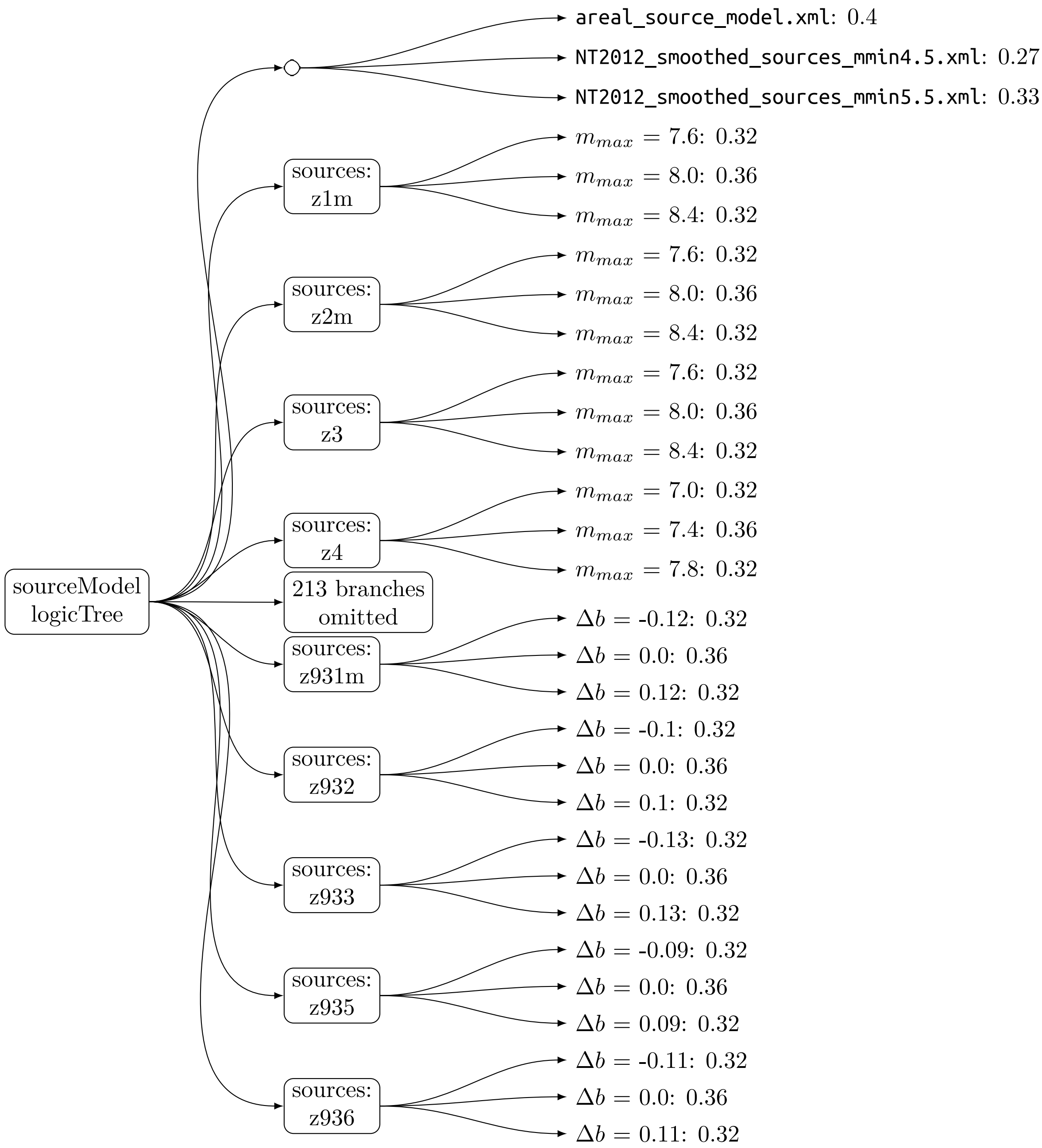

Figure 6: Partial source model logic tree of Nath and Thingbaijam (2012). The full model is encoded in `source_model_logic_tree.xml`

3 Hazard results

3.1 Verification

Validation of PSHA results would require comparison against observed hazard, and would require a much larger dataset than is currently available. In this work we are merely verifying that the current model gives result close to those of Nath and Thingbaijam (2012).

3.2 Sensitivity

3.3 Discussion

4 Conclusions

Acknowledgement

Thanks to Kiran Thingbaijam for clarifications and engaging discussion. Thanks to Amanda for your support.

References

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

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 page 11.)

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 7.)

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 page 13.)

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 9.)

Bilham, R., Gaur, V. K., and Molnar, P. (2001). Himalayan seismic hazard. *Science*, 293(5534):1442–4. (Cited on pages 2 and 21.)

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 11.)

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 page 9.)

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. 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 page 9.)

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 9 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 11.)

Donoho, D. L., Maleki, A., Rahman, I. U., Shahram, M., and Stodden, V. (2009). Reproducible research in computational harmonic analysis. *Computing in Science & Engineering*, 11(1):8–18. (Cited on page 3.)

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 11.)

Fomel, S. and Claerbout, J. F. (2009). Reproducible research. *Computing in Science & Engineering*, 11(1):5–7. (Cited on page 3.)

Frankel, A. (1995). Mapping seismic hazard in the central and eastern united states. *Seismological Research Letters*, 66(4):8–21. (Cited on page 7.)

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.)

Hinsen, K. (2011). A data and code model for reproducible research and executable papers. *Procedia Computer Science*, 4:579–588. (Cited on page 3.)

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 21.)

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 pages 5 and 11.)

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 3, 9, and 11.)

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, 9, 11, 14, 15, 16, and 17.)

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 3, 4, 8, and 21.)

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 3, 4, and 9.)

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 11.)

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 page 9.)

Strasser, F. O., Arango, M., and Bommer, J. J. (2010). Scaling of the source dimensions of interface and intraslab subduction-zone earthquakes with moment magnitude. *Seismological Research Letters*, 81(6):941–950. (Cited on page 7.)

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. K. S. and Nath, S. K. (2011). A seismogenic source framework for the Indian subcontinent. Unpublished manuscript. (Cited on pages 3, 4, 5, and 7.)

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.)

Wells, D. L. and Coppersmith, K. J. (1994). New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement. *Bulletin of the seismological Society of America*, 84(4):974–1002. (Cited on page 7.)

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 5.)

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 9.)

Appendix A Alternative GMPEs

Appendix B Catalogue Evaluation

Appendix C Potential Source Modelling Improvements

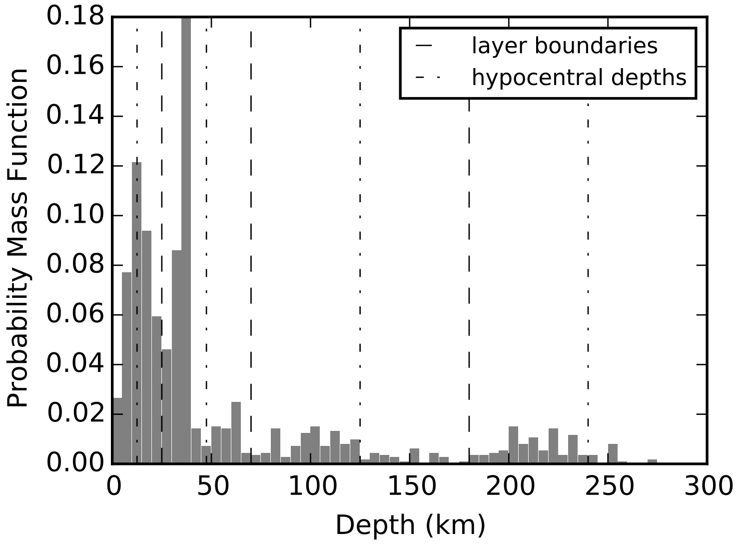

Figure 7: Depth histogram for mainshocks over magnitude 5.5. Mainshock identification is that of Nath et al. (2010). Seismogenic layer boundaries and hypocentral depths used in the current implementation are indicated as dashed and dash-dotted lines respectively.

- 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 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).

Appendix D 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`

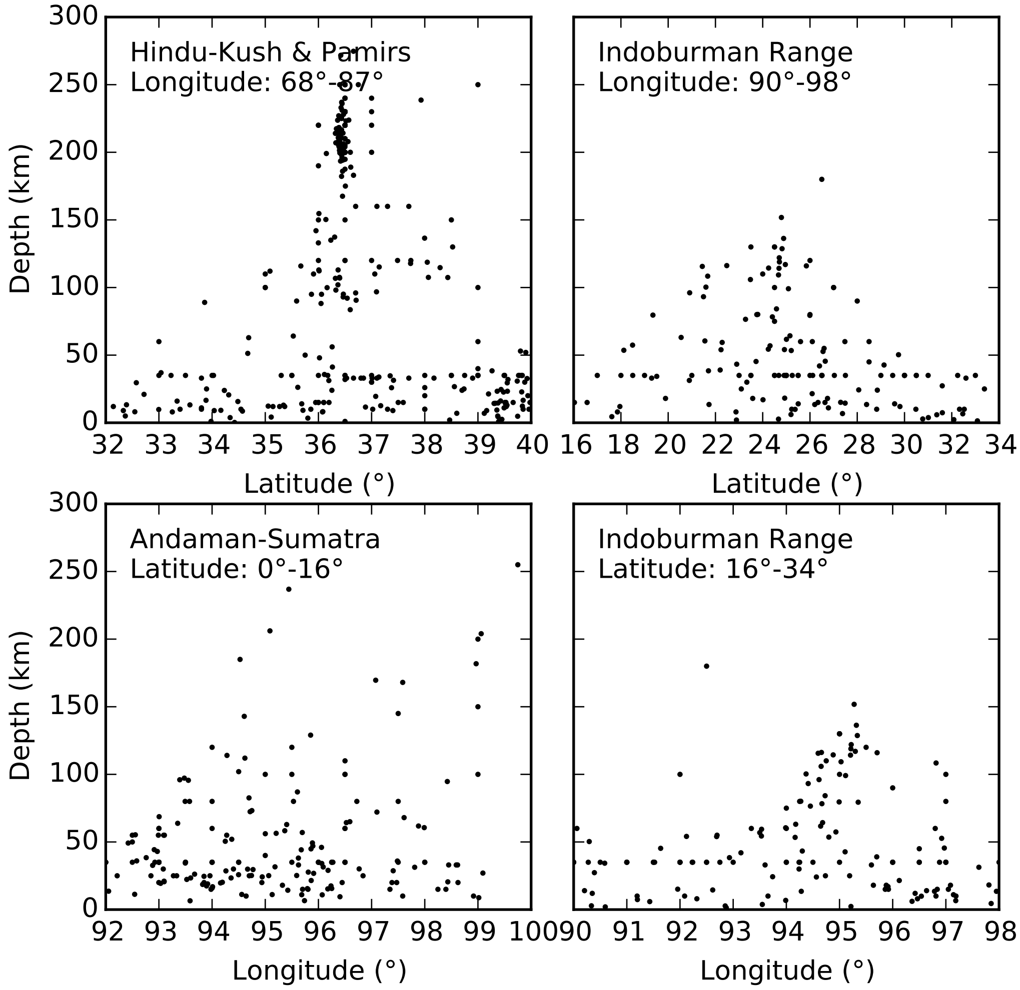

Figure 8: Depth vs. distance for mainshocks in regions with deep events. Subregions are indicated on each map; top left is the Hindu-Kush and Pamir ranges in the northwest of India viewed from the east, top right is the Andoman-Sumatran subduction zone viewed from the south while bottom left and right 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.

Listing 1: phase1–job.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
```