SECOND ELICITATION ROUND: QUANTIFICATION OF THE WEIGHTS OF THE ALTERNATIVE MODELS

Expert's name: ____

Affiliation: ____

Author: TSUMAPS-NEAM TI team

Date: 23 August 2017

Version: 7.0

Table of Contents

Introduction

A diverse range of scientifically acceptable models should be considered for quantifying the epistemic uncertainty in this hazard assessment. Such models should be weighted according to their relevance in the specific application. Then, they will be finally used to quantify the "*informed community distribution*" describing the variability of the results that you would expect if you asked different experts within the technical community to perform the analysis (SSHAC 1997, USNRC 2012). The goal of this elicitation round is to quantify the model's weights.

You may recall that the adopted SPTHA is divided into 4 main STEPs, each of them subdivided in Levels, as reported in Figure 1. The 4 STEPs are the following:

- STEP 1: PROBABILISTIC EARTHQUAKE MODEL
- STEP 2: TSUNAMI GENERATION & MODELING IN DEEP WATER
- STEP 3: SHOALING AND INUNDATION
- STEP 4: HAZARD AGGREGATION & UNCERTAINTY QUANTIFICATION

More details may be found in the attached document "***Doc_P1_S1_ProjectSummary***", which, together with the other documents we provide you for completeness, has been also sent to a group of reviewers for comments.

In real applications, there is the practical necessity to consider only a limited set of alternative models. In TSUMAPS-NEAM, we first provided an extended list of models, forming altogether the "*tree of the alternatives*". Then, in order to reduce the total number of models to be implemented, the tree has been "trimmed" following the feedbacks that you provided through the questionnaire in the first elicitation round.

The results of the first elicitation round and the consequent development of the final tree of alternatives of TSUMAPS-NEAM are described in details in the attached documents "***DOC_P1_S3_Elicitation***" and "***DOC_P1_S4_Prel_Impl_Plan***", respectively. The resulting Alternative Tree is reported in **Figures 2** to **5**, divided in the 4 main STEPs of the analysis. As you can see, alternatives are foreseen only at STEP 1 and STEP 4, and a total of 16 models are to be weighted in this second elicitation round.

In order to combine into a single model quantifying the "informed community distribution", the alternative models of the final tree of alternatives should be weighted according to the credibility of each alternative model and its representativeness within the informed technical community (SSHAC 1997; Marzocchi et al. 2015). The quantification of weights is here based on your opinion as a component of the TSUMAPS-NEAM Pool of Experts (PoE), and it is the goal of this questionnaire.

As in the first elicitation, the group opinion will be here extracted by considering an Analytical Hierarchy Process (AHP). The AHP is a multi-criterion decision model introduced by Thomas Saaty (1980). The AHP takes as input integer numbers expressing the degree of preference of one model over another, ranging from one (equally important / appropriate) to nine (extremely more important / appropriate). These numbers can be seen as the nearest integer to the ratio of the weights assigned to each model. **Table 1** exhibits the information about the fundamental scale to be used in AHP for judgments in pairs for a given criterion adopted for the comparison. The scale enables the expert to incorporate experience and knowledge intuitively and indicates how many times a given model dominates the others with respect to the adopted criterion. Based on experts' judgments, the final scores (positive numerical values) are generated following the STEPs of the AHP, determining the score of each Level and sub-Level.

In this elicitation, compared to the first elicitation, a slightly more sophisticated approach will be adopted. We will use multiple criteria, instead of using just one single criterion. This implies that first, the alternative criteria will be defined and compared (in *Questionnaire Q0*). Then, each question on the alternative models will be repeated once for each criterion (from *Questionnaire Q1 on*). Here we use 2 criteria, so all questions will be posed twice.

Note that different potential quantitative criteria may be also defined further to those based on the PoE's opinions, as for example the performance of models in sanity-checks, statistical and sensitivity tests, etc. If possible, in the future several options will be tested, leading at least to 2 additional alternative methods for quantifying the weights of the alternative models.

In the following, we first present an example of AHP questionnaire (the same of the first elicitation round), to show how this questionnaire should be answered. Then, the main part of the questionnaire starts.

In Question #0, we will ask you to prioritize the criteria to be adopted for the comparison of models. Then, from Questions #1 on, you will be asked to prioritize specific alternative models, by adopting both the criteria defined in Question #0. In each of these questions, you are asked to fill two tables, one per criterion.

Please, review the illustrative example provided in the next pages. Then, to express your opinions, a total of 15 tables should be filled, relative to 8 questions: Q0, Q1, Q2a, Q2b, Q3a, Q3b, Q3c. Please note that each of these questions may contain multiple tables to be filled. In particular, Q0 contains 1 table, while all the other questions contain 2 tables.

Figure 1: Sketch of the workflow.

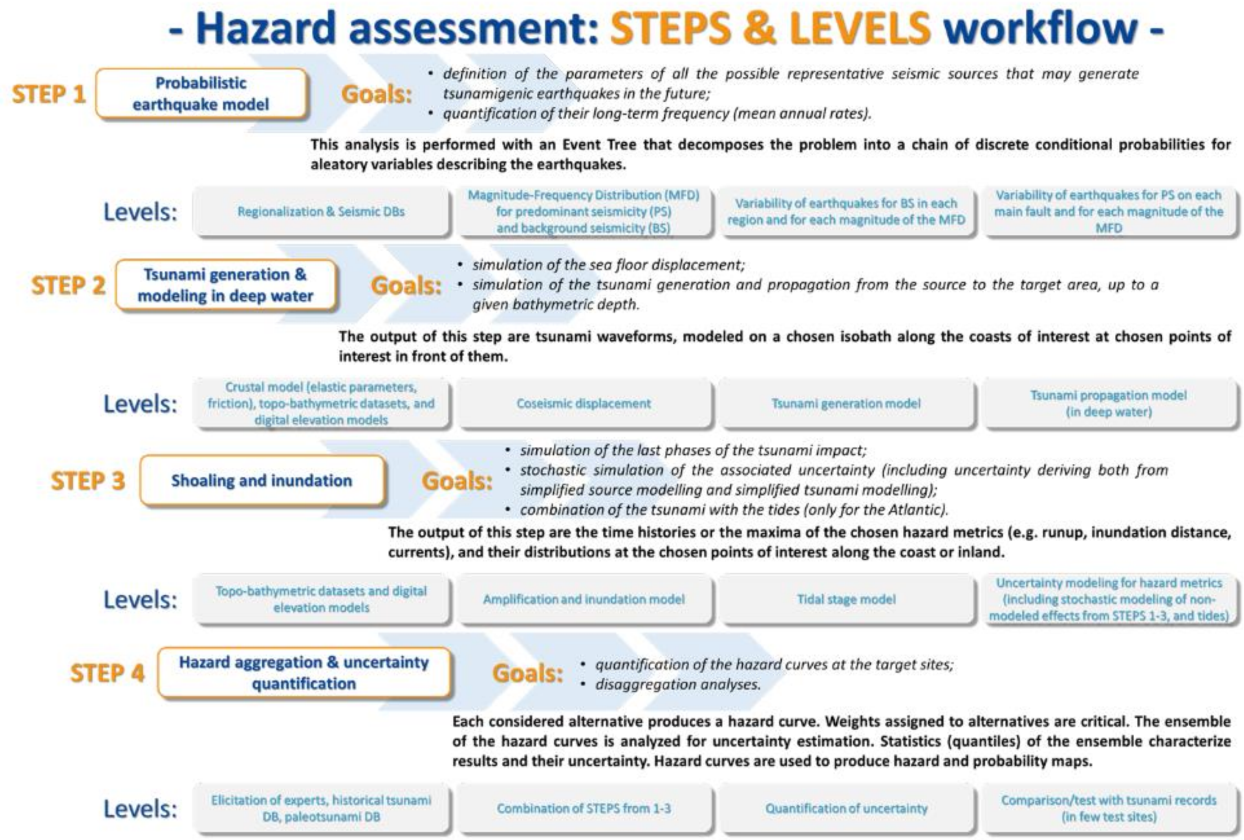
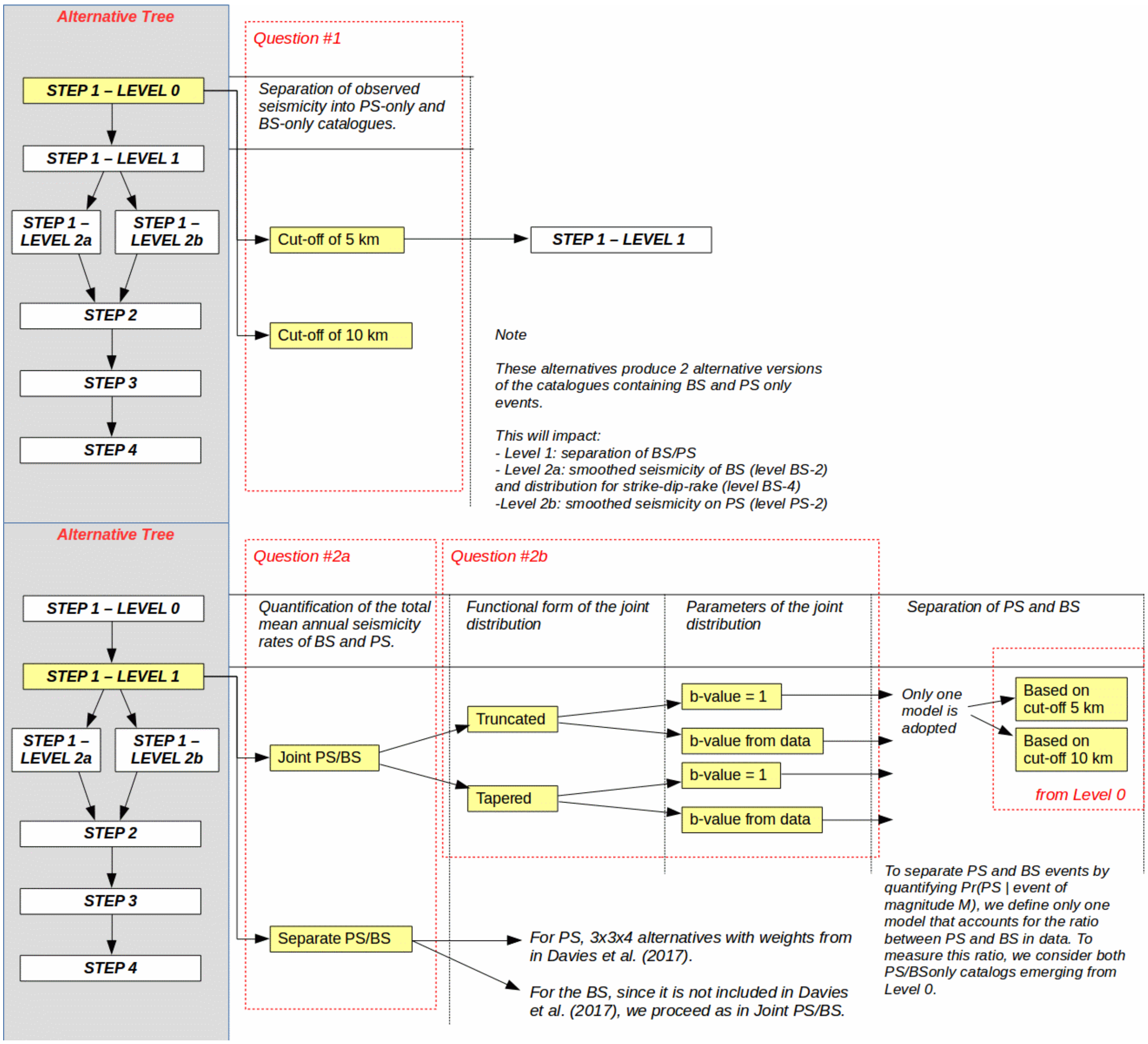
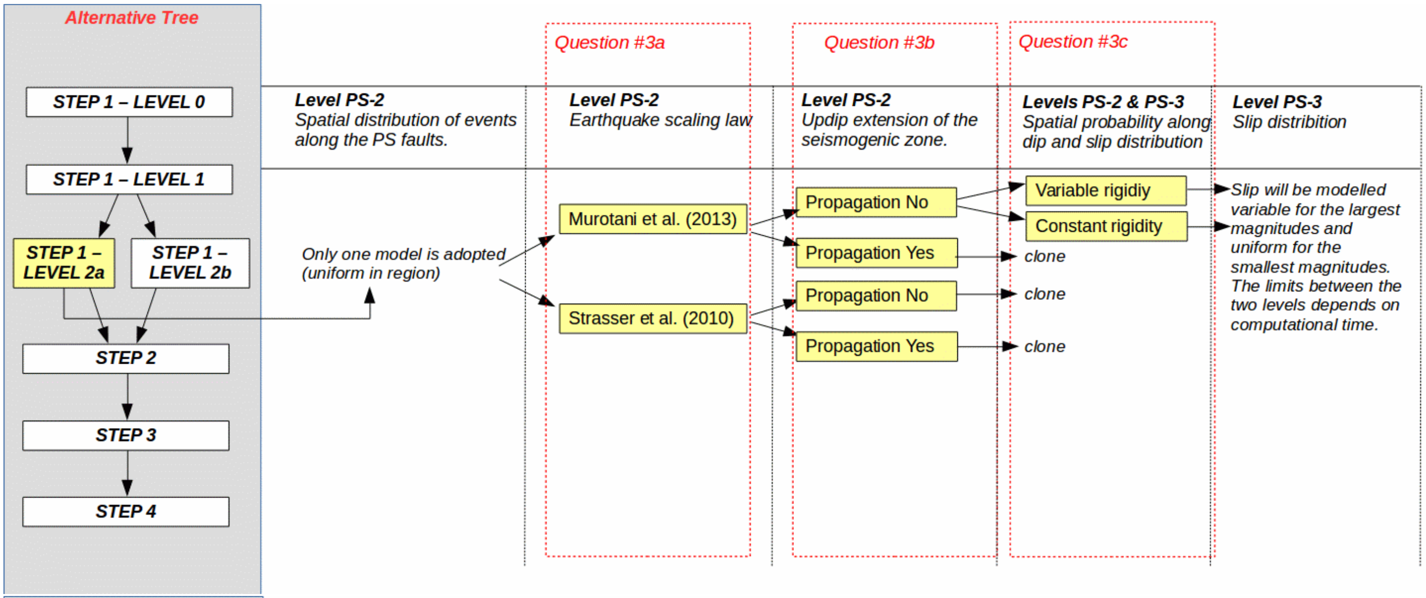


Figure 2: Alternative tree for STEP 1, divided into LEVEL 0 (top panel), LEVEL 1 (second panel from top), LEVEL 2a (third panel) and LEVEL 2b (last panel). At this STEP, a total of 14 models are to be compared, through 6 questions: Q1, Q2a, Q2b, Q3a, Q3b, Q3c.




Alternative Tree
STEP 1 – LEVEL 0
STEP 1 – LEVEL 1
STEP 1 – LEVEL 2a
STEP 1 – LEVEL 2b
STEP 2
STEP 3
STEP 4
Question #3a
Question #3b
Question #3c
Level PS-2
Spatial distribution of events along the PS faults.
Level PS-2
Earthquake scaling law
Level PS-2
Updip extension of the seismogenic zone.
Levels PS-2 & PS-3
Spatial probability along dip and slip distribution
Level PS-3
Slip distribition
Only one model is adopted (uniform in region)
Murotani et al. (2013)
Strasser et al. (2010)
Propagation No
Propagation Yes
Propagation No
Propagation Yes
Variable rigidiy
Constant rigidity
clone
clone
clone
Slip will be modelled variable for the largest magnitudes and uniform for the smallest magnitudes. The limits between the two levels depends on computational time.

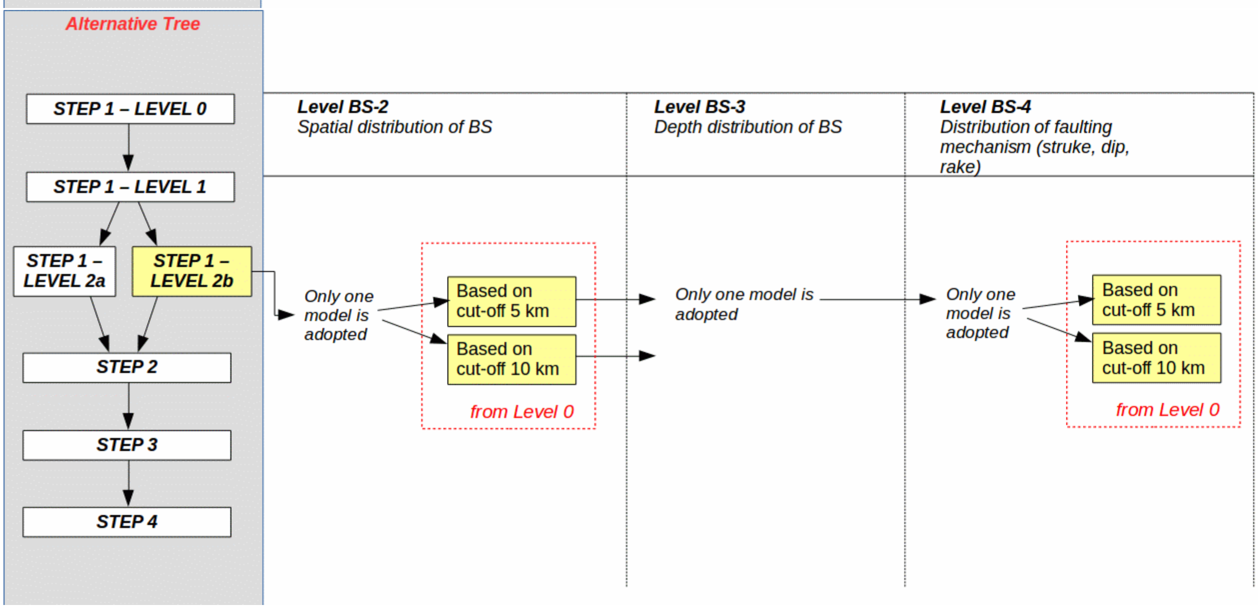

Alternative Tree
STEP 1 – LEVEL 0
STEP 1 – LEVEL 1
STEP 1 – LEVEL 2a
STEP 1 – LEVEL 2b
STEP 2
STEP 3
STEP 4
Level BS-2
Spatial distribution of BS
Level BS-3
Depth distribution of BS
Level BS-4
Distribution of faulting mechanism (struke, dip, rake)
Only one model is adopted
Based on cut-off 5 km
Based on cut-off 10 km
from Level 0
Only one model is adopted
Only one model is adopted
Based on cut-off 5 km
Based on cut-off 10 km
from Level 0

Figure 3: Alternative tree for STEP 2. No alternatives are foreseen at this Level.

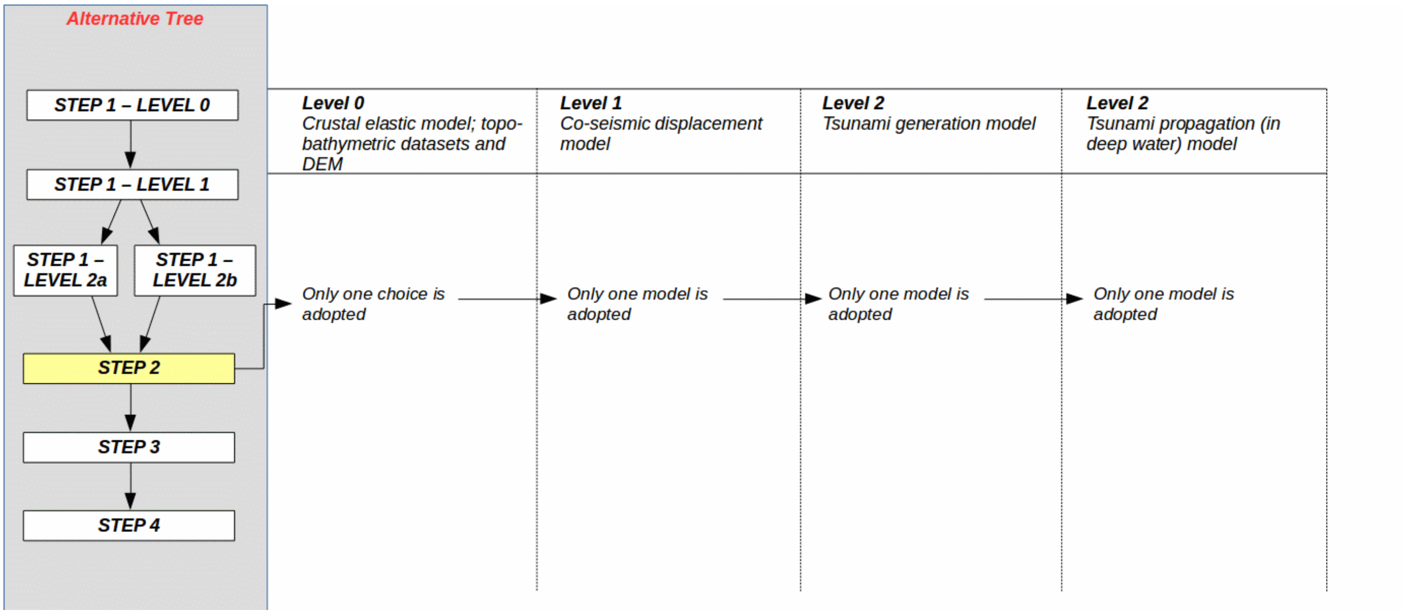


Figure 4: Alternative tree for STEP 3. No alternatives are foreseen at this Level.

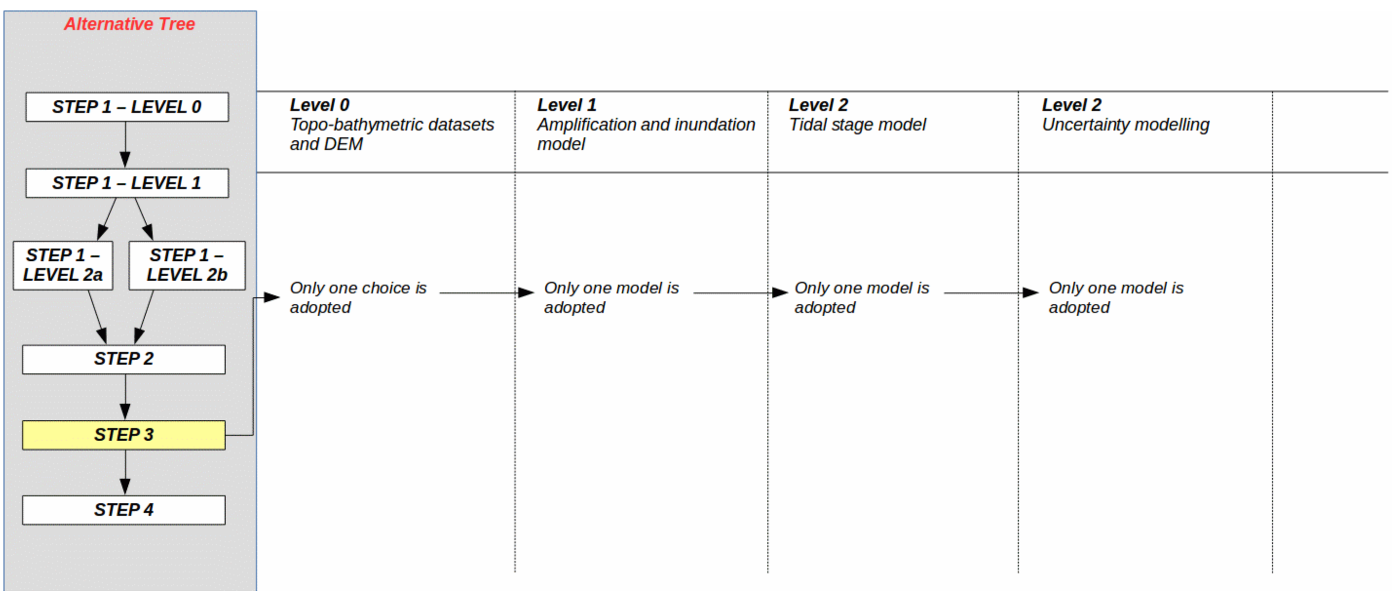


Figure 5: Alternative tree for STEP 3. At this STEP, a total of 2 models are to be compared, through 1 question: Q4.

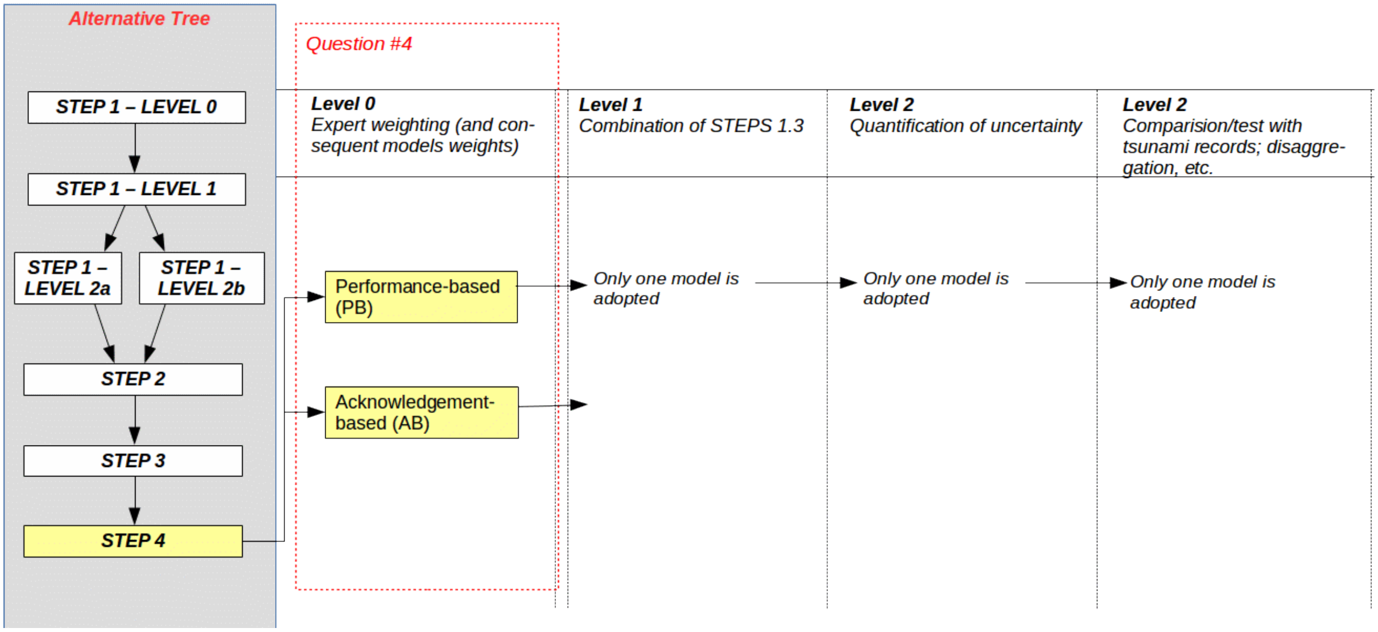


Table 1: Fundamental scale of absolute numbers

Intensity of Importance	Definition	Explanation	Weights of models
1	Equal importance	Two STEPs/Levels/sub-Levels contribute equally to the objective	0.5-0.5
3	Moderate preference	Experience and judgment slightly favor one STEP/Level/sub-Level over another	0.6-0.4 (x1.5)
5	Strong preference	Experience and judgment strongly favor one STEP/Level/sub-Level over another	0.75-0.25 (x3)
7	Very strong preference	A STEP/Level/sub-Level is favored very strongly over another; its dominance demonstrated in practice	0.95-0.05 (x19)
9	Extreme preference	Overwhelming evidence favoring one STEP/Level/sub-Level over another	0.99-0.01 (x99)

Illustrative example of AHP

<u>We provide you</u>: A set of alternatives (say five), as in **Table 2** below.

Table 2: Alternatives to be used for pairwise comparisons.

No.	Model code	Description
1	A1	*Alternative 1*
2	A2	*Alternative 2*
3	A3	*Alternative 3*
4	A4	*Alternative 4*
5	A5	*Alternative 5*

<u>We ask you to provide the relative importance of models in pairs</u>, through a question like: *In your opinion, the alternative in **column A** is either **More important** or **Equal important** or **Less important** than the alternative in **column B** with respect to your personal belief? <u>Please express your opinion by ticking a box with 'X' in each row in **Table 3**</u>. For intensity of importance, please see Table 1.*

Table 3: Pairwise comparisons of alternatives. Red crosses are used to express the expert opinion.

No. of comparisons	**A**	Intensity of importance									**B**
		More important than				Equal	Less important than				
		9	7	5	3	1	3	5	7	9	
1	A1				X						A2
2	A1							X			A3
3	A1					X					A4
4	A1								X		A5
5	A2				X						A3
6	A2			X							A4
7	A2					X					A5
8	A3			X							A4
9	A3								X		A5
10	A4									X	A5

The choices in table 3 stand for:

i) <u>For the first row</u>, if A1 is **more important** with moderate intensity than A2, then tick the box with '**X**' under 3 on the left side of 1.

ii) <u>For the second row</u>, if A1 is **less important** with strong intensity than A3, then tick the box with '**X**' under 5 on the right side of 1.

iii) <u>For the third now</u>, If A1 is **equally important** to A4, then tick the box with '**X**' under 1.

And so on . . .

Question 0: Criteria for the comparison

The comparison among the alternative models can be made adopting different points of view. For example, one expert may strongly believe in one model, but he/she may be also aware that this model still requires some development, or that this model is still not well recognized within the community.

In TSUMAPS-NEAM, we will adopt 2 possible criteria to compare the models, as summarized in the following table:

No.	Model code	Description
1	PREF	Expert's personal preference
2	USED	Most used in the community according to expert's best knowledge

Please, answer to question Q0 on the next page.

QUESTION Q0

In your opinion, the criterion PREF in **column A** is either **More/ Equal / Less appropriate** than the criterion USED in **column B**, to weight the models for quantifying the "community distribution"?

Please express your opinion by ticking a box with '**X**' in each row in the following **Table Q0**. For intensity of importance, please see ***Table 1***.

Table Q1: Pairwise comparisons of S-PTHA STEPs.

No. of comparisons	A	Intensity of importance									B
		More appropriate than				Equal	Less appropriate than				
		9	7	5	3	1	3	5	7	9	
1	PREF										USED

Question 1: Prioritization of alternatives at STEP 1 – Level 0

At STEP 1 – Level 0, the fundamental databases (including seismicity catalogues, fault catalogues, regionalization, convergence rates, etc.) are set for the probabilistic quantification of STEP 1.

At this Level, two alternative methods are used to assign the observed seismicity either to the Prevalent Seismicity (PS) or the Background Seismicity (BS) sources. This serves for assessing the relative proportion of seismic rate and the Probability Density Functions (PDFs) of the faulting mechanism in a region (for example, to assign the observed seismicity either to a subduction interface or to its surroundings; or the rate of strike-slip versus dip-slip faulting in a given portion of the crust).

All the adopted catalogues have been separated in two parts: PS-only and BS-only catalogues. This has been done by adopting two alternative procedures, by using two different seismicity cut-off distances of 5 and 10 km. These two alternatives affect all the Levels of STEP 1 where data from these catalogues are used.

The two alternatives at STEP 1 Level 0 are reported in the following table:

No.	Model code	Description
1	5km	Cut-off distance of 5 km around the PS sources
2	10km	Cut-off distance of 10 km around the PS sources

Please, answer to questions Q1.1 and Q1.2 on the next page.

QUESTION Q1.1

Adopting the criterion C1 (personal preference), the model 5km in **column A** in **Table Q1.1** is **More reliable** or **Equally reliable** or **Less reliable** than the model 10km in **column B** to separate the observed seismicity into BS-only and PS-only catalogues? Please express your opinion by ticking a box with '**X**' in each row in following **Table Q1.1**. For intensity of importance, please see *Table 1*.

Table Q1.1: Pairwise comparisons of alternatives of STEP 1 – LEVEL 0, adopting Criterion C1

No. of comparisons	A	Intensity of importance									B
		More appropriate than				Equal	Less appropriate than				
		9	7	5	3	1	3	5	7	9	
1	5km										10 km

QUESTION Q1.2

Adopting the criterion C2 (most used in the community according to expert's best knowledge), the model 5km in **column A** in **Table Q1.2** is **More reliable** or **Equally reliable** or **Less reliable** than the model 10km in **column B** to separate the observed seismicity into BS-only and PS-only catalogues? Please express your opinion by ticking a box with '**X**' in each row in following **Table Q1.2**. For intensity of importance, please see *Table 1*.

Table Q1.2: Pairwise comparisons of alternatives of STEP 1 – LEVEL 0, adopting Criterion C2

No. of comparisons	A	Intensity of importance									B
		More appropriate than				Equal	Less appropriate than				
		9	7	5	3	1	3	5	7	9	
1	5km										10 km

Question 2: Prioritization of alternatives at STEP 1 – Level 1

At STEP 1 – Level 1, the frequency of the different magnitudes in each region is quantified as the sum of the contribution of Predominant Seismicity (PS) and Background Seismicity (BS). An earthquake belongs to a region if the geometrical centre of its fault lies within this region. The assessment consists of quantifying mean annual rates for a set of discrete magnitude intervals M_j, with reference to the defined exposure time window (50 yr), for both Predominant and Background Seismicity in region R_i, that is $\lambda_i^{(PS)}(M_j)$ and $\lambda_i^{(BS)}(M_j)$, respectively. These two quantifications correspond to the first Level of the ET (PS-1 and BS-1, respectively) as described in Section 2.5 of document ***Doc_P1_S1_Project_Summary***.

At this Level, we have a quite large number of alternative models (a total of 6). As reported in Figure 2, we separate the comparison of these alternative models in 2 questions: Q2a and Q2b.

In Q2a, we consider two alternatives: either rates for PS and BS are quantified jointly, or independently. In Q2b, we enter into the details of the joint quantification of the regional magnitude-frequency (MF) distributions, by considering 4 alternative implementations. No questions will be asked for the separate quantification, since this quantification is based on Davies et al. (2017), from which also the model weights will be adopted.

Question Q2a; Joint vs independent Magnitude-Frequency (MF) distributions for PS/BS

Two main alternatives are considered in quantifying the magnitude-frequency (MF) distributions for PS and BS in each region: either PS/BS distributions are quantified jointly, or independently.

For the joint PS/BS quantification, the MF distribution is obtained in two stages: in stage 1, a common MF for the region is quantified; in stage 2, the MF is split into PS and BS seismicity as a function of the magnitude (Selva et al. 2016), that is:

$$\begin{cases} \boldsymbol{\lambda_i^{(PS)}}(M_j) = \lambda_i(M_j)\,\Pr(PS|M_j, R_i) \\ \boldsymbol{\lambda_i^{(BS)}}(M_j) = \lambda_i(M_j)\,\Pr(BS|M_j, R_i), = \lambda_i(M_j)[1 - \Pr(PS|M_j R_i)] \end{cases}$$

where $\lambda_i(M_j)$ is the total mean annual rate of earthquakes within the region R_i having a magnitude within the interval range M_j, and $\Pr(PS|M_j, R_i)$ represents the probability that a randomly selected event within region R_i and interval M_j belongs to the PS. Both these quantifications are based on a Bayesian formulation, with data coming from the non-declustered seismic catalogue.

For the separate PS-BS quantification, the MF distribution for PS is set as in Davies et al. (2017). This means that for constraining the rate of activity of PS we will use the classical formulation for seismic moment rate $\dot{m}_s$ as given by

$$\dot{m}_s = \chi \dot{m}_g = \chi \mu A \dot{D}$$

where $\dot{m}_g$ is the geologic moment rate, χ is a coefficient that determines how much of the geologic rate is converted into the seismic rate (so called coupling or seismic efficiency), μ is the rigidity or shear modulus, A is the fault area, and $\dot{D}$ is either convergence rate for subduction or slip rate for other faults.

These two alternatives are reported in the following table:

No.	Model code	Description
1	JointMF	The mean annual rates rates for PS and BS are quantified jointly
2	IndepMF	The mean annual rates rates for PS and BS are quantified independently

Please, answer to questions Q2a.1 and Q2a.2 on the next page.

QUESTION Q2a.1

Adopting the criterion C1 (personal preference), the model JointMF in **column A** in **Table Q2a.1** is **More reliable** or **Equally reliable** or **Less reliable** than the model IndepMF in **column B** to quantify the total mean annual seismicity rates of BS and PS? Please express your opinion by ticking a box with '**X**' in each row in following **Table Q2a.1**. For intensity of importance, please see *Table 1*.

Table Q2a.1: Pairwise comparisons of alternatives of STEP 1 – LEVEL 1 (part a), adopting Criterion C1

No. of comparisons	A	Intensity of importance									B
		More appropriate than				Equal	Less appropriate than				
		9	7	5	3	1	3	5	7	9	
1	JointMF										IndepMF

QUESTION Q2a.2

Adopting the criterion C2 (most used in the community according to expert's best knowledge), the model JointMF in **column A** in **Table Q2a.2** is **More reliable** or **Equally reliable** or **Less reliable** than the model IndepMF in **column B** to quantify the total mean annual seismicity rates of BS and PS? Please express your opinion by ticking a box with '**X**' in each row in following **Table Q2a.2**. For intensity of importance, please see *Table 1*.

Table Q2a.1: Pairwise comparisons of alternatives of STEP 1 – LEVEL 1 (part a), adopting Criterion C2

No. of comparisons	A	Intensity of importance									B
		More appropriate than				Equal	Less appropriate than				
		9	7	5	3	1	3	5	7	9	
1	JointMF										IndepMF

Question Q2b: Functional form of the Magnitude-Frequency (MF) distribution and $\boldsymbol{\beta}$ *value*

For the quantification of $\lambda_i(M_j)$, we select the procedure based on the Bayesian formulation by Campbell (1982). This procedure was first suggested for the unbounded Gutenberg-Richter (GR) distribution and later refined by Keller et al. (2014) for the truncated GR distribution. The novelty of our work consists a) in extending the methodology of Keller et al. (2014) to any MF distribution and b) in the simultaneous estimation of all parameters. In this way, potential correlations among the different parameters (for example, a- and b-values) can be accounted for. Following Keller et al. (2014), we include the temporal variability of the completeness period with magnitude, as proposed by Weichert (1980). This allows including different time windows with different magnitude thresholds for the completeness of the catalogue (e.g., M ≥ 4.5 from 1960, M ≥ 5.5 from 1905, M ≥ 6.5 from 1450). Doing this, we can extend backward in time the part of the catalogue used for the estimation, without loosing information from both the recent catalogues based on recent well developed networks with relatively low magnitude threshold for the completeness, and historical catalogues based on the analysis of the historical records with relatively high magnitude threshold for the completeness.

The method described above will be implemented considering the following alternatives.

Both Truncated and Tapered Pareto functional forms are considered as two alternatives for the MF distribution (see Figure 6). The tapered Pareto distribution considers a corner magnitude M_c over which the probability drops, but does not vanish. On the opposite, the truncated Pareto considers a maximum magnitude M_{max} over which the probability is identically 0. The prior distributions will be set as non-informative or slightly informative for λ_0 and M_{max} / M_c (the upper limit for magnitude in the truncated Pareto and the corner magnitude of the tapered Pareto, respectively), considering all the known constraints (for example, maximum magnitude observed in the region).

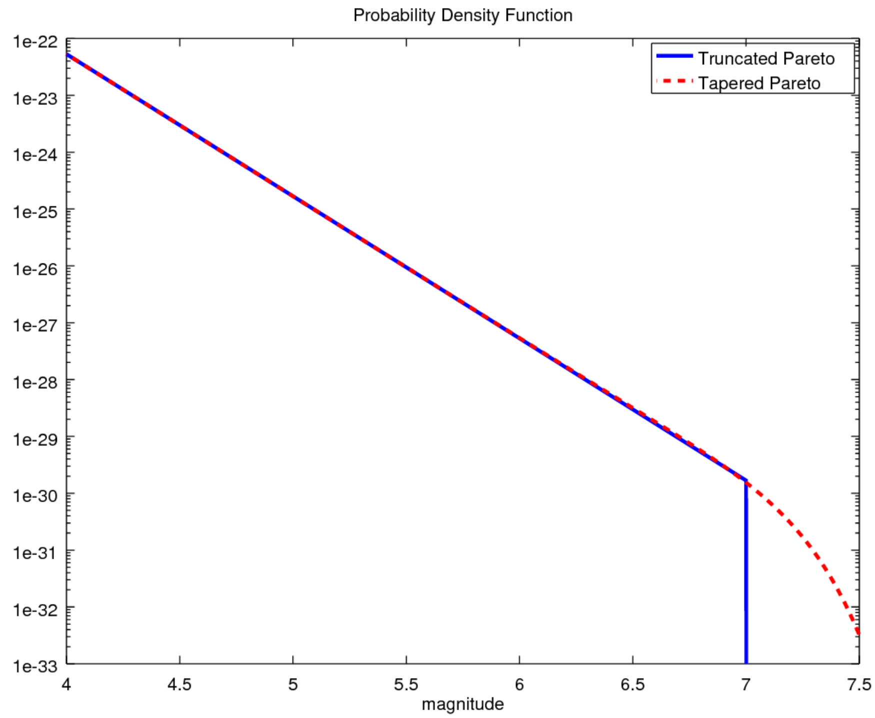


Figure 6: PDFs of Tapered and Truncated Pareto distributions, with parameters Mmax = Mc = 7 and β = 2/3 (equivalent to the b-value = 1).

Two further alternatives are planned for the parameter β (2/3 of the b-value), considering two informative priors based either on worldwide tectonic analogue estimations of Kagan et al. (2010), or forcing the b-value to 1. Note that, in case of few data, the values will be controlled mainly by the prior, that is, from the b-value of the worldwide data from similar tectonic settings.

Within the described approach, we enumerated a total of 4 Bayesian alternatives, as reported in the following table

No.	Model code	Description
1	Tapered & β = 2/3	The tapered distribution (with probability > 0 for all magnitudes) with the parameter β (equivalent to the b-value) set to 2/3 (equivalent to b-value = 1), independently from data.
2	Tapered & β = from data	The tapered distribution (with probability > 0 for all magnitudes) with the parameter β (equivalent to the b-value) set from data.
3	Truncated & β from data	The truncated distribution (with probability = 0 for all M > M_{max}) with the parameter β (equivalent to the b-value) set to 2/3 (equivalent to b-value = 1), independently from data.
4	Truncated & β = 2/3	The truncated distribution (with probability = 0 for all M > M_{max}) with the parameter β (equivalent to the b-value) set from data.

Please, answer to questions Q2b.1 and Q2b.2 on the next page.

QUESTION Q2b.1

Adopting the criterion C1 (personal preference), the model in **column A** in **Table Q2b.1** is **More reliable** or **Equally reliable** or **Less reliable** than the model in **column B** as Functional form of the magnitude frequency distribution? Please express your opinion by ticking a box with '**X**' in each row in following **Table Q2b.1**. For intensity of importance, please see *Table 1*.

Table Q2b.1: Pairwise comparisons of alternatives of STEP 1 – LEVEL 1 (part b), adopting Criterion C1

No. of comparisons	A	Intensity of importance									B
		More appropriate than				Equal	Less appropriate than				
		9	7	5	3	1	3	5	7	9	
1	Tapered & β = 2/3										Tapered & β = from data
2	Tapered & β = 2/3										Truncated & β from data
3	Tapered & β = 2/3										Truncated & β = 2/3
4	Tapered & β = from data										Truncated & β from data
5	Tapered & β = from data										Truncated & β = 2/3
6	Truncated & β from data										Truncated & β = 2/3

QUESTION Q2b.2

Adopting the criterion C2 (most used in the community according to expert's best knowledge), the model in **column A** in **Table Q2b.2** is **More reliable** or **Equally reliable** or **Less reliable** than the model in **column B** as Functional form of the magnitude frequency distribution? Please express your opinion by ticking a box with '**X**' in each row in following **Table Q2b.2**. For intensity of importance, please see *Table 1*.

Table Q2b.2: Pairwise comparisons of alternatives of STEP 1 – LEVEL 1 (part b), adopting Criterion C2

No. of comparisons	A	Intensity of importance									B
		More appropriate than				Equal	Less appropriate than				
		9	7	5	3	1	3	5	7	9	
1	Tapered & β = 2/3										Tapered & β = from data
2	Tapered & β = 2/3										Truncated & β from data
3	Tapered										Truncated

	& β = 2/3										& β = 2/3
4	Tapered & β = from data										Truncated & β from data
5	Tapered & β = from data										Truncated & β = 2/3
6	Truncated & β from data										Truncated & β = 2/3

Question 3: Prioritization of alternatives at STEP1 – Level 2a

This Level deals with the Predominant Seismicity (PS) branch; all the parameters identifying individual sources on either 2D or 3D geometries defined. The PS analysis is subdivided into the 2 sub-Levels that stack on Level PS-1, that are:

- sub-Level PS-2 – Positioning along the PS hosting structure and rupture area
- sub-Level PS-3 – Slip distribution

At the sub-Level PS-2, earthquake positions are discretized by defining a set of coordinates $\{x_c, y_c\}$ over the fault surface. Assessment consists of quantifying the probability $\mathrm{Pr_i}(x_c, y_c, A|M_j)$, that is, the joint probability of a fault centre x_c, y_c and a (maximum) rupture area A for an earthquake of magnitude M_j in the region R_i. We simplify this quantification by computing the A as a function of magnitude M_j from scaling laws, so that $\mathrm{Pr_i}(x_c, y_c, A|M_j) = \mathrm{Pr_i}(x_c, y_c|M_j)$, since no aleatory uncertainty is modelled for A. Average slip can also be estimated from the same scaling law.

At the sub-Level PS-3, we model the aleatory variability of the heterogeneous slip distribution within the rupture area A. We quantify the joint probability of a slip vector field conditioned to the occurrence of an earthquake centred at $\{x_c, y_c\}$ and having rupture area A and magnitude M_j, that is, $\mathrm{Pr_i}(\vec{s}\ |x_c y_c, A, M_j)$. This joint probability distribution should take into account many different constrains, such as total slip, spatial correlation of slip, etc. For simplicity, instead of discretizing the slip vector space and quantifying the joint probability distribution, at this Level, we adopt a Monte-Carlo-like approach. For each rupture area, we build a sample of equally-probable stochastic slip distributions, with conditional probability $\mathrm{Pr_i}(\vec{s}\ |x_c y_c, A, M_j) = 1/n$, where n is the sample size.

We have a quite large number of alternative models (a total of 6=2+2+2 resulting in 2^3=8 combinations). As reported in Figure 2, we separate the comparison of these alternative models in 3 questions: Q3a, Q3b, and Q3c.

In Q3a, we consider two alternatives for the earthquake scaling law: either the Murotani et al. (2013) or the Strasser et al. (2010) scaling law is used.

In Q3b, we consider two alternatives for the updip extension of the seismogenic zone: co-seismic slip is not allowed or allowed to happen at shallow depths under the accretionary wedge.

Q3c is about using or not depth-dependent rigidity, as a model for explaining the observed depth-dependence of normalized earthquake duration (Bilek and Lay, 1999), with an inverse depth-dependence of slip to attain a given seismic moment (Geist and Bilek, 2001). Some details of these implementations can be found in “***Doc_P1_S1_ProjectSummary***” and in “***DOC_P1_S4_Prel_Impl_Plan”.***

The two alternatives for the scaling laws are reported in the following table:

No.	Model code	Description
1	Strasser	Scaling laws from Strasser et al. (2010).
2	Murotani	Scaling laws from Murotani et al. (2013).

Please, answer to questions Q3a.1 and Q3a.2 on the next page.

QUESTION Q3a.1

Adopting the criterion C1 (personal preference), the model Strasser in **column A** in **Table Q3a.1** is **More reliable** or **Equally reliable** or **Less reliable** than the model Murotani in **column B** to quantify the parameters of individual faults with the PS structures? Please express your opinion by ticking a box with '**X**' in each row in following **Table Q3a.1**. For intensity of importance, please see *Table 1*.

Table Q3a.1: Pairwise comparisons of alternatives of STEP 1 – LEVEL PS-1 sub-Level PS-2 (part a), adopting Criterion C1

No. of comparisons	A	Intensity of importance									B
		More appropriate than				Equal	Less appropriate than				
		9	7	5	3	1	3	5	7	9	
1	Strasser										Murotani

QUESTION Q3a.2

Adopting the criterion C2 (most used in the community according to expert's best knowledge), the model Strasser in **column A** in **Table Q3a.2** is **More reliable** or **Equally reliable** or **Less reliable** than the model Murotani in **column B** to quantify the parameters of individual faults with the PS structures? Please express your opinion by ticking a box with '**X**' in each row in following **Table Q3a.2**. For intensity of importance, please see *Table 1*.

Table Q3a.2: Pairwise comparisons of alternatives of STEP 1 – LEVEL PS-1 sub-Level PS-2 (part a), adopting Criterion C2

No. of comparisons	A	Intensity of importance									B
		More appropriate than				Equal	Less appropriate than				
		9	7	5	3	1	3	5	7	9	
1	Strasser										Murotani

The two alternatives concerning the seismogenic depth are reported in the following table:

No.	Model code	Description
1	NUCL	Co-seismic slip is not allowed or allowed to happen at shallow depths under the accretionary wedge.
2	NUCL&PROP	Co-seismic slip is allowed to happen at shallow depths under the accretionary wedge.

Please, answer to questions Q3b.1 and Q3b.2 on the next page.

QUESTION Q3b.1

Adopting the criterion C1 (personal preference), the model NUCL in **column A** in **Table Q3b.1** is **More reliable** or **Equally reliable** or **Less reliable** than the model NUCL&PROP in **column B** to model the updip extension of the seismogenic zone? Please express your opinion by ticking a box with '**X**' in each row in following **Table Q3b.1**. For intensity of importance, please see *Table 1*.

Table Q3b.1: Pairwise comparisons of alternatives of STEP 1 – LEVEL PS-1 sub-Level PS-2 (part b), adopting Criterion C1

No. of comparisons	A	Intensity of importance									B
		More appropriate than				Equal	Less appropriate than				
		9	7	5	3	1	3	5	7	9	
1	NUCL										NUCL&PROP

QUESTION Q3b.2

Adopting the criterion C2 (most used in the community according to expert's best knowledge), the model NUCL in **column A** in **Table Q3b.2** is **More reliable** or **Equally reliable** or **Less reliable** than the model NUCL&PROP in **column B** to model the updip extension of the seismogenic zone? Please express your opinion by ticking a box with '**X**' in each row in following **Table Q3b.2**. For intensity of importance, please see *Table 1*.

Table Q3b.2: Pairwise comparisons of alternatives of STEP 1 – LEVEL PS-1 sub-Level PS-2 (part b), adopting Criterion C2

No. of comparisons	A	Intensity of importance									B
		More appropriate than				Equal	Less appropriate than				
		9	7	5	3	1	3	5	7	9	
1	NUCL										NUCL&PROP

The two alternatives concerning the depth-dependence of rigidity are reported in the following table:

No.	Model code	Description
1	Uniform	Rigidity is uniform with depth (PREM).
2	Depth-depend.	Rigidity varies with depth according to Geist and Bilek 2001.

Please, answer to questions Q3c.1 and Q3c.2 on the next page.

QUESTION Q3c.1

Adopting the criterion C1 (personal preference), the model Uniform in **column A** in **Table Q3c.1** is **More reliable** or **Equally reliable** or **Less reliable** than the model Depth-depen. in **column B** to model the updip extension of the seismogenic zone? Please express your opinion by ticking a box with '**X**' in each row in following **Table Q3c.1**. For intensity of importance, please see *Table 1*.

Table Q3c.1: Pairwise comparisons of alternatives of STEP 1 – LEVEL PS-1 sub-Level PS-3 (part c), adopting Criterion C1

No. of comparisons	A	Intensity of importance									B
		More appropriate than				Equal	Less appropriate than				
		9	7	5	3	1	3	5	7	9	
1	Uniform										Depth-depend.

QUESTION Q3c.2

Adopting the criterion C2 (most used in the community according to expert's best knowledge), the model Uniform in **column A** in **Table Q3c.2** is **More reliable** or **Equally reliable** or **Less reliable** than the model Depth-depend. in **column B** to model the updip extension of the seismogenic zone? Please express your opinion by ticking a box with '**X**' in each row in following **Table Q3c.2**. For intensity of importance, please see *Table 1*.

Table Q3c.2: Pairwise comparisons of alternatives of STEP 1 – LEVEL PS-1 sub-Level PS-3 (part c), adopting Criterion C2

No. of comparisons	A	Intensity of importance									B
		More appropriate than				Equal	Less appropriate than				
		9	7	5	3	1	3	5	7	9	
1	Uniform										Depth-depend.

Question 4: Expert weights method

The relative credibility of alternative implementations is quantified by means of weights. The assessment consists of the quantification of $\mathrm{w_{mnl}}$ where m represents a given alternative model of STEP n and Level l. These weights are subjective and will be quantified through this structured elicitation experiment. The results are obtained by combining the answers of all the experts of the PoE of TSUMAPS, weighting their answers by the expert weights.

In agreement with what we have done analysing the first round of elicitation of the PoE (see the attached ***Doc_P1_S3_Elicitation***), we plan to consider performance-based and acknowledgement-based weights as 2 alternative weighting schemes for experts also for the second elicitation. As a sensitivity test, we will also check the consistency of the results against the equal weights assumption. The meaning of these two alternative weighting schemes are detailed in the attached documents (e.g. ***Doc_P1_S2_PoEKickoff***) and have been discussed and explained in detail to all the members of the PoE during the PoE's kick-off meeting (held in Athens in June 2016).

We only recall here that in the **acknowledgement-based weighting scheme,** a weight is assigned to each expert on the basis of mutual recognition among the experts themselves. In the **performance-based weighting scheme**, the weights on experts' opinion are assigned through experts' relative performance in answering a set of seed questions

The two alternatives at STEP 4 Level 0 are reported in the following table:

No.	Model code	Description
1	AW	Acknowledgement-based weighting scheme
2	PW	Performance-based weighting scheme

Please, answer to questions Q4.1 and Q4.2 on the next page.

QUESTION Q4.1

Adopting the criterion C1 (personal preference), the model AW in **column A** in **Table Q4.1** is **More reliable** or **Equally reliable** or **Less reliable** than the model PW in **column B** to integrate the answers of the different experts for quantifying the model weights? Please express your opinion by ticking a box with '**X**' in each row in following **Table Q4.1**. For intensity of importance, please see *Table 1*.

Table Q4.1: Pairwise comparisons of alternatives of STEP 4 – LEVEL 0, adopting Criterion C1

No. of comparisons	A	Intensity of importance									B
		More appropriate than				Equal	Less appropriate than				
		9	7	5	3	1	3	5	7	9	
1	AW										PW

QUESTION Q4.2

Adopting the criterion C2 (most used in the community according to expert's best knowledge), the model AW in **column A** in **Table Q4.2** is **More reliable** or **Equally reliable** or **Less reliable** than the model PW in **column B** to integrate the answers of the different experts for quantifying the model weights? Please express your opinion by ticking a box with '**X**' in each row in following **Table Q4.2**. For intensity of importance, please see *Table 1*.

Table Q4.2: Pairwise comparisons of alternatives of STEP 4 – LEVEL 0, adopting Criterion C2

No. of comparisons	A	Intensity of importance									B
		More appropriate than				Equal	Less appropriate than				
		9	7	5	3	1	3	5	7	9	
1	AW										PW

References

Bilek, S. L., T. Lay, (1999) Rigidity variations with depth along interplate megathrust faults in subduction zones, Nature, 400, 443–446.

Davies G, Griffin J, Løvholt F, Glymsdal S, Harbitz C, Thio HK, Lorito S, Basili R, Selva J, Geist E, Baptista MA (2017). A global probabilistic tsunami hazard assessment from earthquake sources. From: Scourse, E. M., Chapman, N. A., Tappin, D. R. & Wallis, S. R. (eds) Tsunamis: Geology, Hazards and Risks. Geological Society, London, Special Publications, 456, https://doi.org/10.1144/SP456.5.

Geist, E. L. and Bilek, S. L. (2001), Effect of depth-dependent shear modulus on tsunami generation along subduction zones. Geophysical Research Letters, 28: 1315–1318. doi: 10.1029/2000GL012385.

Kagan, Y.Y., Bird, P., Jackson, D.D. (2010) Earthquake Patterns in Diverse Tectonic Zones of the Globe, Pure Appl. Geophys., 167(6), 721-741, doi:10.1007/s00024-010-0075-3.

Keller M., Pasanisi A., Marcilhac M., Yalamas T., Secanell R., and Senfaute G. (2014), A Bayesian Methodology Applied to the Estimation of Earthquake Recurrence Parameters for Seismic Hazard Assessment, Qual. Reliab. Engng. Int., 30, 921–933, doi: 10.1002/qre.1735

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.

Murotani, S., Satake, K., Fujii, Y., 2013. Scaling relations of seismic moment, rupture area, average slip, and asperity size forM~9 subduction-zone earthquakes. Geophysical Research Letters 40, 5070-5074, doi: 10.1002/grl.50976.

Saaty, TL, 1980, The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation, ISBN 0-07-054371-2, McGraw-Hill

SSHAC (Senior Seismic Hazard Analysis Committee), 1997. Recommendations for probabilistic seismic hazard analysis: Guidance on uncertainty and use of experts, U.S. Nuclear Regulatory Commission Report NUREG/CR-6372.

Strasser, F.O., Arango, M.C., Bommer, J.J., 2010. Scaling of the source dimensions of interface and intraslab subduction-zone earthquakes with moment magnitude. Seismological Research Letters 81, 941-950, doi: 10.1785/gssrl.81.6.941.

USNRC (U.S. Nuclear Regulatory Commission), 2012. Practical Implementation Guidelines for SSHAC Level 3 and 4 Hazard Studies, Prepared by AM Kammerer & JP Ake, NRC Project Manager: R Rivera-Lugo, NUREG-2117.

Weichert DH. (1980) Estimation of the earthquake recurrence parameters for unequal observation periods for different magnitudes. Bulletin of the Seismological Society of America; 70:1337–1356.