PHASE I – STAGE 3: Elicitation Experiment

Author: TSUMAPS-NEAM Technical Integrator (TI) team

Date: 2 June 2017

Version: 1.0

Table of Contents

This page is intentionally left blank.

Executive Summary

In this document, we report the first elicitation experiment of the Pool of Experts (PoE), in which the experts are asked to trim the alternative implementations for the epistemic uncertainty quantification. Many potential alternative implementations were initially proposed at all the STEPs and levels of the analysis (as discussed in ***Doc_P1_S4_Prel_Impl_Plan***, sections 1 and 2). In order to reduce the total number of alternatives to be actually implemented without decreasing the quality of the model, a PoE elicitation is performed to specifically prioritize the STEPs/levels in terms of their potential impact on the total epistemic uncertainty. The quantitative results of the elicitations will support the decisions on which STEPs and levels the Technical Integrator (TI) team should focus the development of alternative implementations.

In ***Section 1***, we introduce the rationale and the selected method for the elicitation.

In ***Section 2***, we discuss the elicitation results.

To complete the information of the present document, we attach the following Appendixes:

- ***Appendix A***: Minutes of the Athens meeting;
- ***Appendix B***: Questionnaire used for the elicitation experiment.

This page is intentionally left blank.

1. Elicitation

1.1 Elicitation preparation

Once the overall framework for the Seismic Probabilistic Tsunami Hazard Analysis (SPTHA) is defined, many alternative implementations of the framework are possible. Alternative formulations are possible depending on alternative parameterizations of a given model, as well as alternative models and assumptions (e.g., Bommer and Scherbaum, 2008; Rougier et al., 2013). If one does not have (enough) data to falsify alternative models, they should be all considered as being *scientifically acceptable*. The quantification of the epistemic uncertainty mainly relies on quantifying the potential variability of the results that depends on the use of scientifically acceptable alternative models in the form of the community distribution (e.g., SSHAC 1997; Bommer 2012; Marzocchi et al. 2015).

In order to constrain the community distribution, it is not necessarily required to implement a huge number of alternative models. Instead, it is required to define and implement a sufficient number of alternative models (and weights) that allow for capturing simultaneously the best estimates that the evaluators can develop and the range of alternatives that should be implemented in view of the limitations of the data and the currently available knowledge (Bommer 2012). The selection of these models should theoretically depend on an extensive sensitivity analysis of the results based on the alternative implementations (Bommer and Scherbaum 2008). However, this is not possible in many cases, due for example to limitations in budget and time (Musson 2012). In case of the TSUMAPS-NEAM project, we notice the additional difficulty of dealing with a relatively young science such as the SPTHA.

In order to define a list of alternative models to be effectively implemented in this project, we first established a reference framework for the SPTHA, in which we identified a finite list of steps and levels to be followed to reach the final results. The framework proposed in Selva et al. (2016) was presented as reference model in the project kick-off meeting. Then, we widely reviewed the framework first through online discussions (Google Group of Task B), and then during the Athens meeting (see ***Appendix A***, in dedicated sessions in Day 1 - afternoon and Day 2 - afternoon). After this meeting, the general framework was finalized (an extensive description of this framework is available in ***Doc_P1_S4_Prel_Impl_Plan***).

Once the SPTHA steps and levels had been established, we could define an elicitation experiment to prioritize them. A STEP/Level is considered more important than another STEP/Level if the epistemic uncertainty associated to that specific STEP/Level is expected to be larger than that for the other STEP/Level or because its influence on the results is larger than the one of the other STEP/Level). Hence, more alternative models should be developed to carefully explore and quantify this epistemic uncertainty for the STEP/Level that is judged to be more important. Such alternative models will either be implemented, if this is feasible within the resources allocated to the project, or the need for their implementation in a future assessment will be clearly reported.

1.2 Elicitation method

There are several structured elicitation processes that are described in pertinent literature with prioritization purpose. We use one procedure that is named Analytic Hierarchy Process (AHP). AHP was originally developed by Saaty (1980); it is a multi-criteria decision-making method that is useful for making decisions when facing complex problems. The hierarchy process breaks down the complex decisions into a series of pairwise comparisons, synthesizes the results, and then helps to take into account both subjective and objective aspects of the decision. Additionally, the process incorporates a useful technique for checking the consistency of the expert judgments, thus reducing the bias in the process of decision making.

The process works by decomposing the decision-making problem into a hierarchy of evaluation criteria and alternative options among which the best decision is to be made. The best decision refers to the goal of the analysis. In general, the structure of the method consists of an overall goal, a group of options or alternatives for reaching the goal, and a group of factors or criteria that relate the alternatives to the goal.

In the first round of the elicitation, we have a simplified scheme with only one criterion. Specifically, we calculate a score for each alternative through the experts' pairwise comparisons of the models with respect to the criterion under consideration. The relative importance of one model over the others is usually expressed with numeric rating from one (equally important) to nine (extremely important) (Saaty 1980) and can be collected into a matrix; the scores are the components of the normalized principal eigenvector of this matrix (Saaty and Hu 1998). Here, we adopted the numeric translation reported in **Table 1 – Column 4 ("Weights of models")**, which was presented to the experts and reported in the introduction of the questionnaire. However, the results are tested for robustness against the classical linear rating 1 to 9 of Saaty (1980), as reported in **Table 1 – Column 5 ("Standard AHP weights")**. The test showed that the variability introduced by these different scales is negligible, as compared to the inter-expert uncertainty, in terms of both final results and consistency of the experts' answers.

Some inconsistencies may arise when many pairwise comparisons are performed (Harker and Vargas 1987), which is typically measured by the Consistency Ratio (CR) (Saaty 1980). A perfectly consistent judgement by experts should always be zero, i.e. CR = 0, but, inconsistencies are tolerated if CR < 0.1 (Saaty 1980). However, it has been suggested to relax this cutoff value up to 0.3 depending on the number of criteria and the kind of project (Goepel 2013).

One important issue in AHP is aggregation of judgements when many experts are involved. Different approaches can be employed to aggregate their individual or group opinions (Forman and Peniwati, 1998), depending on the level of the aggregation and mathematical method used for the aggregation. As for the level of aggregation, the most popular methods consist of either aggregating individual judgments regarding each set of pairwise comparisons to produce an aggregate hierarchy (aggregation of individual judgments - AIJ) or synthesizing each of the individual hierarchies and aggregating the resulting priorities (aggregation of individual priorities – AIP). As for the mathematics of the aggregation, both weighted geometric and arithmetic means are commonly used as aggregation method, considering equal or subjective weights on experts (Goepel, 2013; Forman and Peniwati, 1998; Zio, 1996). Here, we select the aggregation of individual judgments (AIJ), in order to analyze and visualize both individual and group prioritizations. We consider as equal three weighting schemes for the experts (equal (EW), performance-based (PW) and acknowledgement-based weights (AW); see ***Doc_P1_S2_PoEkickoff***). To analyze the results, we take the ensemble distribution of individual priorities of all experts as the main result and, to estimate the group central tendency, we consider both (weighted) arithmetic and geometric means.

1.3 Questionnaire

The questionnaire sent to the experts can be found in ***Appendix B***. The questionnaire is structured into a short introduction followed by 5 questions. Question #1 is focused on prioritizing the 4 STEPs foreseen for the SPTHA framework. Questions #2 through #5 are then dedicated to prioritize the levels and sub-levels inside each one of these STEPS. In appendix to the questionnaire, we also reported a list of the potential alternative implementations at each of the steps and levels. The goal of this appendix was twofold: 1) to help clarifying the meaning of levels, and 2) to specify what kind of alternatives were actually under consideration for the TSUMAPS-NEAM project.

Table 1: Fundamental scale of absolute numbers.

Intensity of Importance	**Definition**	**Explanation**	**Weights of models**	***Standard AHP weights***
1	Equal importance	Two steps/levels/sublevels contribute equally to the objective	0.5-0.5	*0.5-0.5 (x1)*
3	Moderate preference	Experience and judgment slightly favor one step/level/sublevel over another	0.6-0.4 (x1.5)	*0.75-0.25 (x3)*
5	Strong preference	Experience and judgment strongly favor one step/level/sublevel over another	0.75-0.25 (x3)	*0.83-0.17 (x5)*
7	Very strong preference	A step/level/sublevel is favored very strongly over another; its dominance demonstrated in practice	0.95-0.05 (x19)	*0.86-0.14 (x7)*
9	Extreme preference	Overwhelming evidence favoring one step/level/sublevel over another	0.99-0.01 (x99)	*0.90-0.10 (x9)*

2. Results

2.1 Analysis of the results

The questionnaire was sent to the Pool of Experts (PoE) members (15 experts). We received 14 answers, that is, all the experts with only one exception answered the questionnaire. Out of the 14 filled-in questionnaires, each including five questions, we found at least one inconsistency larger than 0.3 only for three experts. We then sent back the questionnaire to these experts, explaining the meaning of the found inconsistencies and asked to review the questionnaire only for the questions for which a high inconsistency was found. The revised answers were received only from one of them. However, since the number of experts with consistent answers is considered sufficient (12 out of 15), we report below the analysis of results considering only these answers. The results are stable with respect to this choice.

The results are used to set the priorities in developing model alternatives. For each question, we report the following plots:

- The empirical CDF of the scores of the proposed alternatives, obtained by considering the prioritization of the different experts as weighted samples; we report one plot for each weighting scheme.
- The parametric variability of the scores of the proposed alternatives, considering arithmetic and geometric means and percentiles 16th, 50th (median) and 84th; we report one plot for each weighting scheme.
- The CR of all the experts, compared with thresholds of 0.1 and 0.3. If one expert had CR > 0.3 for at least one question, the questionnaire was sent back for review of that question.
- The weights of the experts, adopting the different weighting schemes.

The prioritizations obtained by the different weighting schemes are compared, both in terms of central values and of inter-expert distributions. Based on this comparison, the steps and levels of the SPTHA are ranked into three groups:

- **High priority (red):** steps/levels with clear high priority in all weighting schemes. For these steps/levels, alternative implementations are strongly recommended by the PoE. In this case, the alternatives should be carefully selected to represent a range of models that cover the full range of scientifically acceptable modeling alternatives (following SSHAC 2012, “the center, body, and range of technically defensible interpretations”).
- **Medium priority (orange):** steps/levels with either high priority in one (but not all) the weighting schemes, or intermediate priority in all weighting schemes. For these steps/levels an evaluation of the potential consequence of alternative implementations is recommended by the PoE. In this case, some alternative implementations should be considered and/or some sensitivity test should be planned.
- **Low priority (green):** steps/levels with low priority in all weighting schemes. For these steps/levels the PoE suggests a relatively low potential impact of epistemic uncertainty and a single preferred implementation can be considered.

2.2 Specific results

The results show that:

Question #1: Alternatives are strongly encouraged for STEP 1 and STEP 3 only. The potential influence of alternatives in STEP 4 should be tested (Q5 below). Alternatives can be avoided in STEP 2 (Q3 below). The analytic results for Question #1 are reported in ***Section 2.2.1***.

Question #2: Within STEP 1, alternatives are strongly encouraged (Q1 above). From Q2, alternatives are strongly encouraged for 1) the selection of the PS interfaces to be modelled separately, and 2) the quantification of the frequency-magnitude distribution. Alternatives are recommended for 1) the seismic catalog considered, 2) the models for spatial distribution on PS, and 3) the models for slip distribution on PS. The analytic results for Question #2 are reported in ***Section 2.2.2***.

Question #3: Within STEP 2 (if alternatives were to be considered), alternatives are strongly encouraged for 1) Topo-bathymetric datasets and digital elevation models. Alternatives are recommended for 1) coseismic displacement models, 2) tsunami generation models, and 3) tsunami propagation (in deep water) models. The analytic results for Question #3 are reported in ***Section 2.2.3***.

Question #4: For STEP 3, alternatives are strongly encouraged for 1) Topo-bathymetric datasets and digital elevation models, 2) Amplification and inundation models at the points of interest along the coast, and inland, corresponding to the offshore points of STEP 2, and 3) Models of the uncertainty on the tsunami metrics. The analytic results for Question #4 are reported in ***Section 2.2.4***.

Question #5: For STEP 4, alternatives are recommended for 1) the quantification of weights of the experts, and 2) the quantification of the weights of alternative models. The analytic results for Question #5 are reported in ***Section 2.2.5***.

Any choice different from PoE suggestions requires specific justification in the selection of the alternatives actually implemented in the TSUMAPS-NEAM project (see ***Doc_P1_S4_Prel_Impl_Plan***).

2.2.1 Question #1: Prioritization of STEPs

Each of the STEPs contains a number of quantitative assessments that may potentially introduce epistemic uncertainty on the SPTHA results, as summarized in the following table:

No.	Model code	Description
1	STEP1	Definition of the seismic source variability and quantification of the long-run frequencies of all the seismic sources
2	STEP2	Tsunami generation and off-shore propagation
3	STEP3	Near-shore tsunami propagation and inundation
4	STEP4	Computation of the weights of the alternative models developed in STEPs 1 to 3 to measure their credibility, and construction of the “ensemble” model

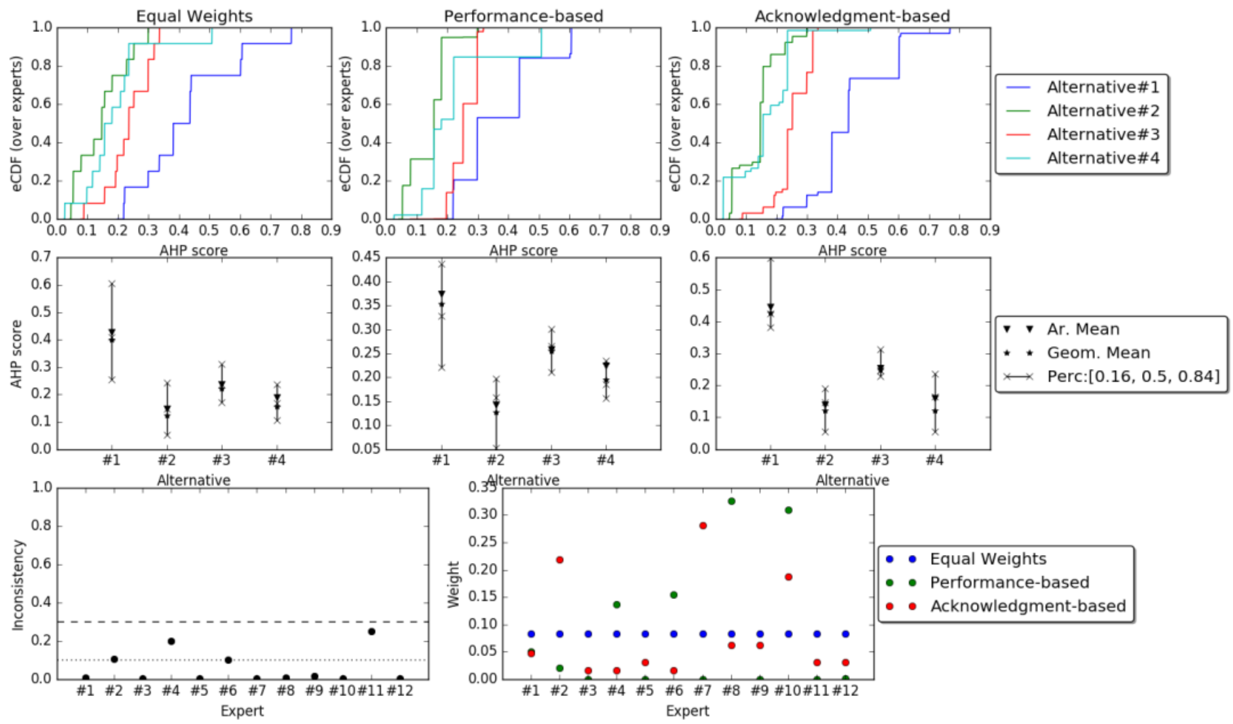

Figure: AHP results, removing highly inconsistent experts (2).

2.2.2 Question #2: Prioritization of Levels in STEP 1

Within the described levels and sublevels, we enumerated a total of 10 groups of quantitative decisions/assessments that may potentially introduce epistemic uncertainty on the STEP 1 results, as reported in the following table:

No.	Model code	Description
1	Region	Level 0 - Regionalization
2	PSDef	Level 0 - Selection of interfaces to be modeled separately
3	SeismicCat	Level 0 - Seismic catalogues
4	FreqMag	Level 1 - Quantification of the Magnitude-frequency (of PS and BS, separately)
5	PS-Pos	Level 2a - Sublevel PS-1: spatial distribution (position and area) and average slip of earthquakes over PS
6	PS-Slip	Level 2a - Sublevel PS-2: slip distribution of PS
7	BS-Pos	Level 2b - Sublevel BS-1/2: hypocentral distribution of BS
8	BS-Mech	Level 2b - Sublevel BS-3: focal mechanism of BS
9	BS-Size	Level 2b - Sublevel BS-4: size of finite faults of BS (scaling laws)
10	BS-Slip	Level 2b - Sublevel BS-5: slip distribution of BS

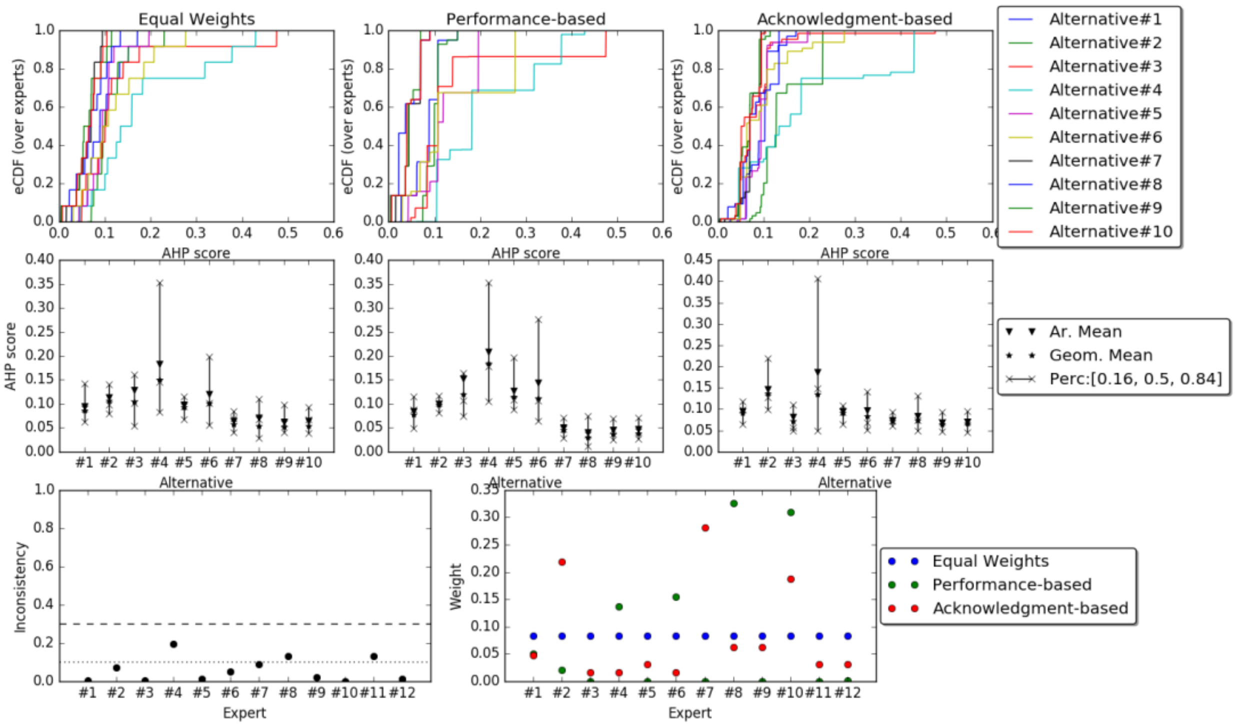

Figure: AHP results, removing highly inconsistent experts (2).

2.2.3 Question #3: Prioritization of Levels in STEP 2

Within the described levels, we enumerated a total of 5 groups of quantitative decisions/assessments that may potentially introduce epistemic uncertainty on the STEP 2 results, as reported in the following table.

No.	Model code	Description
1	Crust	Level 0 - Crustal models (elastic parameters)
2	TopoBath	Level 0 - Topo-bathymetric datasets and digital elevation models
3	CoSeis	Level 1 - Coseismic displacement model
4	TsuGen	Level 2 - Tsunami generation model
5	TsuProp	Level 3 - Tsunami propagation (in deep water) model

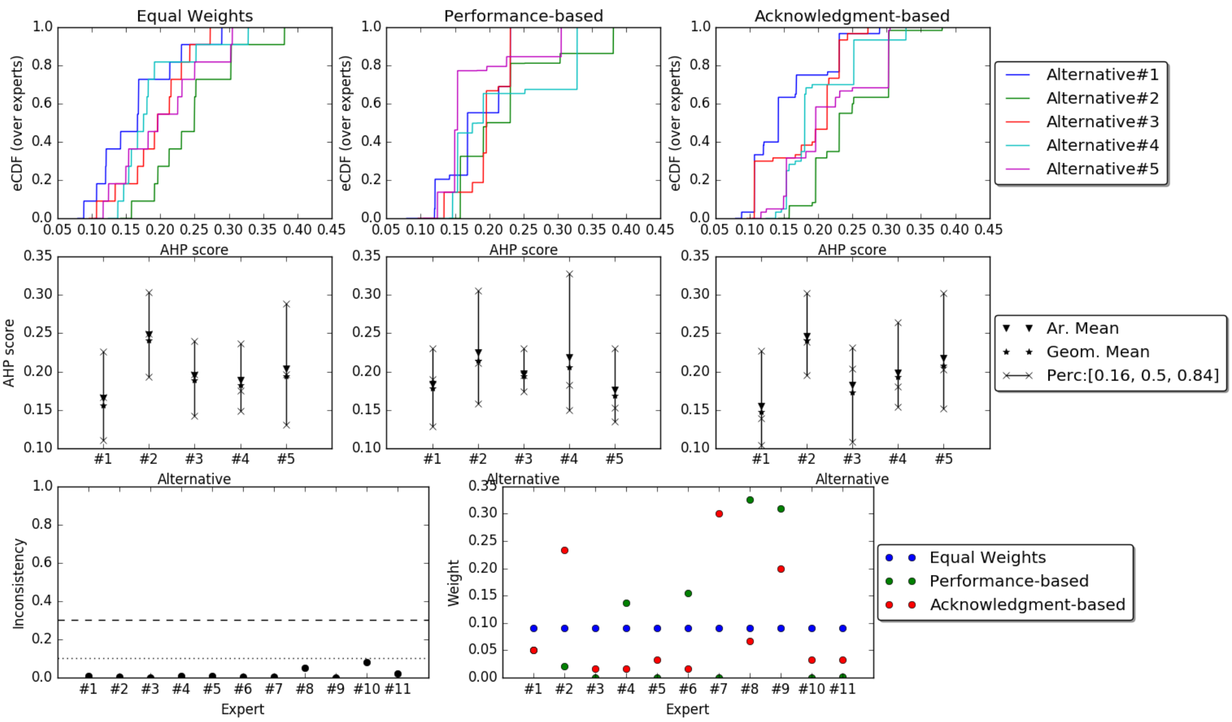

Figure: AHP results, removing highly inconsistent experts (3).

2.2.4 Question #4: Prioritization of Levels in STEP 3

Within the described levels, we enumerated a total of 4 groups of quantitative decisions/assessments that may potentially introduce epistemic uncertainty on the STEP 3 results, as reported in the following table

No.	Model code	Description
1	TopoBath	Level 0 - Topo-bathymetric datasets and digital elevation models
2	Inund	Level 1 - Amplification and inundation models at the points of interest along the coast, and inland, corresponding to the offshore points of STEP 2
3	Tide	Level 2 – Evaluation of the probability of tidal stage at the points of interest
4	Uncertainty	Level 3 - Model the uncertainty on the tsunami metrics

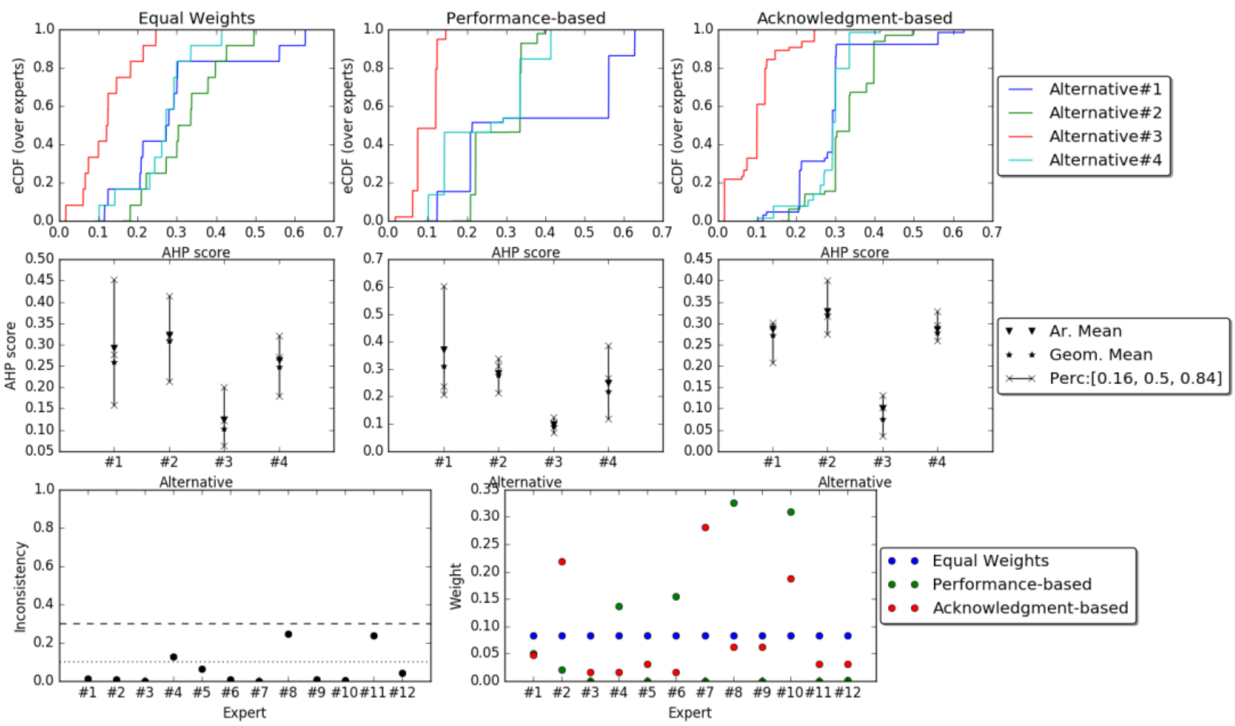

Figure: AHP results, removing highly inconsistent experts (2).

2.2.5 Question #5: Prioritization of Levels in STEP 4

Within the described levels, we enumerated a total of 2 groups of quantitative decisions/assessments that may potentially introduce epistemic uncertainty on the STEP 4 results, as reported in the following table:

No.	Model code	Description
1	WeightsExperts	Level 0 – Quantification of weights of the experts
2	Aggregation	Level 1 – Method for aggregating hazard results within each model
3	WeightsModels	Level 2 – Quantification of the weights of alternative models
4	EpisIntegration	Level 2 – Method for integrating the alternative models into a single model that quantifies also the epistemic uncertainty (e.g., Logic Tree, Ensemble models)

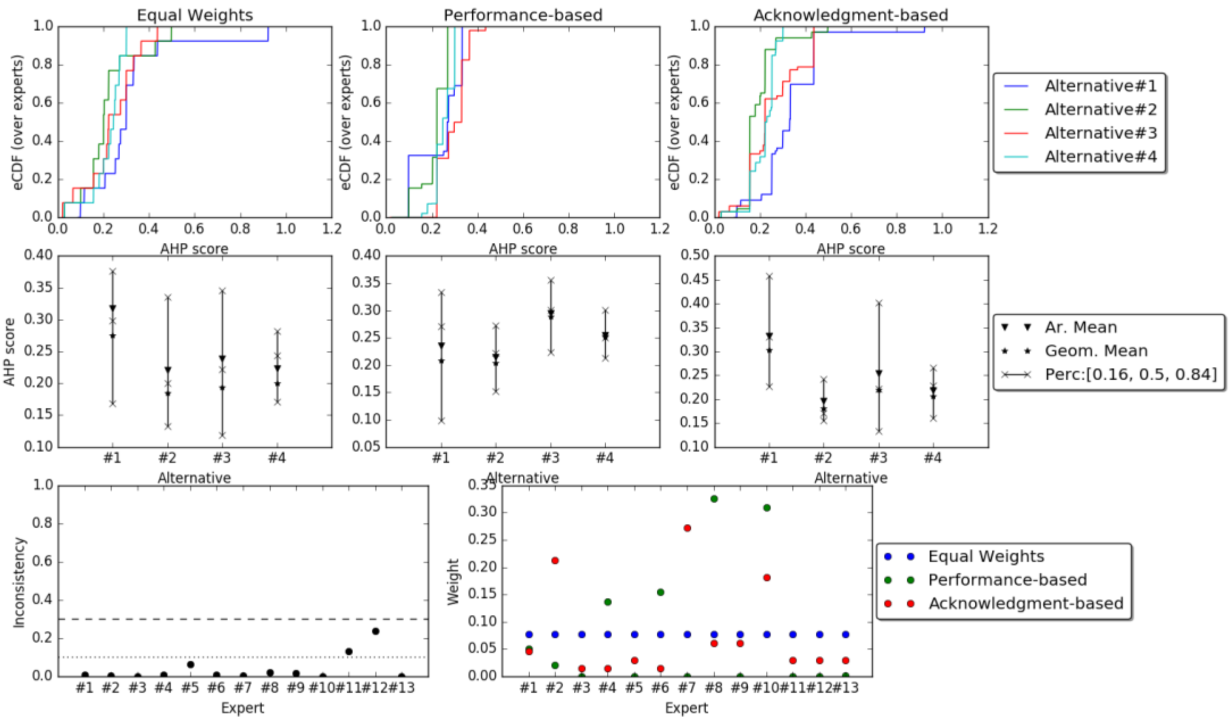

Figure: AHP results, removing highly inconsistent experts (1).

References

Bommer, J.J., Scherbaum, F., 2008. The use and misuse of logic-trees in PSHA. Earthquake Spectra 24 (4), 997–1009.

Bommer JJ (2012). Challenges of Building Logic Trees for Probabilistic Seismic Hazard Analysis, EARTHQUAKE SPECTRA, Vol: 28, 1723-1735, ISSN: 8755-2930.

Forman, E. & Peniwati, K. (1998). Aggregating individual judgments and priorities with the analytic hierarchy process. European Journal of Operational Research, 108(1), 165-169.

Goepel, K.D., 2013, IMPLEMENTING THE ANALYTIC HIERARCHY PROCESS AS A STANDARD METHOD FOR MULTI-CRITERIA DECISION MAKING IN CORPORATE ENTERPRISES – A NEW AHP EXCEL TEMPLATE WITH MULTIPLE INPUTS, Proceedings of the International Symposium on the Analytic Hierarchy Process 2013.

Harker, P., Vargas, L. 1987. The Theory of Ratio Scale Estimation: Saaty's Analytic Hierarchy Process. Management Science. 33(11), 1383–1403.

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.

Musson, R. M. W. (2012). On the nature of logic trees in probabilistic seismic hazard assessment, Earthq. Spectra 28, 1291–1296.

Rougier, J., R. Sparks, and L. J. Hill (2013), Risk assessment and uncertainty in natural hazards, in Risk and Uncertainty Assessment for Natural Hazards, edited by J. C. Rougier, R. S. J. Sparks, and L. J. Hill, pp. 1–18, Cambridge Univ. Press, Cambridge, U. K.

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

Saaty, T.L., Hu, G., 1998, Ranking by Eigenvector Versus Other Methods in the Analytic Hierarchy Process, Appl. Math. Lett. Vol. 11, No. 4, pp. 121-125, 1998

Selva, J., Tonini, R., Molinari, I., Tiberti, M.M., Romano, F., Grezio, A., Melini, D., Piatanesi, A., Basili, R., Lorito, S., 2016. Quantification of source uncertainties in Seismic Probabilistic Tsunami Hazard Analysis (SPTHA), Geophys. J. Int., 205: 1780-1803, DOI: 10.1093/gji/ggw107.

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

SSHAC 2012: 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.

Zio E., 1996. On the use of the analytic hierarchy process in the aggregation of expert judgments, Reliability Engineering & System Safety, Volume 53 (2), 127-138, ISSN 0951-8320