Pure and Applied Geophysics

Epistemic Uncertainty in PSHA and Seismic Hazard Characterization Using the Logic Tree Approach: Part I, Developing the Framework

NARSIRAM GURJAR[1] and DHIMAN BASU[1]

Abstract—Epistemic uncertainty offers alternatives on decision making and various possibilities of computing the hazard integral. Generally, the logic tree approach is used while treating the epistemic uncertainty. Logic tree weight calculation is a subjective decision based on the degree of belief of the analyst on the possible contributors to the epistemic uncertainty and often leads to a different set of values by different researchers. This paper aims to develop a framework of accounting for the epistemic uncertainty in probabilistic seismic hazard analysis (PSHA) by minimizing the subjectivity involved in weight calculation. Guidelines/rules are developed for the weight calculation at each node of the logic tree. Recurrence parameters, magnitude and distance probability distributions, maximum magnitude, and selection of ground motion predictive equations (GMPEs) are considered the possible sources of epistemic uncertainty. A GMPE rule is proposed to be used with the PSHA framework to account for the propagation of epistemic uncertainty. The north-east region of India is chosen for the purpose of illustration. The study region is divided into seven seismic source zones (five in the active crustal region and two in the sub-duction zone). Seismic hazard is characterized in terms of the weighted mean and fractile representation of hazards using the logic tree approach. Only one sample illustration of the results are reported in terms of weighted mean and fractile representation of hazard curves and uniform hazard spectra (UHS). Further illustration of the PSHA results with possible implications from the epistemic uncertainty is reported in the companion paper.

Keywords: Epistemic uncertainty, PSHA, North-east India, GMPE rule, earthquake rupture forecast, sub-earthquake rupture forecast.

List of Symbols

Vs_{30}	Average shear wave velocity in top 30 m
a	Gutenberg-Richter recurrence relation parameter (y-intercept)
b	Gutenberg-Richter recurrence relation parameter (slope)
L	Total number of source zones
q	q^{th} Source zone
a_q	Number of possible candidates for GMPE in q^{th} source zone (finally selected for PSHA)
b_m	Number of possible rules used while defining the maximum magnitude
c	Number of possible rules used to define the probability distribution of magnitudes
d	Number of possible recurrence relations that may define the seismicity
e_r	Number of possible rules used to define the probability distribution of source-to-site distance
N	Number of ERFs
u	u^{th} SERF
p	p^{th} ERF
m	Earthquake magnitude (realization)
r	Source to site distance (realization)
v	v^{th} GMPE
N_{GMPE}	Total number of GMPEs considered
N_{MHC}	Total number of model hazard curves
T^*	Specified or conditional time period
Sa	Spectral acceleration
Sa^*	Given hazard level or specified spectral acceleration
λ	Mean annual rate
$P(\cdot)$	Probability (CDF)
Rup_k	k^{th} Causal rupture scenario
ln	Natural logarithm
$GMPE_j$	j^{th} GMPE

[1] Department of Civil Engineering, Indian Institute of Technology Gandhinagar, Gandhinagar, India. E-mail: dbasu@iitgn.ac.in

N_q	Number of candidates GMPE for the qth source zone screened through the criteria of Bommer et al. (2010)
w_{iq}	Associated weight for i^{th} GMPE (out of N_q) in the qth source zone
$\overline{w}_{jq}$	Normalized weight for j^{th} GMPE (out of a_q) in the qth source zone
s_q	Standard deviation of weights $\left(w_{iq};\ i=1,N_q\right)$ in the qth source zone
w_q	Weight assigned to the qth source zone
w_{SERF}	Weight associated with any typical SERF
w_{ERF}	Weight of the defining ERF
w_{MHC}	Weight associated with any typical MHC
λ_k	k^{th} Realization of the rate
z	z^{th} Fractile
R_{jb}	Joyner-Boore distance (source to site)
R_{epi}	Epicentral distance (source to site)
R_{hypo}	Hypocentral distance (source to site)
R_{rup}	Rupture distance (source to site)
X	Observed sample dataset
x_k	k^{th} Realization of the observed sample data set $(k=1,n)$
M_k	Magnitude associated with k^{th} realization of the observed sample data set
R_k	Source to site distance associated with k^{th} realization of the observed sample data set
T_k	Time period associated with k^{th} realization of the observed sample data set
$\overline{Y}_k$	Median of the intensity measure from GMPE for k^{th} triplet of (M-R-T)
σ_k	Logarithmic dispersion of the intensity measure from GMPE for k^{th} triplet of (M-R-T)
$p(\cdot)$	Probability mass function
$f(\cdot)$	Probability density function
π	Pi
e	Exponential
dx	Small interval of x
B	Base of the logarithm
$\log_B$	Logarithm with base B
LLH_{iq}	Log-likelihood of ith GMPE (out of N_q) in the qth source zone
λ_m	Mean annual rate of exceedance of magnitude m
T_c	Total period of a catalogue
s	Total number of magnitude intervals or sub-catalogues
T_m^s	Completion period in years from recent associated with s^{th} magnitude interval
T_s	Duration of sth sub-catalogue
m_{min}^s	Minimum magnitude of completeness of sth sub-catalogue
n_s	Total number of events in s^{th} sub-catalogue
$m_{n_i}^i$	Magnitude of nth event in i^{th} sub-catalogue
β	Gutenberg-Richter recurrence relation parameter $[\beta = b\ln(10)]$
$L(\cdot)$	Likelihood function
$\hat{\beta}$	Estimated Gutenberg-Richter recurrence relation parameter
$\overline{m}^i$	Mean magnitude of ith sub-catalogue
r_i	Ratio of number of events in i^{th} sub-catalogue to the total number of events in the catalogue (n_i/N_c)
$\hat{N}_i$	Estimated total number of events in i^{th} sub-catalogue between m_{min}^i to m_{max}
N_c	Total number of events in the catalogue
$\hat{N}$	Estimated total number of events in the catalogue
m_{min}	Minimum magnitude of the catalogue
m_{max}	Maximum magnitude of the catalogue
m_{max}^{obs}	Maximum observed magnitude of the catalogue
$\hat{\lambda}_{m_{min}}$	Estimated mean annual rate of exceedance of minimum magnitude
N_m^k	Number of events in k^{th} magnitude interval, within completion period $\left(T_m^k\right)$
λ^{obs}	Observed mean annual rate of exceedance
E_{SL}^q	Mean squared error in the estimation of mean annual rate of exceedance by straight line fitting method in q^{th} source zone

Symbol	Definition
$\lambda^{k,\mathrm{obs}}$	Observed mean annual rate of exceedance in k^{th} magnitude interval
$\hat{\lambda}_m^{k,SL}$	Estimated mean annual rate of exceedance in k^{th} magnitude interval by straight line fitting method
E_{LLh}^q	Mean squared error in the estimation of mean annual rate of exceedance by Log-likelihood method in q^{th} source zone
$\hat{\lambda}_m^{k,LLH}$	Estimated mean annual rate of exceedance in k^{th} magnitude interval by log-likelihood method
E_{LLH}	Mean squared error in the estimation of mean annual rate of exceedance by log-likelihood method over all the source zones
E_{SL}	Mean squared error in the estimation of mean annual rate of exceedance by straight-line fitting method over all the source zones
θ	Slope
W_{SL}	Associated weight of straight-line fitting method
W_{LLH}	Associated weight of log-likelihood method
Mw	Moment magnitude
R^2	Goodness-of-fit measure
M	Earthquake magnitude as random variable
dm_i	Small interval (delta) of earthquake magnitude
$H(\cdot)$	Heaviside step function
$\delta(\cdot)$	Dirac delta function
n	Sample data set of events
W_{g_j}	Associated weight of jth recurrence relation model
$\hat{m}_{\max}$	Estimated maximum magnitude
$\hat{m}_{\max}^2$	Estimated maximum magnitude by second method
L_f	Fault length
L_0	Extended fault length up to perpendicular line to the fault from the site
$X(m)$	Fault rupture length as a function of the earthquake magnitude
D	Perpendicular distance between source (fault line) and the site
$R_{\min}$	Minimum source-to-site distance
$R_{\max}$	Maximum source-to-site distance
R	Source to site distance as random variable
N_{sc}	Total sample count
$W_{f(r)}$	Weight associated with distance probability distribution
C_β	Normalizing coefficient
p_G	Gamma distribution parameter
q_G	Gamma distribution parameter
$\sigma_{\hat{\beta}}$	Standard deviation of estimated Gutenberg-Richter recurrence relation parameter
L_f^j	Length of jth fault
$m_{\max}^j$	Expected maximum magnitude associated with j^{th} fault
m_0	Factored maximum magnitude
$L_f^{\max}$	Maximum fault length in a source zone
L_f^{median}	Median length of all the faults in a source zone
$W_{\hat{m}_{\max}^l}$	Weight associated with l^{th} method of estimating maximum magnitude
N_S	Number of potential earthquake sources in a region around the site
N_M	Number of discrete magnitude segments considered
N_R	Number of discrete distance segments considered
v_i	Mean annual rate of exceedance of events greater than certain magnitude in i^{th} source
Δm	Size of a discrete magnitude segment
g	Gravitational acceleration (9.81 $m/\sec^2$)
Δr	Size of a discrete distance segment
λ_{IM}	Mean annual rate of exceedance associated with IM- Hazard curve
λ_{Rate}	Mean annual rate of exceedance associated with Rate- Hazard curve
Sa_{IM}	Hazard level (Spectral acceleration) associated with IM- Hazard curve
Sa_{Rate}	Hazard level (Spectral acceleration) associated with Rate- Hazard curve
PSHA	Probabilistic seismic hazard analysis

GMPE	Ground motion predictive equation
UHS	Uniform hazard spectra
1D	One dimensional
MECE	Mutually exclusive and collectively exhaustive
LLH	Log-likelihood
SSHAC	Senior Seismic Hazard Analysis Committee
USNRC	United States Nuclear Regulatory Commission
CBR	Center, body, and range
TDI	Technically defensible interpretations
MCE	Maximum considered earthquake
DBE	Design basis earthquake
NEHRP	National earthquake hazard reduction program
ERF	Earthquake rupture forecast
SERF	Sub-earthquake rupture forecast
MHC	Model hazard curve
IM, im	Intensity measure
ACR	Active crustal region
SZ	Subduction zone
SL	Straight line
GR	Gutenberg-Richter
MB	Main and Burton
EDP	Engineering demand parameter
CDF	Cumulative distribution function
z-UHS	Z-fractile representation of uniform hazard spectrum
10PE50	10% Probability of exceedance in 50 years
2PE50	2% Probability of exceedance in 50 years
GMRotI50	50th percentile (median) orientation or rotation independent geometric mean intensity measure
RotD50	50th percentile (median) orientation or rotation dependent intensity measure
km	Kilometer
CFVB15	GMPE by Cauzzi and Faccioli (2015)
SDBK09	GMPE by Sharma et al. (2009)
KNMF06	GMPE by Kanno et al. (2006)
ASB14	GMPE by Akkar et al. (2014)
OZBE04	GMPE by Özbey et al. (2004)
ZHAO16c	GMPE by Zhao et al. (2016)
ZHAO06	GMPE by Zhao et al. (2006)

ASK14	GMPE by Abrahamson et al. (2014)
IDRI14	GMPE by Idriss (2014)
BSSA14	GMPE by Boore et al. (2014)
CB14	GMPE by Campbell and Bozorgnia (2014)
CY14	GMPE by Chiou and Youngs (2014)
MZAS06	GMPE by McVerry et al. (2006)
NATH12	GMPE by Nath et al. (2012)
ANBA13	GMPE by Anbazhagan et al. (2013)
GUPTA10	GMPE by Gupta (2010)
YOUNG97	GMPE by Youngs et al. (1997)
BCHY16	GMPE by Abrahamson et al. (2016)
ZHAO16	GMPE by Zhao et al. (2016)
ATBO03	GMPE by Atkinson and Boore (2003)
LNLE08	GMPE by Lin and Lee (2008)
ARRO10	GMPE by Arroyo et al. (2010)
NDMA10	GMPE from National disaster management authority report (2010)
SZ2	Second source zone in subduction zone
ACR5	5th source zone in active crustal region
PDF	Probability density function
PGA	Peak ground acceleration
CS	Conditional spectra
GCR	Generalized causal rupture
GCIM	Generalized conditional intensity measure

1. Introduction

Uncertainties are generally categorized into two groups, namely, the aleatory variability and epistemic uncertainty, though the physical basis of distinction is somewhat debatable (Bommer & Scherbaum, 2008; Der Kiureghian & Ditlevsen, 2009). Aleatory variability refers to the intrinsic/congenital component, which is irreducible (stems from the Latin 'alea' meaning dice). Epistemic uncertainty is subjective, stemming from incomplete knowledge and, in fact, reducible with more relevant information. In addition to the location and magnitude of future earthquakes and scatter in the GMPEs (precisely the residuals), several other components can also be considered under aleatory variability, for example, the focal depth, and is generally included in the PSHA directly

through the hazard integral. Sources of epistemic uncertainty that stems from the decision making can be categorized into two broad groups, namely, (i) the decisions pertaining to earthquake rupture prediction, i.e., seismic source characterization; and (ii) the decision on ground motion characterization models (Bradley, 2009; Rodriguez-Marek et al., 2021). Decision making in seismic source characterization includes the possible alternatives of source zone boundaries, recurrence models and their parameters, assigning maximum magnitude to each fault, rules for defining magnitude distribution, and distance metrics. Decision making is also involved in source-to-site distance characterization whenever fault geometry is not completely known, such as strike or dip. (Yagh-maei-Sabegh & Ebrahimi-Aghabagher, 2019). Decision making on the selection of GMPEs was detailed by Cotton et al. (2006) and subsequently revised by Bommer et al. (2010), while Scherbaum et al. (2009) reported a framework for ranking the selected GMPEs. Accounting for the effects of local site amplification also needs decision making and hence is attributed to the epistemic uncertainty. Most GMPEs account for this effect in an average sense through period-dependent site amplification factors over the prediction at the bedrock level. Average shear wave velocity in the top 30 m, denoted as Vs_{30}, is often used as a proxy of the shear wave velocity profile at the site. Another alternative is to arrive at the detailed velocity profile through geotechnical and geophysical investigations followed by a 1D (one dimensional) site response analysis with reasonable accuracy (Rodriguez-Marek et al., 2021).

Epistemic uncertainty offers alternatives on decision making and various possibilities of computing the hazard integral and subsequently results in a suite of hazard curves. Generally, the logic tree approach, derived from the probability tree, is employed while treating the epistemic uncertainty. Opinion differs in interpreting the weights assigned to each terminal branch of the logic/probability tree. While the probability tree works on the total probability theorem under the assumption of mutually exclusive and collectively exhaustive (MECE) events, the weights associated with each terminal branch represent its probability of occurrence. Along this line, the weighted mean representation of the hazard curve is

only recommended to be used in seismic design (McGuire et al., 2005; Musson, 2005). Logic tree interpretation was introduced by Kulkarni et al. (1984) and subsequently defended by Abrahamson and Bommer (2005) though used by various researchers. The weights associated with the terminal branches represent the relative degree of belief of the analyst on each model instead of frequentist interpretation. Subsequently, these weights are considered subjective probability of the consequence of each model and enables a fractile representation of the hazard curve. Marzocchi et al. (2015) presented a nice example illustrating the difference of two interpretations. Bommer and Scherbaum (2008) recommended two possible interpretations of the fractile/weighted mean representation of a hazard (as the case may be) from the suite of hazard curves contingent on the intended application. Bommer et al. (2005) discussed the implementation of the logic tree to account for the propagation on epistemic uncertainty through selection and compatibility of GMPEs. More recent investigation recommends a representative suite approach for GMPEs instead of multiple GMPEs with weights (Atkinson et al., 2014), though the framework is arguably still in its infancy and requires unmanageable complexity for generalization. Finally, populating the nodes of the logic tree with possible alternatives needs thorough understanding on the part of analyst. For example, providing equal weights should be avoided, correlated alternatives (such as a- and b-values of recurrence relation) should not be selected as independent branches, aleatory variables (alternatives on focal depth) should not be considered in the logic tree, etc. Bommer and Scherbaum (2008) discussed the possible use and misuse of the logic tree in PSHA.

A large number of PSHA studies have been reported using the logic tree approach in different parts of the world. In most of the studies, logic tree weights are taken as either equal or based on degree of belief from expert/group of experts. For example, Bradley (2009) and Kalakonas et al. (2020) considered equal weights for GMPEs. Delavaud et al. (2012) considered the logic tree weights for GMPEs as a combination of the log-likelihood (LLH) method (Scherbaum et al., 2009) and judgement from six different experts. Anderson (2018) considered the

equal weights for GMPEs and random/expert judgement-based weights for branches at other nodes of the logic tree. Grünthal et al. (2018) considered the equal weights at most of the nodes of a logic tree and expert judgement at the remaining. Anbazhagan et al. (2019) considered the logic tree weight by the LLH method for GMPEs and expert judgement for branches at other nodes of the logic tree. Tromans et al. (2019) considered weights based on expert judgement of the analyst. Epistemic uncertainty in site response analysis is considered by Rodriguez-Marek et al. (2021) and logic tree weights are taken as equal and discrete representations of a log-normal distribution.

Assigning equal weights to different terminal branches of a logic tree reflects more on the 'lack of relevant information' rather than 'equal confidence' on each alternative. Logic tree weight calculation is a subjective decision based on the degree of belief of the analyst on the possible contributors to the epistemic uncertainty and often leads to a different set of values by different researchers. However the weight should be logically deducible in a peer review process and, in other words, conform to the terminal goal of SSHAC (1997): 'the center, the body, and the range of technical interpretations that the larger technical community would have if they were to conduct the study' and that of USNRC (2018): 'the center, the body, and the range of technically defensible interpretations of the available data, methods and models (CBR of TDI)'.

This paper aims to develop a framework of accounting for the epistemic uncertainties in PSHA. Minimizing the subjectivity involved in weight calculation is another complementary objective. Hence, guidelines/rules are developed for the weight calculation at each node of the logic tree. Sources of epistemic uncertainty considered here are (1) recurrence relation parameter, (2) magnitude probability distribution, (3) distance probability distribution, (4) maximum magnitude, and (5) selection of GMPEs. Proposed methodology of accounting for the epistemic uncertainty and constructing the weighted mean and fractile representation of hazards using the suite of hazard curves (denoted as model hazard curves; MHCs) as outcome from various branches of the logic tree is explained under a separate heading. So North-east India is chosen for the purpose of

illustrating the proposed framework. While one set of sample results is used in the paper for the purpose of illustrating the application of proposed framework, further details on hazard analysis and results are presented in the companion paper (Gurjar and Basu, 2022).

For example, topography, tectonic setup, earthquake catalogue, seismicity, and delineation of source zones of the study region are reported in the companion paper. PSHA is then performed in the companion paper using the logic tree approach and taking into account the sources of epistemic uncertainties addressed above. Weighted mean and fractile representation of hazard of the region are estimated at 2% (maximum considered earthquake, MCE) and 10% (design basis earthquake, DBE) probability of exceedance in 50 years for the NEHRP (National Earthquake Hazard Reduction Program) soil types of B, C, and D. Hazard analysis also reports the weighted mean and fractile representation of uniform hazard spectra (UHS) for 112 district headquarters from seven states of North-east India.

2. Proposed Methodology of Accounting for the Epistemic Uncertainty and Characterization of Seismic Hazard

A step-by-step framework is proposed here to enable inclusion of the epistemic uncertainty in seismic hazard characterization.

2.1. Defining the Sources of Epistemic Uncertainty: Earthquake Rupture Forecast

Let L be the number of source zones and $q(=1, L)$ denotes a typical source zone with a_q number of possible candidates for GMPE using a standard selection procure, for example, the log-likelihood method. Further, b_m, c, d, and e_r are the number of possible rules used to define the maximum magnitude, probability distribution of magnitude, sets of recurrence relation parameters, and probability distribution of source-to-site distance, respectively. Any possible combination of these four rules constitutes a scenario that is defined here as earthquake rupture forecast (ERF). Clearly, $N = b_m.c.d.e_r$ is the

number of ERFs that define the sources of epistemic uncertainty. Ideally, N is the number of hazard curves that can be generated to describe the distribution of seismic hazard at a site.

One single GMPE may not adequately describe the attenuation of intensity measure for the entire study area if comprising different source zones. In such a case, each source zone may be best described by an individual set of GMPEs. Given this understanding, PSHA framework must include the provision of using multiple GMPEs for the hazard estimation. The GMPE rule is required to combine the contribution of applicable GMPEs if the end objective is to compute the fractile hazard (considering different sets of GMPEs in different source zones). However, the GMPE rule is not required if the end objective is only to compute the weighted mean rate hazard. Further, it should be noted that any typical ERF above does not explicitly refer to one GMPE or a set of GMPEs. One typical ERF should be assigned with a 'GMPE rule' that identifies the GMPE for each source zone. Hence, any ERF is further discretized into a set of sub-earthquake rupture forecasts (SERFs) conforming to the assigned GMPE rule as described below.

2.2. GMPE Rule and Its Implementation: Sub-Earthquake Rupture Forecast

The applicable GMPE rule may have multiple interpretations. For example: (i) considering only one GMPE from all the source zones and replacing only one at a time, and (ii) considering one GMPE from any one source zone and weighted attributes of all possible GMPEs in other source zones. The former leads to a large number of MHCs (same as the multiplication of the number of GMPEs over all the contributing source zones). In comparison, the latter leads to the same number of MHCs as the summation of the number of GMPEs over all the contributing source zones. However, both the interpretations lead to a consistent weighted mean hazard.

The first interpretation includes only a single GMPE from each source zone and hence does not account for the epistemic uncertainty (due to the selection of GMPEs) in an MHC. In contrast, the second interpretation (i.e., the proposed GMPE rule

recommending a single GMPE from one source zone and weighted attributes of all possible GMPEs in other source zones) indirectly accounts for the epistemic uncertainty (due to the selection of GMPEs) in an MHC. Hence, the second interpretation is considered in this paper for further processing.

Consider, for example, the uth SERF and the pth ERF. Note that the GMPE rule is defined for the SERFs and hence remains the same in all ERFs, i.e., regardless of p. The GMPE rule is implemented as follows:

(a) u^{th} SERF: All ruptures ($m - r$ pair) in q^{th} source zone will be governed by *only* v^{th} of the a_q number of possible candidates for GMPE, whereas other ruptures in any other source zone ($1, L$ but $\neq q$) will be contributed by the weighted attributes of all possible GMPEs in the respective source zone.

(b) Repeat this rule for all the GMPEs in the qth source zone, i.e., $v = 1, a_q$. This will lead to a set of a_q number of SERFs.

(c) Follow a similar assignment in all source zones $q = 1, L$. This will lead to a set of $\sum_{q=1}^{L} a_q = N_{GMPE}$ number of SERFs. Here, N_{GMPE} denotes the total number of GMPEs considered provided no GMPE is common between two zones. However, $\sum_{q=1}^{L} a_q$ holds for the number of SERFs in general.

Therefore, any of the $\sum_{q=1}^{L} a_q$ number of SERFs is defined as the qth source zone governed by its vth GMPE, $q = 1, L$, and $v = 1, a_q$. Clearly, the number of possible MHCs contributing to the epistemic uncertainty with due consideration of the attributes from GMPEs is given by $N_{MHC} = N \sum_{q=1}^{L} a_q$.

2.3. Construction of a Typical Model Hazard Curve

PSHA has been the focus of interdisciplinary research and is reported by a wide spectrum of researchers. Contingent on the discipline specific needs, many often simplified frameworks are adopted that defy the basic rule of probability. Regardless of the scenario, the following general assumptions are made in this paper for an MHC in the absence of epistemic uncertainty:

(a) Number of occurrences of an earthquake exceeding a magnitude within a time interval (defined as the recurrence model) and hence the number of exceedance of a specific hazard level in a specific site within a time interval follows Poisson distribution. By virtue, 'Poisson's distribution generates itself under the addition of independent random variables' and hence, under the assumption of mutually exclusive and collectively exhaustive causal ruptures, the exceedance rate contributed by all causal ruptures can be added to calculate that of a specific hazard level in a specific site.

(b) Given the occurrence of a causal rupture, aleatory variability of an intensity measure (IM) closely

two scenarios, assuming that the set of considered GMPEs constitutes a mutually exclusive and collectively exhaustive set, one may write using the theorem of total probability

$$P(\ln Sa(T^*) > \ln Sa^*|Rup_k) = \sum_{j=1}^{\sum_{q=1}^{L} a_q} P(\ln Sa(T^*) \\ > \ln Sa^*|GMPE_j, Rup_k)P(GMPE_j|Rup_k). \tag{2}$$

Equation (2) precisely accounts for the propagation of epistemic uncertainty through GMPE, which when substituted into Eq. (1) results in

$$\lambda(\ln Sa(T^*) > \ln Sa^*) \\ = \sum_k \left[\sum_{j=1}^{\sum_{q=1}^{L} a_q} P(\ln Sa(T^*) > \ln Sa^*|GMPE_j, Rup_k)P(GMPE_j|Rup_k) \right] \lambda(Rup_k). \tag{3}$$

follows log-normal distribution. This is specified through GMPEs by predicting the logarithmic mean and standard deviation of the considered IM.

Defining the geomean spectral acceleration at any specified time period (T^*) as the IM, $Sa(T^*)$, the rate of exceedance of a given hazard level Sa^* can be expressed as

$$\lambda(\ln Sa(T^*) > \ln Sa^*) \\ = \sum_k P(\ln Sa(T^*) > \ln Sa^*|Rup_k)\lambda(Rup_k). \tag{1}$$

Here $\lambda(Rup_k)$ denotes the rate of triggering the kth rupture.

While accounting for the epistemic uncertainty in the proposed framework, the contribution of the kth rupture is reiterated here for better clarity: (i) with one single GMPE to some SERFs contingent on the containing source zone; and (ii) with weighted GMPEs to remaining SERFs. Regardless of these

Here $P(GMPE_j|Rup_k)$ defines, given the triggering of the kth rupture, the probability that the jth GMPE would control the propagation of intensity measure from the source to the site. According to the GMPE rule, $P(GMPE_j|Rup_k)$ is 1.0 if the kth rupture contributes to one single GMPE, and otherwise, the weight should be specified for all GMPEs such that sum of the weights is unity. Finally, Eq. (3) is the manifestation of total probability theorem and the weight discussed here has the 'frequentist interpretation'.

2.4. Construction of a Fractile Representation of Hazard Curve

Prior to describing the construction of a fractile representation of a hazard curve, a discussion on assigning weight to each MHC is relevant.

2.4.1 Assigning Weight to MHCs

With respect to the two 'irreconcilable attitudes', namely, the probability tree vs. the logic tree, and the resulting controversy between the weighted mean vs. fractile representation of a hazard, this paper stands by the fractile representation of hazard. In other words, each terminal branch of the logic tree is considered as a 'possible alternative' without any 'frequentist interpretation', and hence the weight assigned to the resulting MHC reflects merely the confidence of analyst on that alternative. Nevertheless, the weights assigned to each MHC should be logically deducible in a peer review process and, in other words, conform to the terminal goal of SSHAC (1997): 'the center, the body, and the range of technical interpretations that the larger technical community would have if they were to conduct the study', and that of USNRC (2018): 'the center, the body, and the range of technically defensible interpretations of the available data, methods and models (CBR of TDI)'.

A typical terminal branch of the logic tree in this paper comprises five nodes as shown in Fig. 1a. Referring to Sect. 2.1, the first four nodes represent the alternative ERFs, and referring to Sect. 2.2, the fifth node considers the alternative SERFs. Nodes associated with ERFs are recommended to be populated with weights conforming to the standard practice but in compliance with the CBR of TDI. Some site-specific information is needed that precludes further generalization at this stage. However, the consideration of SERFs representing the propagation of epistemic uncertainty through GMPEs is somewhat unique in this paper and demands a generalized discussion on populating and assigning the weights. First, the candidate GMPEs will be selected conforming to the selection criteria recommended by Bommer et al. (2010). Next, the selected GMPEs will be ranked according to the log-likelihood (LLH) as reported by Scherbaum et al. (2009). The procedure computes the LLH of each GMPE that passed the selection criteria of Bommer et al. (2010). The weight is next assigned to each candidate inversely proportional to the LLH and all GMPEs with weight higher the 'uniform weight' (reciprocal of the number of candidates) are selected. These two steps are repeated for each source zone.

To explain it further, let N_q be the number of candidate GMPEs for the qth source zone screened through the criteria of Bommer et al. (2010) and w_{iq} represents the associated weights, $i = 1, N_q$. Let a_q be the number of GMPEs with weight $w_{iq} > 1/N_q$ and finally selected for PSHA. The weights assigned to these selected a_q number of GMPEs are rescaled such that sum of the weights is unity. Let us denote these final weights as $\overline{w}_{jq}$, $j = 1, a_q$.

Further, an additional step is considered to quantify the confidence in the selected set of GMPEs (through the criteria of Bommer et al., 2010) in any source zone. It should be noted that this additional step is useful only for the estimation of fractile hazard and does not affect the estimation of weighted mean hazard. A hypothetical case of equal LLH (or weights) for all the GMPEs within a source zone is first considered for further illustration. In other words, all the GMPEs lead to a nearly identical exceedance rate and hence do not account for the epistemic uncertainty effectively. Conversely, the greater the variation in weights within a source zone, the better is the effectiveness in capturing the epistemic uncertainty in that source zone.

Consider, for example, only the qth source zone. The weight (W_q) is a random variable with realizations, $w_{1q}, w_{2q}, \cdots, w_{N_q q}$, depending on (i) the total number of GMPEs considered, and (ii) the observed data set. Parameters of the random variable can be estimated using the moments of different orders. Consider the variance $\left(\sigma_q^2 \right)$ representing the confidence in capturing the epistemic uncertainty (and a spike representing no epistemic uncertainty). In other words, confidence in the sets of selected GMPEs across the source zones in capturing the epistemic uncertainty can be considered directly proportional to the variance of weights, which is denoted as the source zone weightage in this study. For example, the variance for the qth source zone may be estimated as

$$ s_q^2 = \frac{1}{N_q - 1} \sum_{i=1}^{N_q} \left(w_{iq} - \frac{1}{N_q} \sum_{i=1}^{N_q} w_{iq} \right)^2 . \quad (4) $$

Rationale behind this assignment is intrinsic to the motivation of adopting the logic-tree framework.

Assigning equal weights to all terminal branches of the logic tree reflects more on the 'lack of relevant information' rather than 'equal confidence' (should not be confused with 'equally likely' in the probability tree) in each alternative. Therefore, the weight assigned to the qth source zone is proposed to be

$$w_q = \frac{s_q^2}{\sum_{q=1}^{L} s_q^2}. \qquad (5)$$

Finally, the weight associated with any of the $\sum_{q=1}^{L} a_q$ number of SERFs, defined through the qth source zone governed by its jth GMPE, consists of
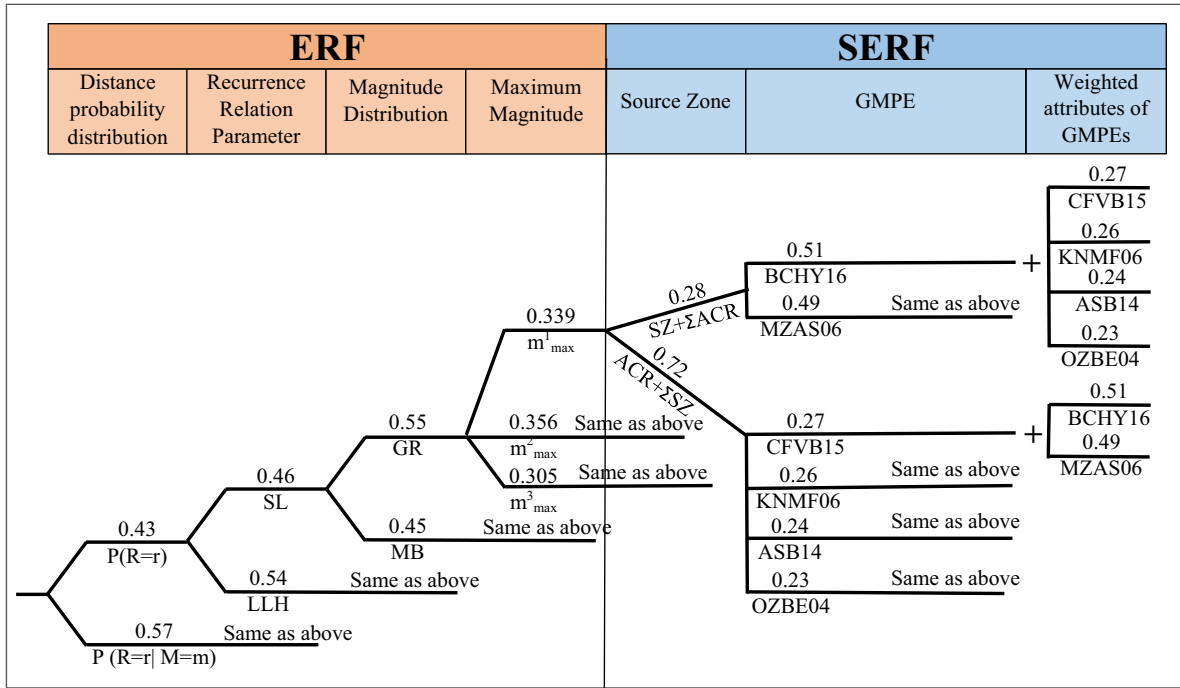
(a) Logic tree diagram used for PSHA and assigning weights to MHCs
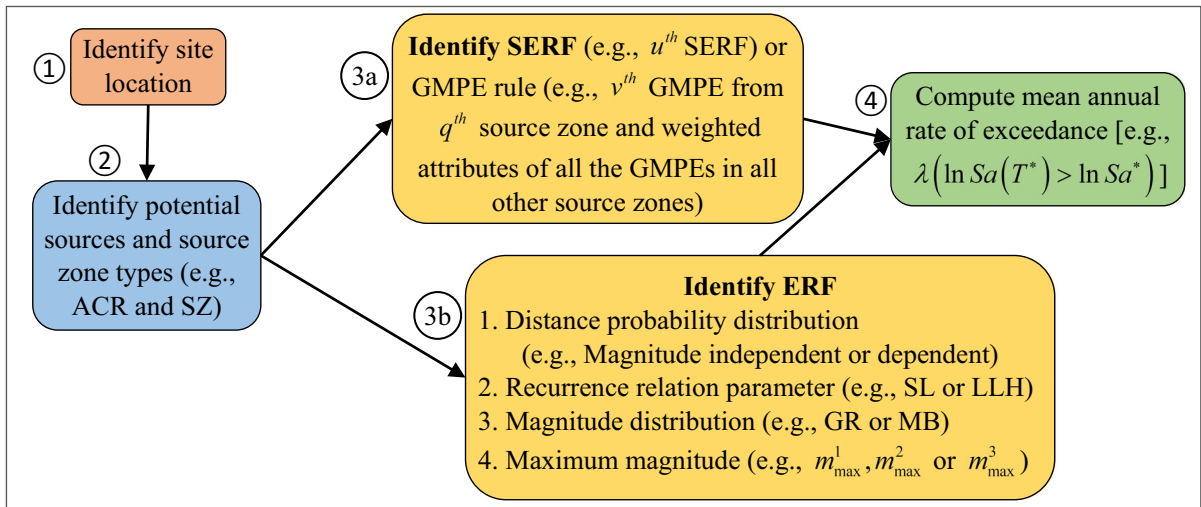
(b) Sample illustration of MHC from logic tree branch

Figure 1
Logic tree diagram

two parts, i.e., within the source zone $(\overline{w}_{jq})$ and across the source zone (w_q), and is expressed as

$$w_{SERF} = w_q\overline{w}_{jq}. \qquad (6)$$

Therefore, the weight associated with any typical MHC is given by

$$w_{MHC} = w_{ERF}w_{SERF}. \qquad (7)$$

where w_{ERF} denotes the weight of the defining ERF.

One should not be confused between the (i) seven total source zones (five for active crustal region, ACR, and two for subduction zone, SZ) defined initially (refer Fig. 1 in the companion paper) and (ii) the only two (one each for ACR and SZ) considered in Fig. 1a. The reason is the paucity of available strong motion data for ranking and selection of GMPEs in the north-east region of India to divide it into seven source zones. Hence, the entire ACR is considered for the selection and ranking of one set of GMPEs, whereas the entire SZ is considered for another set. Therefore, the GMPE rule is implemented using two source zones, whereas all other sources of epistemic uncertainty account for the seven source zones.

One may also argue that the source zones should not be considered as logic tree branches (Fig. 1a). However, as per the GMPE rule associated with a particular SERF, total hazard at the site is governed by a single GMPE from one source zone (contribution 1) and the weighted attributes of all the GMPEs in all other source zones (contribution 2). Hence, each branch of the GMPE in any of the source zones (i.e., contribution 1) is further added to the weighted attributes of all the GMPEs in all other source zones (i.e., contribution 2). Since the GMPE rule considers each source zone twice, it is represented by two branches $(SZ + \Sigma ACR$ and $ACR + \Sigma SZ$; here, ΣACR and ΣSZ represent the weighted attributes of ACR and SZ, respectively) such that the weighted sum is unity. This is also consistent with the estimation of weighted mean hazard. Figure 1b sequentially represents the steps to be followed in the logic tree application for performing PSHA: (i) identification of the site where PSHA is required to be performed, (ii) identification of the sources in the surrounding contributing hazard to the site, (iii) identification of the applicable SERF and ERF based on the type of source zones, and (iv) computation of the mean annual rate of exceedance contributed from all potential sources in the surrounding.

2.4.2 Choice of Dissecting Ensemble MHCs

Consider all the N_{MHC} number of MHCs that are collectively defined as the ensemble, referring back to the frequentist interpretation, two possible alternatives for conditional random variable exist contingent on the possible usage. For example, conditioned to the exceedance of IM by a specified level, 'rate' can be treated as a random variable if the EDP-hazard (engineering demand parameter-hazard) or loss computation is the end objective. This is denoted as the 'vertical dissection'. Alternatively, conditioned to a given exceedance rate, IM can be considered a random variable if the design seismic hazard is the end objective. This is denoted as the 'horizontal dissection'.

With design seismic hazard as the end objective, consider the ensemble MHC and a rate of exceedance, say, $\lambda = \lambda_k$ of an IM describing the seismic hazard. A vector of the IM is constructed with one element from each MHC. Since the IM vector is treated as independent, possible alternatives (independent random variables) from various terminal branches of the logic tree with branch factors indicating their probability of occurrence, referring to Fig. 2a, the weighted mean representation of hazard is given by

$$E(IM) = \mu_{IM} = \int_{-\infty}^{\infty} im.f(im)\,dim$$
$$= \sum_{i=1}^{N_{MHC}} im_i f(im_i)\,dim_i = \sum_{i=1}^{N_{MHC}} W_{MHC}^i im_i, \quad (8)$$

Here, im_i $(i = 1, n)$ is the realization of the IM vector from horizontal dissection of all the MHCs and W_{MHC}^i $(i = 1, n)$ is the branch factor (weight) from the logic tree associated with i^{th} MHC.

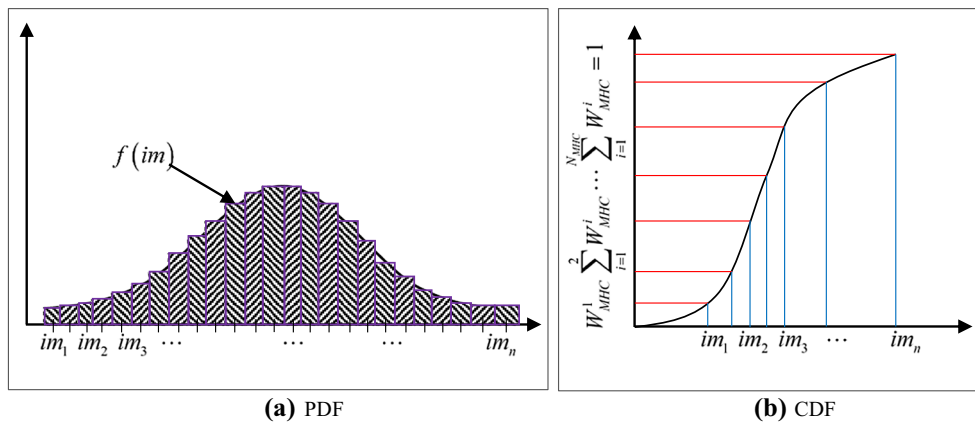Similarly, the fractile representation of hazard is given by

(a) PDF **(b)** CDF

Figure 2
Weighted mean and fractile representation of hazard

$$P\left(IM \le im_p\right) = \Phi\left(im_p\right) = \int\limits_{-\infty}^{im_p} f(im)\,\mathrm{dim}$$

$$= \sum_{i=1}^{im_p} W_{MHC}^i \Rightarrow im_p$$

$$= \Phi^{-1}\left(\sum_{i=1}^{im_p} W_{MHC}^i\right). \qquad (9)$$

Here, im_p is the pth fractile estimate of the hazard, which can be estimated from cumulative distribution function (CDF) of the IM vector constructed non-parametrically (Fig. 2b) considering cumulative branch factors $\left(\sum W_{MHC}^i\right)$ as the cumulative probability. Similar process is repeated for several discrete realization of rate of exceedance ($k = 1, 2, \cdots$) and the same z-fractile estimate is noted. Rate of exceedance when plotted against these z-fractile estimates leads to a fractile representation of the hazard curve, which is denoted here for better clarity as the fractile representation of the IM-hazard curve. A similar procedure can be adopted with vertical dissection and the resulting fractile hazard curve is defined here as the fractile representation of rate-hazard curve.

2.5. Construction of a Fractile Representation of the Uniform Hazard Spectrum and Design Seismic Hazard

IM is next considered at several discrete time periods and the associated z-fractile representation of

IM-hazard curves is constructed. These z-fractile representations of IM-hazard curves are utilized to construct the z-fractile representation of the uniform hazard spectrum (z-UHS) representing a specified rate of exceedance. While the rate of exceedance is usually considered 10PE50 (475-year return period, DBE) or 2PE50 (2475-year return period, MCE), the design seismic hazard accounting for the epistemic uncertainty may be characterized by mean or appropriate fractile representation of UHS.

Further, there is a dividend in studying this fractile representation of IM-hazard. The ratio between the appropriate fractile and weighted mean representation of hazard (not median!) explains an alternate viewpoint of the 'importance factor' commonly used in seismic standards. In such a case, it would be interesting to see the importance factor as period- and site-dependent, as opposed to a constant.

Referring to Sect. 2.4, the weight calculation for a typical MHC (w_{MHC}), resulting from any branch of the logic tree (Fig. 1a) is given by Eq. (7) which requires the w_{SERF} and w_{ERF} to be calculated. Here w_{SERF} is given by Eq. (6) and can be calculated if w_{iq} (associated weights of GMPEs, $i = 1, N_q$ in q^{th} source zone) are known. Hence, the selection of GMPEs and their weight calculation (w_{iq} and $\overline{w}_{jq}$) are explained in the next section.

3. GMPE Selection and Weight Calculation

Proper selection of GMPE is an important step for evaluating the seismic hazard at any site. A few GMPEs are reported to date for North-east India including the Himalayan region due to the limited recordings of strong ground motion data, such as Sharma et al. (2009), Nath et al. (2012), Anbazhagan et al. (2013), and Gupta (2010). However, several GMPEs have been reported around the world for similar tectonic features and site characteristics. Douglas (2020) reported a nice collection of GMPEs developed all around the world to the date of publication.

Generally, three to four GMPEs are good enough for the consideration of epistemic uncertainty due to selection of GMPEs (Stewart et al., 2013). Selection criteria recommended by Bommer et al. (2010) is applied to the database of GMPEs (Douglas, 2020) in the reverse chronological order (starting from the recent), to the selection of ten to 15 GMPEs in each source zone. Hence, a set of 14 and 12 GMPEs are selected for ACR and SZ, respectively, and summarized in Appendix A. All these screened GMPEs use different measures of source-to-site distance such as Joyner-Boore distance $\left(R_{jb}\right)$, epicentral distance $\left(R_{\mathrm{epi}}\right)$, hypocentral distance $\left(R_{\mathrm{hypo}}\right)$ and rupture distance $\left(R_{\mathrm{rup}}\right)$.

Final selection of the GMPEs is based on the ranking using an information theoretic or log-likelihood (LLH) approach reported by Scherbaum et al. (2009), and implementation of which in the present context is briefly described here. Let N_q be the number of candidates GMPEs for ranking in the qth source zone after preliminary screening. Recorded strong motion data at PESMOS (Program for Excellence in Strong Motion Studies) and COSMOS (Consortium of Organizations for Strong Motion Observation Systems) stations in the north-east region with moment magnitude greater than 4 and R_{jb} less than 300 km (reduced if GMPE is not applicable) are considered for log-likelihood estimation. However, R_{jb} is not reported in the database. Therefore, R_{hypo} and R_{epi} are first computed from the information of latitude and longitude of the source and station and focal depth reported in the database. Associated R_{jb} is next constructed using the algorithm of Petersen et al.

(2008). Appendix B summarizes the selected database which is used in the ranking process.

The candidate GMPEs use different intensity measures for the horizontal component of seismic excitation including GMRotI50, RotD50 and geomean. Five percent damped absolute acceleration response spectra are constructed using GMRotI50, RotD50 and geomean as the intensity measure for each recording. Geomean is defined in this paper over the 'as-recorded' pair (without any rotation for along and normal to the principal plane). Other intensity measures such as random, vectorial composition, and horizontal unspecified, if used in the GMPEs, are assumed as geomean in this paper. The observed sample data set for each GMPE is constructed using the associated spectral ordinates sampled at the period of 0.005 s over a period range of zero to the specified maximum.

Let the observed sample data set be defined as $X = \{x_k\}$, with $k = 1, 2, 3, \cdots, n$. Here k^{th} realization is associated with a given triplet of magnitude (M_k), source-to-site distance (R_k) and the time period (T_k). Associated with this k^{th} triplet, each GMPE returns the median $\left(\overline{Y}_k\right)$ and logarithmic dispersion (σ_k) of the respective intensity measures that are assumed to be log normally distributed. Hence, conditioned to a governing GMPE, the probability of kth observed data is given by

$$p\left(X = x_k \big| \overline{Y}_k, \sigma_k\right) = f(x_k)dx$$
$$= \frac{1}{\sigma_k\sqrt{2\pi}} e^{-\frac{[\ln(x_k) - \ln(\overline{Y}_k)]^2}{2\sigma_k^2}} dx. \quad (10)$$

It should be noted that source-to-site distance for the observed events is available in R_{hypo}, R_{epi} and also in R_{jb} using the algorithm of Petersen et al. (2008). In addition, R_{rup} is also required for some GMPEs and, in such cases, R_{jb} is converted into R_{rup} as recommended by Tavakoli et al. (2018). Average sample log-likelihood is next given by the negative of the average of logarithm of joint probability as follows:

$$\mathrm{LLH} = -\frac{1}{n}\sum_{k=1}^{n} log_B\{f(x_k)\}. \quad (11)$$

Here B denotes the base of the logarithm which is taken as 2. Associated weight for the ith (out of N_q) GMPE in q^{th} source zone is given by

Table 1

Ranking and final selection of candidate GMPEs for ACR and SZ

Active crustal region (ACR)					Subduction zone (SZ)				
Rank	GMPE	LLH	w_{iq}	$\overline{w}_{jq}$	Rank	GMPE	LLH	w_{iq}	$\overline{w}_{jq}$
1	**CFVB15**	**2.5883**	0.0989	**0.27**	1	ANBA13	1.6415	0.1097	Rejected (Other)
2	SDBK09	2.6459	0.0950	Rejected (Other)	2	GUPTA10	1.7633	0.1008	Rejected (Other)
3	**KNMF06**	**2.6497**	0.0948	**0.26**	3	YOUNG97	1.7877	0.0991	Rejected (Other)
4	**ASB14**	**2.7590**	0.0879	**0.24**	4	**BCHY16**	**1.9403**	0.0891	**0.51**
5	**OZBE04**	**2.8519**	0.0824	**0.23**	5	**MZAS06**	**1.9888**	0.0862	**0.49**
6	ZHAO16c	2.8619	0.0818	Rejected (Cap)	6	ZHAO16	2.0596	0.0821	Rejected (LLH)
7	ZHAO06	2.8916	0.0802	Rejected (Cap)	7	KNMF06	2.0900	0.0804	Rejected (LLH)
8	ASK14	2.9104	0.0791	Rejected (Cap)	8	ZHAO06	2.1132	0.0791	Rejected (LLH)
9	IDRI14	3.0194	0.0734	Rejected (Cap)	9	ATBO03	2.1372	0.0778	Rejected (LLH)
10	BSSA14	3.0212	0.0733	Rejected (Cap)	10	LNLE08	2.2562	0.0716	Rejected (LLH)
11	CB14	3.4281	0.0553	Rejected (LLH)	11	ARRO10	2.3363	0.0678	Rejected (LLH)
12	CY14	3.7508	0.0442	Rejected (LLH)	12	NDMA10	2.5996	0.0564	Rejected (LLH)
13	MZAS06	4.2482	0.0313	Rejected (LLH)					
14	NATH12	4.7302	0.0224	Rejected (LLH)					

$$w_{iq} = \frac{B^{-LLH_{iq}}}{\sum_{i=1}^{N_q} B^{-LLH_{iq}}}. \qquad (12)$$

Higher the weight better is the ranking and Table 1 presents the GMPEs according to ranking for both ACR and SZ. The GMPEs with weight less than uniform weight, i.e., $w_{iq} < 1/N_q$, are rejected (Anbazhagan et al., 2016). GMPEs are also rejected based on the exclusive requirements of this paper. For example, Sharma et al. (2009) is only applicable to rock and soil sites (NEHRP B and C) and $R_{jb} \leq 100$ km, and hence is not considered in ACR. Similarly, some of the GMPEs are rejected in SZ: 1) Anbazhagan et al. (2013) is applicable to the bedrock only; 2) Gupta (2010) is developed based on only three earthquake data and applicable to Intraslab only; 3) Youngs et al. (1997) is applicable to rock and soil sites. Finally, a cap of four GMPEs is imposed on both ACR and SZ and that leads to i) Rank-1, -3, -4 and -5 GMPEs for the ACR; and ii) Rank-4 and -5 GMPEs for the SZ. The GMPEs that are finally selected for PSHA are highlighted by bold letters in Table 1 with their rescaled weights ($\overline{w}_{jq}$).

4. Earthquake Rupture Forecast (ERF)

The logic tree consists of two parts, ERF and SERF, as shown in the Fig. 1a. SERF refers to the GMPE rule comprising two sections, (i) within the source zone (explained in the Sect. 3), and (ii) across the source zone (explained in the Sect. 2.4). Hence, this section is devoted to the ERF only. ERF is primarily contributed from the first four nodes of the logic tree, namely, (i) distance probability distribution, (ii) recurrence relation parameters, (iii) magnitude distribution, and (iv) maximum magnitude. Any possible combination from these four nodes leads to one realization of ERF. The weight calculations for the first four nodes of the logic tree (Sect. 2.1, Fig. 1a) are explained in the subsequent sections. Finally, the weight associated with any realization of ERF, w_{ERF}, is given by the product of the weights contributed from the constituting nodes of the logic tree.

It should be noted that the logic tree starts in a sequence: (i) distance probability distribution, (ii) recurrence relation parameters, (iii) magnitude distribution, (iv) maximum magnitude and (v) selection of GMPEs. However, these sections are explained in a different sequence. Nevertheless, it is instructive to note the interdependence of some of these sections. For example, (i) weight calculation for maximum
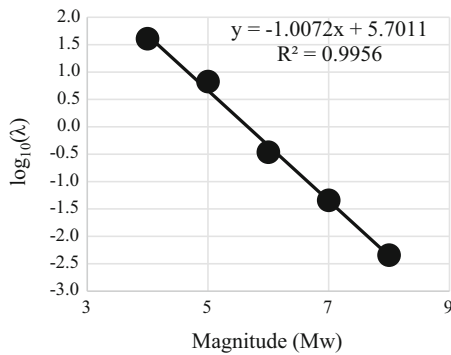
Figure 3

Frequency magnitude relationship for ACR Zone 1 by straight line fitting

magnitude depends on the magnitude distribution and the recurrence relation parameters, and (ii) the weight calculation for magnitude distribution depends on the maximum magnitude and the recurrence relation parameters. Hence, for better clarity, the independent sections, for example, selection of GMPEs (Sect. 3) and recurrence relation parameters (Sect. 4.1) are discussed first. Interdependent sections are explained next, for example, magnitude distribution (Sect. 4.2), distance probability distribution (Sect. 4.3), and maximum magnitude (Sect. 4.4).

4.1. Estimation of Recurrence Relation Parameters

Gutenberg-Richter (1944) recurrence relation is expressed as $\log_{10}(\lambda_m) = a - bm$ with a and b as the recurrence parameters, whereas λ_m is the mean annual rate of exceedance of magnitude m. Two of the most widely used methods of computing the recurrence parameters, namely (i) the least squares

method of fitting a straight line (SL) and (ii) the maximum log-likelihood (LLH) method (Kijko & Smit, 2012), are considered in this paper as the possible contributors to epistemic uncertainty. While the first method is generally used and a sample illustration for the first zone of ACR is presented in Fig. 3, the maximum log-likelihood method requires some discussion on fundamentals for its implementation in this paper. This is explained with an illustration presented in Table 2 for the first zone of ACR.

Let a catalogue over a period of T_c years be known for the qth source zone. The catalogue is discretised into a set of s magnitude intervals and Stepp's (Stepp, 1972) method is employed for the evaluation of associated completion periods $(T_m^1, T_m^2, T_m^3, \ldots, T_m^s)$. With reference to Table 2 (rows 2 and 3), the completion periods of 30, 50, 190 and 220 years from recent (31/07/2020) are obtained for magnitude intervals of 4–5, 5–6, 6–7 and > 7, respectively. The information of completion period and associated magnitude intervals are recast to construct a set of sub-catalogues of duration $T_1, T_2, T_3, \ldots, T_s$ with minimum (complete) magnitude in each sub-catalogue as $m_{min}^1, m_{min}^2, m_{min}^3, \ldots, m_{min}^s$, respectively, and total number of events as $n_1, n_2, n_3, \ldots, n_s$, respectively. Unlike Stepp's method, here the time is measured from the oldest, and hence, $m_{min}^1 > m_{min}^2 > m_{min}^3 > \cdots > m_{min}^s$. Rows 4 and 5 of Table 2 illustrate this calculation from the results of Stepp's method. The complete catalogue is then reconstructed by compiling all the sub-catalogues

Table 2

Sub-catalogues for ACR Zone-1

Sub-catalogue →	C1	C2	C3	C4
Starting date	31/7/1800	31/7/1830	31/7/1970	31/7/1990
End date	30/7/1830	30/7/1970	30/7/1990	31/7/2020
Duration of sub-catalogue (T_i)	30	140	20	30
Minimum magnitude $\left(m_{min}^i\right)$	7.7	6	5	4
Mean magnitude ($\overline{m}^i$)	7.7	6.4093	5.3032	4.5427
Number of events	1	54	125	1222
r_i	7.13×10^{-4}	0.0385	0.0892	0.8716
$\hat{\beta}_i$	Inf	2.4434	3.2982	1.8426

such that: (i) duration $\sum_{i=1}^{s} T_i = T_c$; and (ii) number of events $\sum_{i=1}^{s} n_i = N$.

Let $m_1^i, m_2^i, m_3^i, \ldots, m_{n_i}^i$ represent the observed magnitudes in the ith sub-catalogue. Under the assumption of mutual independency, the joint probability density function for the ith sub-catalogue conditioned to $\beta = b \ln(10)$ is given by

$$f(m|\beta) = f\left(m_1^i|\beta\right)f\left(m_2^i|\beta\right)f\left(m_3^i|\beta\right)\cdots f\left(m_{n_i}^i|\beta\right)$$
$$= \prod_{j=1}^{n_i} f\left(m_j^i|\beta\right). \tag{13}$$

Therefore, likelihood of the entire catalogue may be constructed as

$$L(\beta) = \prod_{i=1}^{s} \prod_{j=1}^{n_i} f\left(m_j^i|\beta\right). \tag{14}$$

The parameter β can be evaluated by maximizing the likelihood and, alternatively, the logarithm of the likelihood. Hence, by taking natural logarithm of Eq. (14) and for the maxima,

$$\ln[L(\beta)] = \sum_{i=1}^{s} \sum_{j=1}^{n_i} \ln\left[f\left(m_j^i|\beta\right)\right] \to \frac{d}{d\beta}\ln[L(\beta)] = 0. \tag{15}$$

Using the lower bound Gutenberg-Richter recurrence relation, one may write,

$$f\left(m_j^i\Big|\hat{\beta}\right) = \hat{\beta}e^{-\hat{\beta}\left(m_j^i - m_{\min}^i\right)} \to \ln\left[f\left(m_j^i\Big|\hat{\beta}\right)\right]$$
$$= \ln\left(\hat{\beta}\right) + \hat{\beta}\left(m_{\min}^i - m_j^i\right). \tag{16}$$

Using Eq. (16) into Eq. (15) and after some simplifications, one may write

$$\sum_{i=1}^{s} \sum_{j=1}^{n_i} \frac{1}{\hat{\beta}} + \sum_{i=1}^{s} \sum_{j=1}^{n_i}\left[-\left(m_j^i - m_{\min}^i\right)\right] = 0$$
$$\to \frac{\sum_{i=1}^{s} n_i}{\hat{\beta}} = \sum_{i=1}^{s} n_i\left[\left(\frac{1}{n_i}\sum_{j=1}^{n_i} m_j^i\right) - m_{\min}^i\right]. \tag{17}$$

Further, assuming $\left(\frac{1}{n_i}\sum_{j=1}^{n_i} m_j^i\right) - m_{\min}^i = \overline{m}^i - m_{\min}^i = \frac{1}{\hat{\beta}_i}$ and $r_i = n_i/\sum_{i=1}^{s} n_i$, Eq. (17) may be solved for

Table 3

Seismicity parameters for the north-east region

Zone	Straight line fitting (SL) method		Log-likelihood (LLH) method	
	a value	b value	a value	b value
SZ1	5.96	0.99	4.84	0.80
SZ2	3.82	0.73	3.99	0.78
ACR1	5.70	1.01	4.91	0.84
ACR2	5.14	0.90	4.23	0.74
ACR3	4.38	0.85	3.48	0.72
ACR4	3.74	0.76	3.49	0.70
ACR5	5.23	0.99	3.55	0.71

$$\frac{1}{\hat{\beta}} = \sum_{i=1}^{s} \frac{r_i}{\hat{\beta}_i}. \tag{18}$$

Equation (18) is known as the generalized Aki-Utsu b value estimator and also well explained by Kijko and Smit (2012).

Estimated mean annual rate of exceedance corresponding to $m_{\min}$ is given by

$$\hat{\lambda}_{m_{\min}} = \hat{N}/T_c. \tag{19}$$

Here, $\hat{N} = \sum_{i=1}^{s} \hat{N}_i$ and $\hat{N}_i$ is the total number of events in the ith sub-catalogue between $m_{\min}^i$ to $m_{\max}$, which can be estimated using the probability distribution of magnitude as follows

$$\hat{N}_i = \frac{n_i}{F_M\left(m_{\min}^i \leq m \leq m_{\max}|m_{\min}, m_{\max}\right)}$$
$$= \frac{n_i}{1 - F_M\left(m_{\min} \leq m \leq m_{\min}^i|m_{\min}, m_{\max}\right)}. \tag{20}$$

Here, $F_M\left(m_{\min} \leq m \leq m_{\min}^i|m_{\min}, m_{\max}\right) = \frac{1 - e^{-\beta\left(m_{\min}^i - m_{\min}\right)}}{1 - e^{-\beta(m_{\max} - m_{\min})}}$ is given by the doubly truncated Gutenberg-Richter recurrence relation with $m_{\min} = 4$ (events below which can be ignored due to lack of engineering importance) and $m_{\max} = m_{\max}^{obs}$ (maximum observed magnitude), and n_i is the number of events in the ith sub-catalogue between $m_{\min}^i$ to $m_{\max}$. Values of a and b can be estimated using the relations $\hat{\beta} = b\ln(10)$ and $\hat{\lambda}_{m_{\min}} = 10^{a - bm_{\min}}$.

With reference to Table 2, for the sample illustration, $N_c = 1402$ and $m_{\max}^{obs} = 8$, with

Table 4

Observed and estimated mean annual rate (λ) from (a, b values) of SL and LLH method for ACR Zone-1

Magnitude (Mk)	Completion period (T_m^k)	N_m^k	N_m^k/T_m^k	$\lambda^{k,obs}$	$\hat{\lambda}_m^{k,SL}$	$\hat{\lambda}_m^{k,LLH}$
4	30	1016	33.8667	40.5921	47.0245	35.0418
5	50	319	6.3800	6.7255	4.6253	5.0427
6	190	57	0.3000	0.3455	0.4549	0.7257
7	220	9	0.0409	0.0455	0.0447	0.1044
8	220	1	0.0045	0.0045	0.0044	0.0150

Table 5

Mean squared error between the observed and estimated rates of SL and LLH methods

Zone	Mean squared error	
	Straight line fitting method (E_{SL})	Log-likelihood method (E_{LLH})
SZ1	150.47	150.22
SZ2	1.32	2.02
ACR1	9.16	6.76
ACR2	15.86	15.79
ACR3	0.36	5.14
ACR4	0.23	0.35
ACR5	14.29	10.21

Figure 4
Zero-intercept best-fit straight-line relationship between the mean squared error of SL and LLH methods

$m_{\min} = 4$. Using Eqs. (18) and (19), one may calculate $\hat{\beta} = 1.9386$ and $\hat{\lambda}_{m_{\min}} = 35.0418$, respectively, and finally $b = 0.84$ and $a = 4.91$. Recurrence parameters obtained by both the methods for all the source zones are reported in Table 3.

4.1.1 Weight Calculation for the Process of Estimating Recurrence Relation Parameters

Weights associated with each method of estimating recurrence relation parameters for the logic tree approach are proposed to be inversely proportional to mean squared error over all the source zones. For example, consider one of the source zones, say q^{th}. Let T_m^k be the completion period of the kth magnitude interval with N_m^k number of events within the completion period. Table 4 presents a sample illustration for the first zone of ACR. Recurrence parameters are estimated using the procedure described above. Associated mean rate of occurrence
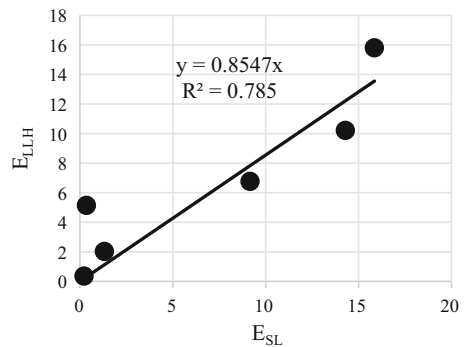
is estimated as $\lambda_m = 10^{(a-bm)}$ for both the methods (for example, the last two columns of Table 4). The observed mean annual rate of exceedance (λ^{obs}) is estimated using the cumulative sum of N_m^k/T_m^k starting from the highest magnitude interval (for example, column 5 of Table 4). Mean squared error for both the methods, namely, SL and LLH, are given

by $E_{SL}^q = \frac{1}{s} \sum_{k=1}^{s} \left(\lambda^{k,obs} - \hat{\lambda}_m^{k,SL} \right)^2$ and $E_{LLH}^q = \frac{1}{s} \sum_{k=1}^{s} \left(\lambda^{k,obs} - \hat{\lambda}_m^{k,LLH} \right)^2$, respectively. Table 5 compares the mean squared error using both the methods for each source zone.

Both errors in all source zones are compared and the best-fit linear relation is constructed with a constraint of zero intercept, i.e., $E_{LLH} = \theta \times E_{SL}$ (Fig. 4). The first source zone of SZ is neglected as an outlier and θ is estimated as 0.85. Finally, the weights are computed as inversely proportional to the mean squared error as follows:

(a) Probability density function
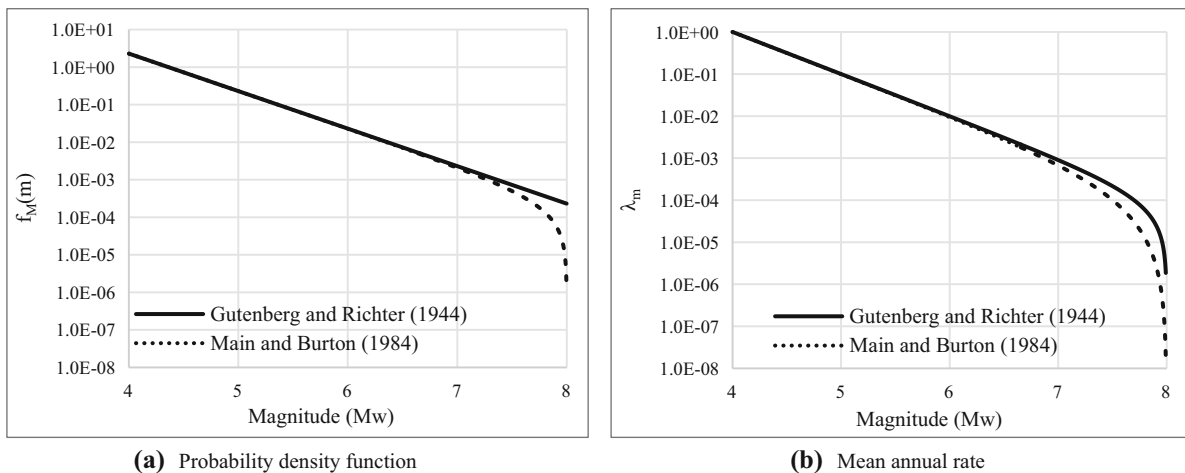

(b) Mean annual rate

Figure 5
Variation of probability density function and mean annual rate with magnitude for Gutenberg-Richter (1944) and Main and Burton (1984) recurrence relations [$m_{\min} = 4$, $m_{\max} = 8$, $a = 4$ and $b = 1$]

$$W_{SL} = \frac{1/E_{SL}}{(1/E_{SL} + 1/E_{LLH})} = \frac{\theta}{1+\theta} = \frac{0.85}{1+0.85}$$
$$= 0.46, \quad W_{LLH} = \frac{1/E_{LLH}}{(1/E_{SL} + 1/E_{LLH})} = \frac{1}{1+\theta}$$
$$= \frac{1}{1+0.85} = 0.54.$$

(21)

not have any documented characteristic events in any of the source zones, the characteristic earthquake model is not included. Therefore, two alternatives of magnitude distribution are considered in this paper as the possible contributors to epistemic uncertainty, namely, (i) bounded Gutenberg-Richter (GR) (1944) and (ii) Main and Burton (MB) (1984). Associated magnitude distributions, $f_1(m_i)$ and $f_2(m_i)$, respectively, are given by

$$p(M = m_i | m_{\max}, \beta, m_{\min} \leq m_i \leq m_{\max}) = f_1(m_i)dm_i = \frac{\beta e^{-\beta(m_i - m_{\min})}}{1 - e^{-\beta(m_{\max} - m_{\min})}} dm_i \quad (\text{Gutenberg} - \text{Richter}, 1944) \quad (22)$$

$$p(M = m_i | m_{\max}, \beta, m_{\min} \leq m_i \leq m_{\max}) = f_2(m_i)dm_i$$
$$= \frac{\beta\left[e^{-\beta(m_i - m_{\min})} - e^{-\beta(m_{\max} - m_{\min})}\right]H(m_{\max} - m_i) + \left[e^{-\beta(m_i - m_{\min})} - e^{-\beta(m_{\max} - m_{\min})}\{1 + \beta(m_{\max} - m_i)\}\right]\delta(m_{\max} - m_i)}{\left[1 - e^{-\beta(m_{\max} - m_{\min})} - \beta e^{-\beta(m_{\max} - m_{\min})}(m_{\max} - m_{\min})\right]} dm_i$$

(Main and Burton, 1984)

(23)

4.2. Magnitude Distribution/Recurrence Relation

Recurrence relation defines a relation between the mean annual rate of exceedance and magnitude and hence the magnitude distribution as well. Since the study is applied to the north-east region which does

Similarly, the mean annual rate of exceedance, $\lambda_1(m_i)$ and $\lambda_2(m_i)$ respectively, are given by

$$\lambda_1(m_i) = v\frac{e^{-\beta(m_i - m_{\min})} - e^{-\beta(m_{\max} - m_{\min})}}{1 - e^{-\beta(m_{\max} - m_{\min})}} \quad (24)$$

(Gutenberg-Richter, 1944)

$$\lambda_2(m_i) = v\Big[e^{-\beta(m_i - m_{\min})} - e^{-\beta(m_{\max} - m_{\min})} \{1 + \beta(m_{\max} - m_i)\}\Big]H(m_{\max} - m_i) \quad (25)$$

(Main and Burton 1984) Here, $v = e^{(\alpha - \beta m_{min})}$ and $\alpha = a1n$ (10).

Probability density function (PDF) (magnitude distribution) and mean annual rate of exceedance are plotted against magnitude in Fig. 5a and 5b, respectively, for the purpose of illustration.

4.2.1 Weight Calculation for Magnitude Distribution

Weights associated with each alternative are computed for the logic tree using a log-likelihood approach similar to that employed in the case of GMPEs. Let $M = \{m_i\}$, with $i = 1, 2, 3, \cdots, n$ be the sample data set of observed events (seismic catalogue) in any of the source zones, say q^{th}. Since all the observed events are independent, denoting $g(m_i) = f(m_i)dm_i$ as the likelihood of model $g[g_1$ or $g_2]$ for the observed sample m_i, the associated joint probability is given by $L(g|M) = \prod_{i=1}^{n} g(m_i)$. Negative of the average sample log-likelihood is considered here to define the log-likelihood (LLH):

$$LLH(g) = -\langle log_B(L(g|M))\rangle = -\frac{1}{n}\sum_{i=1}^{n} log_B(g(m_i)). \quad (26)$$

Here, B refers to the base of the logarithm and is taken as 2 (can also be taken as 10, Scherbaum et al., 2009). The negative sign is considered here, as the logarithm of probability is always non-positive.
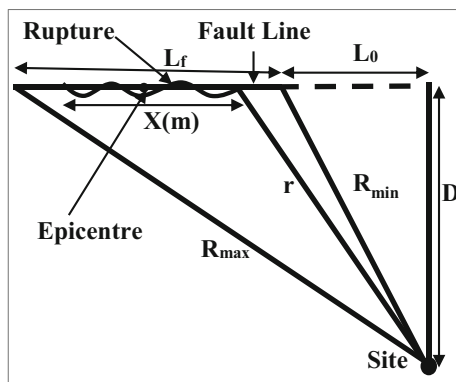
Clearly, the LLH is superior and is the method of recurrence relation. Associated weightage of recurrence relations is given by

$$W_{g_j} = B^{-LLH(g_j)} \Big/ \sum_{j=1}^{2} B^{-LLH(g_j)}. \quad (27)$$

Since PDFs of both methods are identical at low magnitudes with a little difference near the maximum magnitude, none of the observed events ($m_{\min}$ to $m_{\max}^{obs}$) should be considered while calculating the LLH, which would otherwise appear as nearly identical. It is reasonable to consider events in the range of $[\hat{m}_{\max} - 1.5 \sim m_{\max}^{obs}]$ for the computation of LLH (Fig. 5a). Weights are calculated in this paper separately for each source zone by considering the observed events within magnitude range $[\hat{m}_{\max} - 1.5 \sim m_{\max}^{obs}]$. Four possible combinations are also considered for each source zone, for example, two sets of recurrence parameters $[(a1, b1), (a2, b2)$ or $\beta_1, \beta_2]$ and two sets of maximum magnitude $[\hat{m}_{\max}^{1}$ and $\hat{m}_{\max}^{2}]$. Three alternatives for maximum magnitudes are discussed later in Sect. 4.4. However, the third maximum magnitude is dropped here since it is dependent on the fault length and, therefore, has an unreasonably higher order of complexity. Final weights for both the magnitude distributions are taken as the weighted average (by number of events) over all the source zones and all possible combinations.
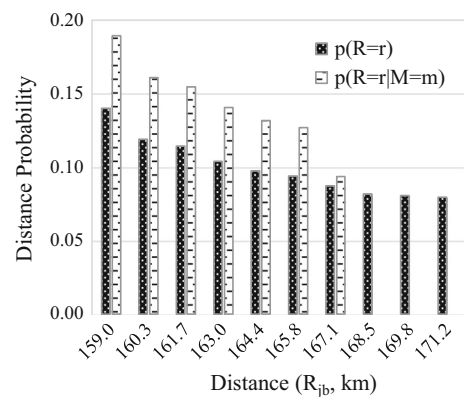

Figure 6
Fault rupture model

Figure 7
Source-to-site distance probability distribution (arbitrary source and site, Mw = 6.4)

4.3. Distance Probability Distribution

Distribution of source-to-site distance also contributes to the epistemic uncertainty. While assuming all the points on a fault plane are equiprobable to rupture (Iyengar & Ghosh, 2004), two alternatives are considered in this paper as the possible contributors to epistemic uncertainty: the distribution of source-to-site distance is independent or dependent on magnitude of the event. Magnitude independent probability distribution is based on the point-source model, and hence, rupture length of the fault due to

All the faults are assumed to be linear with known length in the study region and R_{jb} is considered as the measure of source-to-site distance (R) in this paper. $P(R|M = m)$ due to a rupture segment uniformly distributed over the fault plane was reported by Der Kiureghian and Ang (1977) and simplified by Iyengar and Ghosh (2004) for the hypocentral distance. The same model is extended in this paper to account for R_{jb} as the distance metric. While Fig. 6 illustrates the conceptual development of CDF for a fault located at one side of the site, the mathematical model is given by

$$P(R < r|M = m) = \left. \begin{array}{ll} P(R < r|M = m) = & 0 \quad for \quad r < \left(L_0^2 + D^2\right)^{0.5} \\ \dfrac{\left(r^2 - D^2\right)^{0.5} - L_0}{L_f - X(m)} \quad for \quad \left(L_0^2 + D^2\right)^{0.5} < r < \left[D^2 + \{L_f + L_0 - X(m)\}^2\right]^{0.5} \\ P(R < r|M = m) = & 1 \quad for \quad r > \left[D^2 + \{L_f + L_0 - X(m)\}^2\right]^{0.5} \end{array} \right\}. \tag{28}$$

causal magnitude for the earthquake is ignored. The magnitude-dependent model accounts for the finite rupture length, and hence, the distribution is conditioned to the magnitude of event.

Here, $X(m)$ is the rupture length in km given by Wells and Coppersmith (1994) for different fault types but not exceeding the fault length:

$$X(m) = \begin{cases} \mathrm{Min}\left(L_f, 10^{-3.22 + 0.69m}\right) & \mathrm{Min}\left(L_f, 10^{-3.55 + 0.74m}\right) & \mathrm{Min}\left(L_f, 10^{-2.01 + 0.50m}\right) & \mathrm{Min}\left(L_f, 10^{-2.86 + 0.63m}\right) \\ \text{Undefined} & \text{Strike - slip} & \text{Normal} & \text{Reverse} \end{cases}. \tag{29}$$

Other recent recommendations may also be used in place of Wells and Coppersmith (1994) leading to another source of epistemic uncertainty, which, however, is not considered in this paper for simplicity.

For a general case with the fault located at both sides of the site, conditional probabilities for the segments on each side are calculated separately and added with a weight as the fraction of total fault length. The same formulation also applies to the magnitude-independent model with $X(m) = 0$. Sample comparison of both probability density functions (PDF), i.e., $p(r)$ and $p(r|m)$, is presented in Fig. 7 for an arbitrary source and site in the study region with moment magnitude M_w 6.4.

4.3.1 Weight Calculation for the Type of Distance Probability Distribution

Proposed procedure for weight calculation of distance probability distribution is slightly complicated. However, it is also based on the log-likelihood approach. Let $M = \{m_i\}$, with $i = 1, 2, 3, \cdots, n$ be the sample data set of observed events (seismic catalogue) in the

unnecessary computation. Area of influence is defined in this paper as the maximum radial distance of 300 km around a site under the assumption that an event above certain magnitude ($m_{\min}$) triggering outside it would cause negligible effect to the structure located at the site.

Each observed event with the source assigned will have a total of 24 possible combinations of calculating the PDF: 12 combinations each for $g_j(r_i, m_i) = p(r_i)f(m_i)dm_i dr_i$ and $g_j(r_i, m_i) = p(r_i|m_i)f(m_i)dm_i dr_i$. For example, three alternatives for maximum magnitude (Sect. 4.4), two for recurrence parameters and two for magnitude distribution. Hence, three probability density functions, namely, $p(r)$ and $p(r|m)$ (as above) and $f(m)$ (as in Sect. 4.2), are computed in each of the 24 possible combinations and evaluated for each observed event with a given $m - r$ pair within the area of influence for each site followed by the computation of log-likelihood. This step is repeated for all the site locations considered (2302 locations with 0.1° grid spacing) in the study region followed by an estimate of the sample size independent average log-likelihood as follows:

Here, n_k is the total number of events within the

$$
\left.
\begin{aligned}
\text{LLH}_{g(r)} &= -\frac{1}{N_{sc}} \left[\sum_{k=1}^{2302} \sum_{j=1}^{12} \sum_{i=1}^{n_k} log_B\{g_j(r_i, m_i)\} \right] \quad \text{with} \quad g_j(r_i, m_i) = p(r_i)f(m_i)dm_i dr_i \\
\text{LLH}_{g(r|m)} &= -\frac{1}{N_{sc}} \left[\sum_{k=1}^{2302} \sum_{j=13}^{24} \sum_{i=1}^{n_k} log_B\{g_j(r_i, m_i)\} \right] \quad \text{with} \quad g_j(r_i, m_i) = p(r_i|m_i)f(m_i)dm_i dr_i
\end{aligned}
\right\}. \tag{30}
$$

magnitude interval $\text{Min}\left(\hat{m}_{\max}^1, \hat{m}_{\max}^2\right) - 1.5 \sim m_{\max}^{obs}$ in the complete study region including all the source zones. Each observed event must be assigned to a particular source (fault) since the distance probability distribution depends on relative location of source and site. With limited information available, one may assign an event to the nearest fault or a few faults (say 3 or 5) nearby. The former approach is adopted throughout this paper to avoid subjectivity. It is desirable to consider only the events occurring within the area of influence around a particular site to avoid

influence area of kth site, $p(r)$ and $p(r|m)$ are magnitude independent and dependent distance probabilities, respectively, and calculated using the CDF given by Eq. (28), whereas $f(m)$ is the magnitude probability density function as defined in Eqs. (22) and (23). Further, $1/N_{sc}$ is used for a sample size independent estimate with N_{sc} as total sample count ($2302 \times 12 \times n_k$) and $B = 2$ is taken as the base of logarithm. Associated weights are calculated as:

Table 6

Maximum magnitude for all the source zones by different methods

Zone	$m_{\max}^{obs}$	Kijko and Singh (2011)					Max (M1:M5)	Gupta (2002)	Max $\left(\hat{m}_{\max}^1, \ \hat{m}_{\max}^2\right)$
		M1	M2	M3	M4	M5	$\hat{m}_{\max}^1$	$\hat{m}_{\max}^2$	$\hat{m}_{\max}^3$
SZ1	8.20	8.35	8.35	8.35	8.35	8.31	8.35	8.70	8.70
SZ2	8.00	8.85	9.50	8.81	9.32	8.12	9.50	8.50	9.50
ACR1	8.00	8.16	8.16	8.15	8.16	8.12	8.16	8.50	8.50
ACR2	8.60	9.03	9.09	9.02	9.08	8.91	9.09	9.10	9.10
ACR3	8.10	8.66	8.79	8.63	8.74	8.34	8.79	8.60	8.79
ACR4	8.30	8.77	8.84	8.74	8.80	8.49	8.84	8.80	8.84
ACR5	7.50	7.65	7.65	7.64	7.64	7.70	7.70	8.00	8.00

$$W_{g(r)} = \frac{B^{-LLH_{g(r)}}}{B^{-LLH_{g(r)}} + B^{-LLH_{g(r|m)}}}, \quad W_{g(r|m)} = \frac{B^{-LLH_{g(r|m)}}}{B^{-LLH_{g(r)}} + B^{-LLH_{g(r|m)}}}. \tag{31}$$

4.4. Maximum Magnitude

Opinion differs while considering the maximum magnitude in PSHA, and several terminologies are interchangeably used in the prior art, for example, maximum characteristic, expected or possible earthquake magnitude. However, writing a simple terminology such as 'maximum magnitude' is often preferred followed by explaining the way of defining it (Abrahamson, 2000). Seven possible alternatives (five alternatives from Gupta, 2002; Kijko & Singh, 2011; and proportional to fault length) are initially selected for the estimation of maximum magnitude. Since five alternatives from Kijko and Singh, (2011) result in nearly identical estimation, the maximum of these five is considered a single alternative. Finally, three possible alternatives (maximum of five alternatives from Gupta, 2002; Kijko & Singh, 2011; and proportional to fault length) are considered in this paper as the contributors to epistemic uncertainty as explained below.

A. **Maximum of five methods from** Kijko and Singh (2011)

Several procedures for maximum magnitude estimation (statistically) are well explained and summarized by Kijko and Singh (2011), out of which following five methods are selected in this paper:

(a) Tate–Pisarenko:

$$\hat{m}_{\max} = m_{\max}^{obs} + \frac{1 - e^{-\hat{\beta}(\hat{m}_{\max} - m_{\min})}}{\hat{N}\hat{\beta} e^{-\hat{\beta}\left(m_{\max}^{obs} - m_{\min}\right)}}. \tag{32}$$

(b) Kijko–Sellevoll:

$$\hat{m}_{\max} = m_{\max}^{obs} + \int_{m_{\min}}^{m_{\max}} \left[\frac{1 - e^{-\hat{\beta}(m - m_{\min})}}{1 - e^{-\hat{\beta}(\hat{m}_{\max} - m_{\min})}} \right]^{\hat{N}} dm. \tag{33}$$

(c) Tate–Pisarenko–Bayes:

$$\hat{m}_{\max} = m_{\max}^{obs} + \frac{1}{\hat{N}\hat{\beta}C_\beta} \left(\frac{p_G}{p_G + m_{\max}^{obs} - m_{\min}} \right)^{-(q_G+1)}$$

$$C_\beta = \frac{1}{1 - [p_G/(p_G + \hat{m}_{\max} - m_{\min})]^{q_G}}; p_G$$

$$= \frac{\hat{\beta}}{\left(\sigma_{\hat{\beta}}\right)^2}; q_G = \left(\frac{\hat{\beta}}{\sigma_{\hat{\beta}}}\right)^2; \sigma_{\hat{\beta}} = \frac{\hat{\beta}}{\sqrt{N}}. \tag{34}$$

(d) Kijko–Sellevoll–Bayes:

$$+ \left(C_\beta\right)^{\hat{N}} \int_{m_{\min}}^{m_{\max}} \left[1 - \{p_G/(p_G + m - m_{\min})\}^{q_G}\right]^{\hat{N}} dm \tag{35}$$

(e) Non-parametric with Gaussian kernel

After arranging observed magnitudes in ascending order,

$$\hat{m}_{max} = m_{max}^{obs} + \sum_{i=1}^{N_c-1} \left(i/N_c\right)^{N_c} (m_{i+1} - m_i). \quad (36)$$

Here, $\hat{\beta}$ and $\hat{N}$ are given by Eqs (18) and (20), respectively, and N_c is the total number of events observed between minimum magnitude ($m_{min} = 4$) to maximum observed magnitude (m_{max}^{obs}). Maximum estimated magnitude ($\hat{m}_{max}$) appears on both sides of the governing expression in the first four methods [(a) to (d)], and hence, an iterative scheme is adopted with initial guess as $\hat{m}_{max} = m_{max}^{obs}$. Care should be exercised while implementing these methods: any one or even more of these methods may not necessarily converge owing to the incompleteness or insufficient number of events in some catalogues leading to unrealistic estimates of maximum magnitude even exceeding 10 and such cases should be discarded.

B. **Maximum observed magnitude + 0.5** (Gupta, 2002)

This is the simplest method recommended by Gupta (2002) for the maximum magnitude of a source zone:

$$\hat{m}_{max} = m_{max}^{obs} + 0.5 \quad (37)$$

III. Proportional to fault length

Maximum magnitude of earthquake depends on fault area (Wyss, 1979) and hence may be considered proportional to the fault length in some sense. Maximum magnitude for the jth fault of length L_f^j is proposed in this paper as the third alternative:

$$m_{max}^j = \hat{m}_{max}^3 - m_0 \left(1 - L_f^j / L_f^{max}\right); \quad m_0$$
$$= \hat{m}_{max}^3 \left[L_f^{median} / L_f^{max}\right]. \quad (38)$$

Here, $\hat{m}_{max}^3$ is the larger of the above two methods; L_f^{max} and L_f^{median} are maximum and median length of all the faults in a source zone. Maximum magnitudes estimated by the first and second methods considered here are summarized in Table 6 along with larger of the two. Note that the third method allows different maximum magnitudes in the faults located even within the same source zone and hence is not included in Table 6.

4.4.1 Weight Calculation for Maximum Magnitude

Weight associated with each alternative of maximum magnitude may be calculated in a similar way as explained with the distance probability model. A set of 12 possible combinations (four for each alternative of maximum magnitude, i.e., two for magnitude distributions with two for recurrence parameters) are considered. The framework is applied over the entire study area considering the events within the magnitude interval $Min(\hat{m}_{max}^1, \hat{m}_{max}^2) - 1.5 \sim m_{max}^{obs}$ for respective source zones. Log-likelihood is estimated as follows:

$$LLH_{f(m|\hat{m}_{max}^l)}$$
$$= -\frac{1}{N_{sc}} \left[\sum_{j=1}^{2} \sum_{k=1}^{2} \sum_{i=1}^{n_e} log_B \left\{ f_j\left(m_i|\hat{m}_{max}^l, \beta_k\right) dm_i \right\} \right] ;$$
with $l = 1, 2, 3$.

$$(39)$$

Here, n_e is the number of events within the magnitude range considered and $N_{sc} = 2 \times 2 \times n_e$ denotes the total sample count. Weights associated with each alternative of maximum magnitude is given by

$$W_{\hat{m}_{max}^l} = \frac{B^{-LLH_{f(m|\hat{m}_{max}^l)}}}{\sum_{l=1}^{3} B^{-LLH_{f(m|\hat{m}_{max}^l)}}}. \quad (40)$$

Sample illustration of an MHC contributed from a typical branch of logic tree and its weight calculation is explained in the next section.

5. Typical Model Hazard Curve and Weight Calculation

Construction of a typical model hazard curve with weight calculation is explained in the Sect. 2. This section presents one sample illustration for better clarity.

5.1. Model Hazard Curve

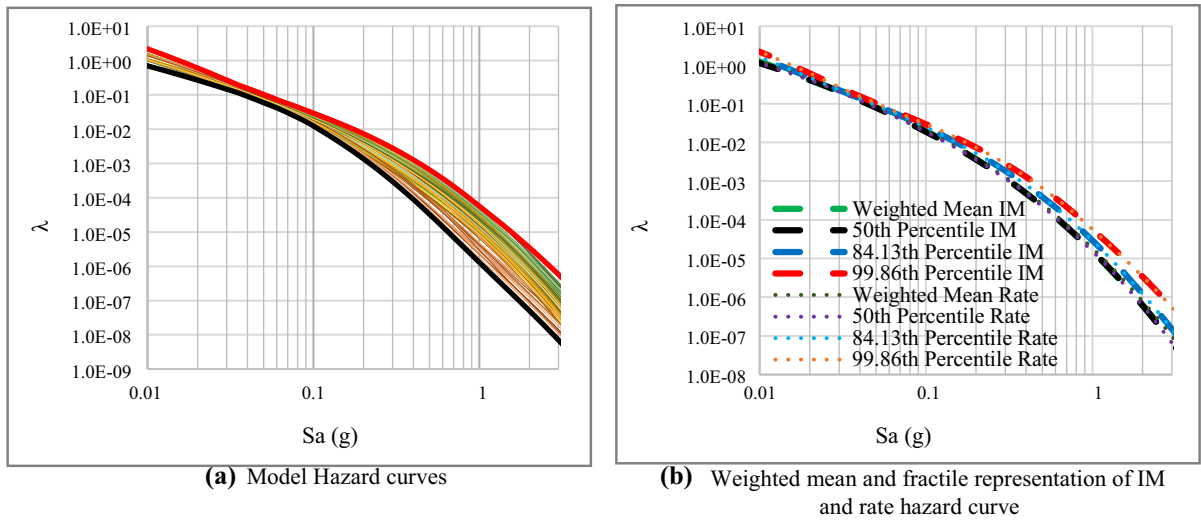Hazard integral or PSHA computational framework in terms of discrete summations can be expressed as (Kramer, 1996)

(a) Model Hazard curves

(b) Weighted mean and fractile representation of IM and rate hazard curve

Figure 8
Hazard curves at Site class B, $T = 0$ for Champhai (Mizoram)

(a) Variation in the ratio of exceedance frequencies
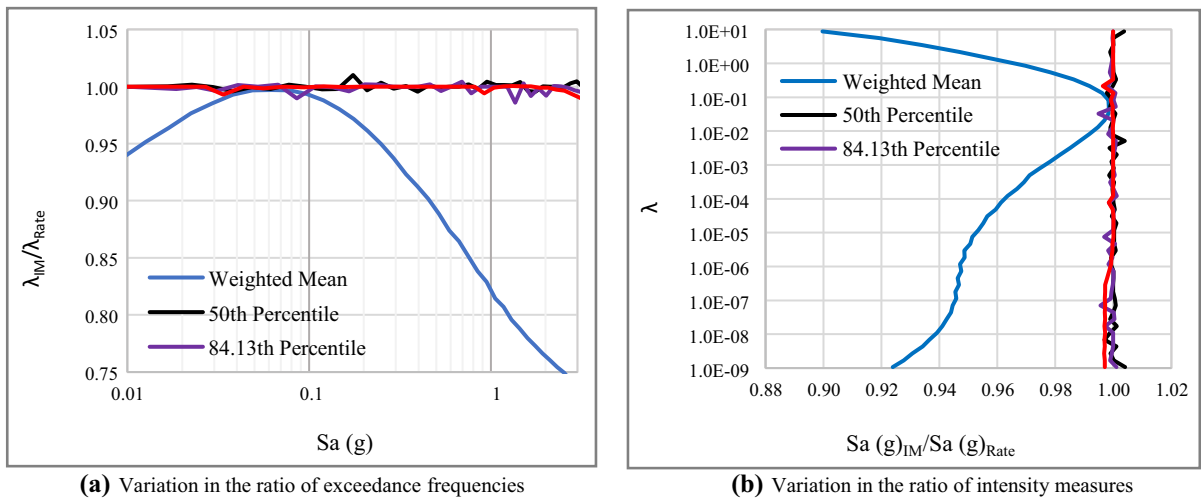

(b) Variation in the ratio of intensity measures

Figure 9
Difference between weighted mean and fractile representation of hazard curves due to horizontal and vertical dissection at Site class B, $T = 0$ for Champhai (Mizoram)

$$\lambda\left(\ln Sa(T^*) > \ln Sa^*\right)$$

$$= \sum_{i=1}^{N_S} \sum_{j=1}^{N_M} \sum_{k=1}^{N_R} v_i(m_{\min}) P\left(\ln Sa(T^*) > \ln Sa^* \big| m_j, r_k\right)$$

$$f\left(m_j, r_k\right) \Delta m \Delta r. \tag{41}$$

Here, N_S represents the number of potential earthquake sources in a region around the site; N_M and N_R are the number of discrete magnitude and distance segments considered in the ith potential earthquake source; $v_i(m_{\min})$ is the mean annual rate of exceedance of events with magnitude greater than $m_{\min}$ in the ith source; $P\left(\ln Sa(T^*) > \ln Sa^* \big| m_j, r_k\right)$ is given by GMPE for an earthquake of magnitude

(a) Model UHS **(b)** Weighted mean and fractile representation of UHS
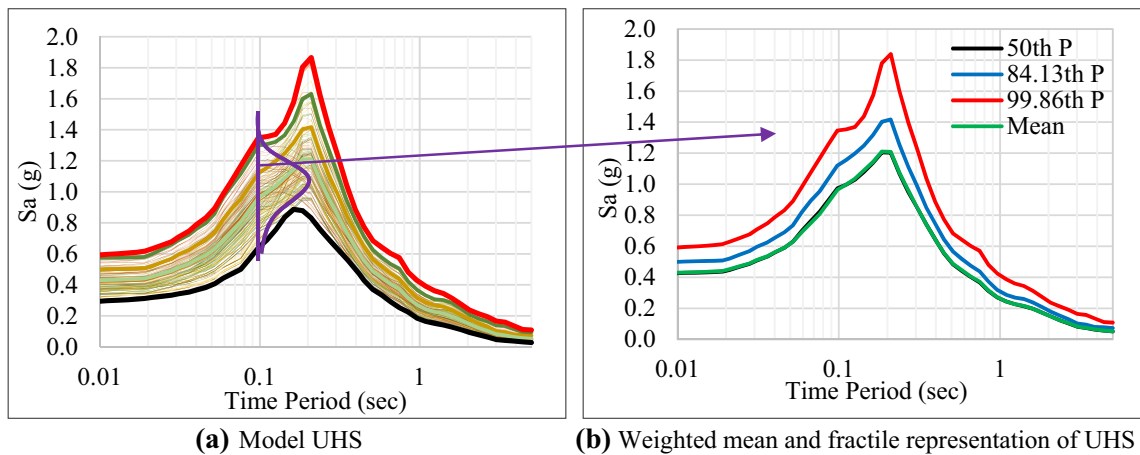
Figure 10
Uniform hazard spectra (UHS) at Site class B, 2%PE in 50 years, for Champhai (Mizoram)

$Mw = m_j$ triggered at a distance of $R_{jb} = r_k$; and $f(m_j, r_k)\Delta m \Delta r$ is the probability of triggering an earthquake of magnitude m_j at a distance of r_k. In the case of magnitude-independent distance probability distribution, $f(m_j, r_k) = f(m_j)p(r_k)$ with $f(m_j)$ and $p(r_k)$ as the probability density functions of magnitude and distance, respectively, given by Eq. (22)/ Eq. (23) and Eq. (28) with $X(m) = 0$, respectively. In the case of magnitude-dependent distance probability distribution, $p(r_k)$ is replaced by $p(r_k|m_j)$, which can be obtained from the first derivative of Eq. (28).

In this study, North-east India is divided into two types of source zones, namely, SZ and ACR, which are further subdivided into two and five source zones, respectively. Referring to Fig. 1a, one branch of the logic tree with the following details is considered here for sample illustration: (i) magnitude-independent distance probability distribution; (ii) recurrence parameters using straight line method; (iii) magnitude distribution by Gutenberg-Richter (1944); and (iv) maximum magnitude as m^1_{max}(Table 6). This explains one alternative for the first four nodes of the logic tree, which is defined as ERF (Sect. 2.1). The selected ERF is further divided into a set of SERFs conforming to the GMPE rule (Sect. 2.2).

Rate of exceedance of an intensity measure given a threshold, $\lambda(\ln Sa(T^*) > \ln Sa^*)$ (denoted as λ hereafter for brevity) is first computed by summing up the contributions from all the potential sources in SZ around a particular site considering (i) ERF as above and (ii) BCHY16 (Table 1) as the GMPE. Next, λ contributed by the potential sources of ACR with same ERF and all four GMPEs (in ACR) are computed separately followed by taking a weighted (Table 1) sum. Total exceedance rate is given by the sum of the two computed above. Repeating the same procedure for different values of threshold intensity measure $(\ln Sa^*)$ enables the construction of one MHC corresponding to a single terminal branch of the logic tree.

Further, GMPEs considered in this computation if uses the source-to-site distance other than R_{jb}, a conversion of distance metric is performed prior to using the GMPE: R_{jb} is converted into (i) R_{rup} and R_{hypo} using the algorithm of Kaklamanos et al. (2011); and (ii) R_{epi} using the recommendation of Tavakoli et al. (2018).

5.2. Weight Calculation

Weight associated with each MHC (w_{MHC}) is required for the computation of weighted mean representation of rate hazard curve. With reference to Fig. 1, for the same MHC considered above, $w_{ERF} = 0.43 \times 0.46 \times 0.55 \times 0.339 = 0.0368$ and $w_{SERF} = 0.28 \times 0.51 = 0.1428$. Therefore, weight of the MHC is computed as $w_{MHC} = 0.0368 \times 0.1428 = 5.266 \times 10^{-3}$.

6. Results and Discussion

PSHA is performed in the companion paper using the logic tree approach to consider the possible effects of epistemic uncertainty. The sources of epistemic uncertainty considered are the (i) recurrence parameters, (ii) magnitude distribution, (iii) distance probability distribution, (iv) maximum magnitude and (v) selection of GMPEs. PSHA is carried out for three NEHRP soil sites of class B (V_{s30} 760–1500 m/s), C (V_{s30} 360–760 m/s) and D (V_{s30} 180–360 m/s) with an average shear wave velocity of 1100, 550 and 250 m/s, respectively. Only one sample illustration of PSHA results out of 112 district headquarters of seven states in North-east India is reported in this paper in terms of the weighted mean and fractile representation of hazard curves and UHS for brevity (50th, 84.13th and 99.86th percentile). Computation of weighted mean and fractile representation of hazards may lead to an alternate perspective of 'importance factor' used in seismic design. As opposed to a structure specific decision, importance factor is likely to be dependent on the natural period and seismicity of the location as well. Further exploration of PSHA results and the 'importance factor' is reported in the companion paper.

6.1. Weighted Mean and Fractile Representation of Hazard Curves

A typical MHC represents the mean annual rate of exceedance of a selected intensity measure at a particular site. Several MHCs can be generated from various terminal branches of the logic tree as a possible alternative. A sample illustration of the MHCs (peak ground acceleration, PGA, $T = 0$) at site class B for Champhai district in Mizoram is presented in Fig. 8a as an outcome from different branches of the logic tree. Thick outer lines indicate the spread of the hazard curves. Figure 8b represents the weighted mean and different fractile estimates of the hazard curves illustrated in Fig. 8a. Both IM and rate hazard curves resulting from horizontal and vertical dissections, respectively, are included in Fig. 8b.

Variation in the ratio of mean annual rate of exceedance of IM to rate hazard curves (Fig. 8b) with respect to IM is presented in Fig. 9a. Similarly, Fig. 9b presents the variation in the ratio of intensity measures from IM to rate hazard curves (Fig. 8b) with respect to the mean annual rate of exceedance. As evident from the Figs. 9a and b, the choice of IM or rate hazard does not affect the fractile representation of MHC. Similar comparison is also included in Fig. 9 for the weighted mean representation of MHC which is somewhat sensitive to the definition of IM vs. rate hazard. Hence, care should be exercised while selecting and adopting the weighted mean representation of hazard, which is governed by the end objective such as characterization of seismic hazard and estimation of collapse fragility.

The end objective in this paper is to estimate the design seismic hazard, and hence, the horizontal dissection is chosen with IM as a random variable conditioned to a mean annual rate of exceedance. Therefore, weighted mean and any percentile/fractile representation of IM will be referred to as weighted mean and percentile/fractile representation of hazard (dropping IM) hereafter for brevity.

6.2. Uniform Hazard Spectra

Given one terminal branch of the logic tree, the hazard curves are constructed using spectral ordinates at different periods as the IM. The spectral ordinates, when plotted against the time periods conditioned to a mean annual rate of exceedance, result in the UHS associated with the considered terminal branch of the logic tree. Similarly, UHS associated with all branches of the logic tree are constructed. Figure 10a presents a sample illustration of 2475 years UHS for the Champhai district in Mizoram with site class B while thick outer lines indicate the spread. Figure 10b presents the associated weighted mean and different fractile estimates (by vertical dissection) of the UHS. Similar studies are carried out in the companion paper for all of North-east India for all three site classes (B, C and D) and at both MCE and DBE levels.

6.3. Results Presented in the Companion Paper

This paper presents the framework for conducting PSHA considering possible sources of epistemic

uncertainty using the logic tree approach. The framework is implemented over the north-east region of India. Only one sample illustration is presented here to explain the application of the proposed framework. Detailed comparison of results with previous studies and Indian standard recommendations, in terms of weighted mean representation of PGA hazard and UHS for NEHRP soil types B, C and D at DBE are presented in the companion paper. Further, a hazard map of the complete region is prepared by dividing the area in a grid spacing of 0.1° (about 10 km). Two return periods (MCE and DBE) with three soil types are considered and maps are presented in Appendix 2 of the companion paper. Ratio of fractile-to-weighted mean representation of hazard is presented as an alternative viewpoint of importance factor used in the seismic design. Weighted mean and fractile representation of target conditional spectra (CS) with different conditional time periods (0.2–1.6 s in increments of 0.2 s) are also estimated by two methods (GCR-CS and GCIM-CS, considering GMPE specific generalized causal rupture and all contributing causal ruptures, respectively) using the logic tree approach and results are compared. Weighted mean representation of target CS (mean and standard deviation) are calculated at 112 district headquarters, but only mean spectra (without standard deviation) are reported in the Appendix 3 of the companion paper.

developed for computing the weights associated with the MHCs. The weighted mean representation of hazard is calculated by taking weighted sum of IM as outcome from various branches of the logic tree. Similarly, fractile estimate is used to construct the fractile representation of hazard.

Following are the key contributions/conclusions of this paper based on the limited investigation carried out:

1. Relevant guidelines are proposed for the weight calculation at each node of the logic tree (except for GMPEs) to minimize the subjectivity based on the degree of belief of the analyst.
2. A GMPE rule is proposed to account for the propagation of epistemic uncertainty.
3. Characterization of seismic hazard in terms of weighted mean and fractile representation of hazard with a new interpretation of importance factor in seismic design is envisioned.
4. Horizontal or vertical dissection of the hazard curves should be determined based on the end objective. While the fractile representation of hazard remains insensitive, weighted mean representation of hazard is somewhat sensitive to the choice of dissection.

Some of the contributors of epistemic uncertainty such as Vs30 and fault rupture characteristics are not included in the proposed framework.

7. Conclusion

A framework is developed for PSHA to account for the contribution from possible sources of epistemic uncertainty that includes (i) recurrence parameters, (ii) magnitude distribution, (iii) distance probability distribution, (iv) maximum magnitude and (v) selection of GMPEs. The log-likelihood approach is employed to arrive at the weights associated with various contributors of epistemic uncertainty. A somewhat different approach is applied for the recurrence parameters and weights are assumed to be inversely proportional to the mean squared error. A GMPE rule is proposed to be used with the PSHA framework to account for the propagation of epistemic uncertainty. A procedure is also

Acknowledgements

Center for Seismology, India; and IRIS Data management centre, New York, USA.

Funding

Availability of Data and Material

Data may be available from the corresponding author through making reasonable request.

Code Availability

Custom code is developed in MATLAB environment and not available for sharing.

Declarations

Conflict of interest Authors declare that the manuscript is free from any Conflicts of interest and Competing interests.

REFERENCES

Abrahamson, Norman A., (2000). State of The Practice of Seismic Hazard Evaluation. Paper presented at the ISRM International Symposium, Melbourne, Australia

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

Abrahamson, N., Gregor, N., & Addo, K. (2016). BC Hydro ground motion prediction equations for subduction earthquakes. *Earthquake Spectra, 32*(1), 23–44.

Abrahamson, N. A., Silva, W. J., & Kamai, R. (2014). Summary of the ASK14 ground motion relation for active crustal regions. *Earthquake Spectra, 30*(3), 1025–1055.

Akkar, S., Sandıkkaya, M. A., & Bommer, J. J. (2014). Empirical ground-motion models for point-and extended-source crustal earthquake scenarios in Europe and the Middle East. *Bulletin of Earthquake Engineering, 12*(1), 359–387.

Anbazhagan, P., Bajaj, K., Matharu, K., Moustafa, S. S. R., & Al-Arifi, N. S. N. (2019). Probabilistic seismic hazard analysis using the logic tree approach—Patna district (India). *Natural Hazards and Earth Systems Sciences, 19*, 2097–2115.

Anbazhagan, P., Kumar, A., & Sitharam, T. G. (2013). Ground motion prediction equation considering combined dataset of recorded and simulated ground motions. *Soil Dynamics and Earthquake Engineering, 53*, 92–108.

Anbazhagan, P., Sreenivas, M., Ketan, B., Moustafa, S. S. R., & Al-Arifi, N. S. N. (2016). Selection of ground motion prediction equations for seismic hazard analysis of Peninsular India. *Journal of Earthquake Engineering, 20*, 699–737.

Anderson, J. G. (2018). Quantifying the epistemic uncertainty in the probabilistic seismic hazard from two major faults in Western Nevada. *Earthquake Spectra, 34*, 549–568.

Arroyo, D., García, D., Ordaz, M., Mora, M. A., & Singh, S. K. (2010). Strong ground-motion relations for Mexican interplate earthquakes. *Journal of Seismology, 14*(4), 769–785.

Atkinson, G. M., Bommer, J. J., & Abrahamson, N. A. (2014). Alternative approaches to modeling epistemic uncertainty in ground motions in probabilistic seismic-hazard analysis. *Seismological Research Letters, 85*(6), 1141–1144.

Atkinson, G. M., & 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.

Bommer, J. J., Douglas, J., Scherbaum, F., Cotton, F., Bungum, H., & 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., & Scherbaum, F. (2008). The use and misuse of logic trees in probabilistic seismic hazard analysis. *Earthquake Spectra, 24*(4), 997–1009.

Bommer, J. J., Scherbaum, F., Bungum, H., Cotton, F., Sabetta, F., & Abrahamson, N. A. (2005). On the use of logic trees for ground motion prediction equations in seismic hazard assessment. *Bulletin of the Seismological Society of America, 95*(2), 377–389.

Boore, D. M., Stewart, J. P., Seyhan, E., & Atkinson, G. M. (2014). NGA-West2 equations for predicting PGA, PGV, and 5 damped PSA for shallow crustal earthquakes. *Earthquake Spectra, 30*(3), 1057–1085.

Bradley, B. A. (2009). Seismic hazard epistemic uncertainty in the San Francisco bay area and its role in performance-based assessment. *Earthquake Spectra, 25*(4), 733–753.

Campbell, K. W., & Bozorgnia, Y. (2014). NGA-West2 ground motion model for the average horizontal components of PGA, PGV, and 5 damped linear acceleration response spectra. *Earthquake Spectra, 30*(3), 1087–1115.

Cauzzi, C., Faccioli, E., Vanini, M., & Bianchini, A. (2015). Updated predictive equations for broadband (0.01–10 s) horizontal response spectra and peak ground motions, based on a global dataset of digital acceleration records. *Bull Earthquake Engineering, 13*(6), 1587–1612.

Chiou, B.S.-J., & Youngs, R. R. (2014). Update of the Chiou and Youngs NGA model for the average horizontal component of peak ground motion and response spectra. *Earthquake Spectra, 30*(3), 1117–1153.

Cotton, F., Scherbaum, F., Bommer, J. J., & Bungum, H. (2006). Criteria for selecting and adjusting ground-motion models for specific target applications: applications to Central Europe and rock sites. *Journal of Seismology, 10*(2), 137–156.

Delavaud, E., Cotton, F., Akkar, S., Scherbaum, F., Danciu, L., Beauval, C., Drouet, S., Douglas, J., Basili, R., Sandikkaya, M. A., Segou, M., Faccioli, E., & Theodoulidis, N. (2012). Toward a ground-motion logic tree for probabilistic seismic hazard assessment in Europe. *Journal of Seismology, 16*, 451–473.

Der Kiureghian, A., & Ang, A.H.-S. (1977). A fault rupture model for seismic risk analysis. *Bulletin of the Seismological Society of America, 67*, 1173–1194.

Der Kiureghian, A., & Ditlevsen, O. (2009). Aleatory or epistemic? Does it matter? *Structural Safety, 31*(2), 105–112.

Douglas, J., (2020). Ground motion prediction equations 1964-2020. http://www.gmpe.org.uk. Assessed 1 Oct 2020.

Grünthal, G., Stromeyer, D., Bosse, C., Cotton, F., & Bindi, D. (2018). The probabilistic seismic hazard assessment of Germany—version 2016, considering the range of epistemic uncertainties and aleatory variability. *Bull Earthquake Eng, 16*, 4339–4395.

Gupta, I. D. (2002). The state-of-the-art in seismic hazard analysis. *ISET Journal Earthquake Technological, 39*(4), 311–346.

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

Gurjar, N., & Basu, D. (2022). Epistemic uncertainty in PSHA and seismic hazard characterization using logic tree approach: Part II-Implementation over North-East India. Pure and Applied Geophysics (Companion Paper-II; Accepted 3 September 2022)

Gutenberg, B., & Richter, C. F. (1944). Frequency of earthquakes in California. *Bulletin of the Seismological Society of America, 34*, 185–188.

Idriss, I. M. (2014). An NGA-West2 empirical model for estimating the horizontal spectral values generated by shallow crustal earthquakes. *Earthquake Spectra, 30*(3), 1155–1177.

Iyengar, R. N. and Ghosh, S., (2004). Seismic hazard mapping of Delhi city. 13th world conference on Earthquake Engineering Canada, Paper No. 180.

Kaklamanos, J., Baise, L. G., & Boore, D. M. (2011). Estimating unknown input parameters when implementing the NGA ground-motion prediction equations in engineering practice. *Earthquake Spectra, 27*, 1219–1235.

Kalakonas, P., Silva, V., Mouyiannou, A., & Rao, A. (2020). Exploring the impact of epistemic uncertainty on a regional probabilistic seismic risk assessment model. *Natural Hazards, 104*, 997–1020.

Kanno, T., Narita, A., Morikawa, N., Fujiwara, H., & 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.

Kijko, A., & Singh, M. (2011). Statistical tools for maximum possible earthquake magnitude estimation. *Acta Geophysica, 59*, 674–700.

Kijko, A., & Smit, A. (2012). Extension of the Aki-Utsu b-value estimator for incomplete catalogs. *Bulletin of the Seismological Society of America, 102*, 1283–1287.

Kramer, S. L. (1996). *Geotechnical earthquake engineering.* Prentice Hall.

Kulkarni, R. B., Youngs, R. R., and Coppersmith, K. J., (1984). Assessment of confidence intervals for results of seismic hazard analysis. In Proceedings, Eighth World Conference on Earthquake Engineering, vol. 1, San Francisco, pp. 263–270.

Lin, P. S., & 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.

Main, I. G., & Burton, P. W. (1984). Information theory and the earthquake frequency-magnitude distribution. *Bulletin of the Seismological Society of America, 74*, 1409–1426.

Marzocchi, W., Taroni, M., & Selva, J. (2015). Accounting for epistemic uncertainty in PSHA: Logic tree and ensemble modeling. *Bulletin of the Seismological Society of America, 105*(4), 2151–2159.

McGuire, R. K., Cornell, C. A., & Toro, G. R. (2005). The case for the mean hazard curve. *Earthquake Spectra, 21*, 879–886.

McVerry G. H., J. X. Zhao, N. A. Abrahamson, and P. G. Somerville, (2006). New Zealand acceleration response spectrum attenuation relations for crustal and subduction zone earthquakes. Bulletin of the New Zealand Society for Earthquake Engineering, vol. 39, no. 1, Art. No. 1.

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

Nath, S. K., Thingbaijam, K. K. S., Maiti, S. K., & Nayak, A. (2012). Ground-motion predictions in Shillong region, northeast India. *Journal of Seismology, 16*(3), 475–488.

NDMA, (2010). Development of probabilistic seismic hazard map of india. Technical report by National disaster management authority, Government of India, Delhi, India.

Özbey, C., Sari, A., Manuel, L., Erdik, M., & Fahjan, Y. (2004). An empirical attenuation relationship for Northwestern Turkey ground motion using a random effects approach. *Soil Dynamics and Earthquake Engineering, 24*(2), 115–125.

Petersen, Mark D., Frankel, Arthur D., Harmsen, Stephen C., Mueller, Charles S., Haller, Kathleen M., Wheeler, Russell L., Wesson, Robert L., Zeng, Yuehua, Boyd, Oliver S., Perkins, David M., Luco, Nicolas, Field, Edward H., Wills, Chris J., and Rukstales, Kenneth S., (2008). Documentation for the 2008 Update of the United States National Seismic Hazard Maps: U.S. Geological Survey Open-File Report–1128, 61 p.

Rodriguez-Marek, A., Bommer, J. J., Youngs, R. R., Crespo, M. J., Stafford, P. J., & Bahrampouri, M. (2021). Capturing epistemic uncertainty in site response. *Earthquake Spectra, 37*(2), 921–936.

Scherbaum, F., Delavaud, E., & Riggelsen, C. (2009). Model selection in seismic hazard analysis: An information-theoretic

perspective. *Bulletin of the Seismological Society of America, 99*(6), 3234–3247.

Senior Seismic Hazard Analysis Committee (SSHAC), (1997). Recommendations for Probabilistic Seismic Hazard Analysis: Guidance on Uncertainty and Use of Experts, U.S. Nuclear Regulatory Commission, U.S. Dept. of Energy, Electric Power Research Institute, NUREG/CR6372, UCRL-ID-122160, Vols. 1/2.

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

Stepp J. C., (1972). Analysis of completeness of the earthquake sample in the Puget Sound area and its effect on statistical estimates of earthquake hazard. In Proc. Of the 1st Int. Conf. on Microzonazion, Seattle, Vol. 2, pp. 897–910.

Stewart J.P., Douglas J., Javanbarg M.B., Di Alessandro C., Bozorgnia Y., Abrahamson N.A., Boore D.M., Campbell K.W., Delavaud E., Erdik M., and Stafford P.J. (2013). GEM-PEER Task 3 Project: Selection of a global set of ground motion prediction equations, PEER Report No. 2013/222, Pacific Earthquake Engineering Research Center, University of California, Berkeley, CA.

Tavakoli, B., Sedaghati, F., & Pezeshk, S. (2018). An analytical effective point-source-based distance-conversion approach to mimic the effects of extended faults on seismic hazard assessment. *Bulletin of the Seismological Society of America, 108*, 742–760.

Tromans, I. J., Aldama-Bustos, G., Douglas, J., Lessi-Cheimariou, A., Hunt, S., Daví, M., Musson, R. M. W., Garrard, G., Strasser, F. O., & Robertson, C. (2019). Probabilistic seismic hazard assessment for a new-build nuclear power plant site in the UK. *Bull Earthquake Engineering, 17*, 1–36.

USNRC, (2018). Updated implementation guidelines for SSHAC hazard studies. NUREG-2213, October 2018. Washington, DC: US Nuclear Regulatory Commission.

Wells, D. L., & 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, 1994*(84), 974–1002.

Wyss, M. (1979). Estimating maximum expectable magnitude of earthquakes from fault dimensions. *Geology, 7*(7), 336–340.

Yaghmaei-Sabegh, S., & Ebrahimi-Aghabagher, M. (2019). Quantification of source-to-site distance uncertainty in ground motion models. *Natural Hazards, 99*, 287–306.

Youngs, R. R., Chiou, S. J., Silva, W. J., & Humphrey, J. R. (1997). Strong ground motion attenuation relationships for subduction zone earthquakes. *Seismological Research Letters, 68*(1), 58–73.

Zhao John X., Fei Jiang, Pan Shi, Hao Xing, Haifeng Huang, Ruibin Hou, Yingbin Zhang, Pengcheng Yu, Xiaowen Lan, David A. Rhoades, Paul G. Somerville, Kojiro Irikura, Yoshimitsu Fukushima, (2016a). Ground-Motion Prediction Equations for Subduction Slab Earthquakes in Japan Using Site Class and Simple Geometric Attenuation Functions. Bulletin of the Seismological Society of America. 106 (4): 1535–1551.

Zhao, J. X., Zhang, J., Asano, A., Ohno, Y., Oouchi, T., Takahashi, T., Ogawa, H., Irikura, K., Thio, H. K., Somerville, P. G., Fukushima, Y., & Fukushima, Y. (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.

Zhao, J. X., Zhou, S., Zhou, J., Zhao, C., Zhang, H., Zhang, Y., Gao, P., Lan, X., Rhoades, D., Fukushima, Y., Somerville, P., & Irikura, K. (2016b). Ground-motion prediction equations for shallow crustal and upper-mantle earthquakes in Japan using site class and simple geometric attenuation functions. *Bulletin of the Seismological Society of America, 106*(4), 1552–1569.