EXPERT JUDGEMENT BASICS AND INTRODUCTION TO THE STROMBOLI CASE STUDY

A. Tadini[(1)], A. Bevilacqua[(1)], M. De' Michieli Vitturi[(1)], A. Neri[(1)], W. Aspinall[(2)]

(1) Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Pisa, Italia
(2) University of Bristol, School of Earth Sciences, UK

December 19th 2022

EXPERT JUDGEMENT BASICS

Doubly stochastic approaches in volcanology

A volcano can be assumed as a **random system** that must be assessed with **uncertain information**

- Epistemic (imperfect knowledge of the system)
- Aleatoric (intrinsic randomness of the system)

The **forecast** of its behavior cannot be easily constrained by using simple probability models.

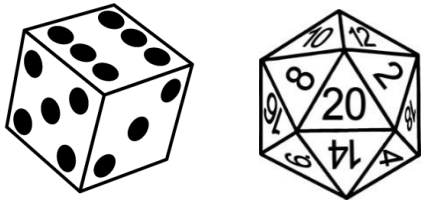
By adopting a **doubly stochastic approach,** the ill-constrained parameters of the probability models are themselves represented as additional random variables.

Therefore, all the probability estimates have their own confidence intervals.

Example: assume to roll an unknown dice, which could have 6 or 20 faces with equal chances.

The probability P of the event of getting a number N > 3 is 50% in the first case, but it is 85% in the second.

Following a doubly stochastic approach, we might say that
P is 67.5% in mean, with an uncertainty range from 50% to 85%.



Type of dice → epistemic unc.
Number on a face → aleatoric unc.

Expert judgement methods

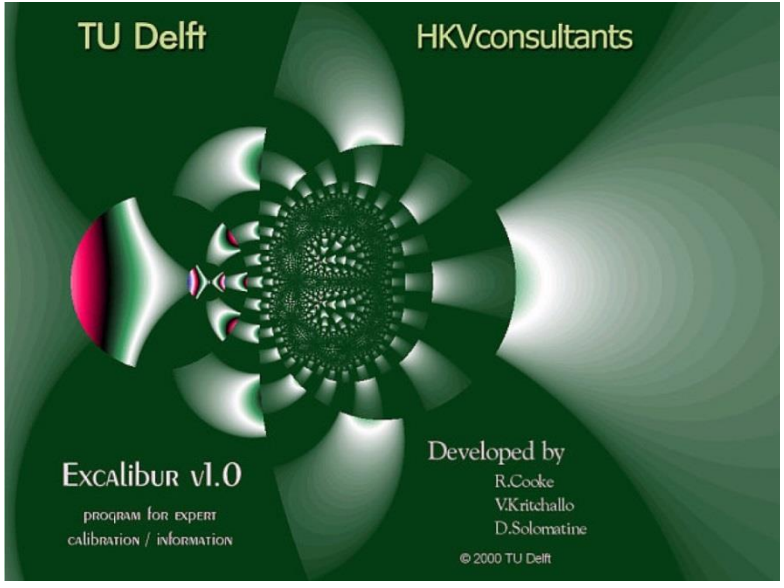
If physical models or statistical procedures are not applicable, the quantification of epistemic uncertainty can be based directly on **expert judgement.**

Expert elicitations (EE) are aimed at producing robust quantitative estimates relying on the views of a pool of experts. For example (according to our experience):

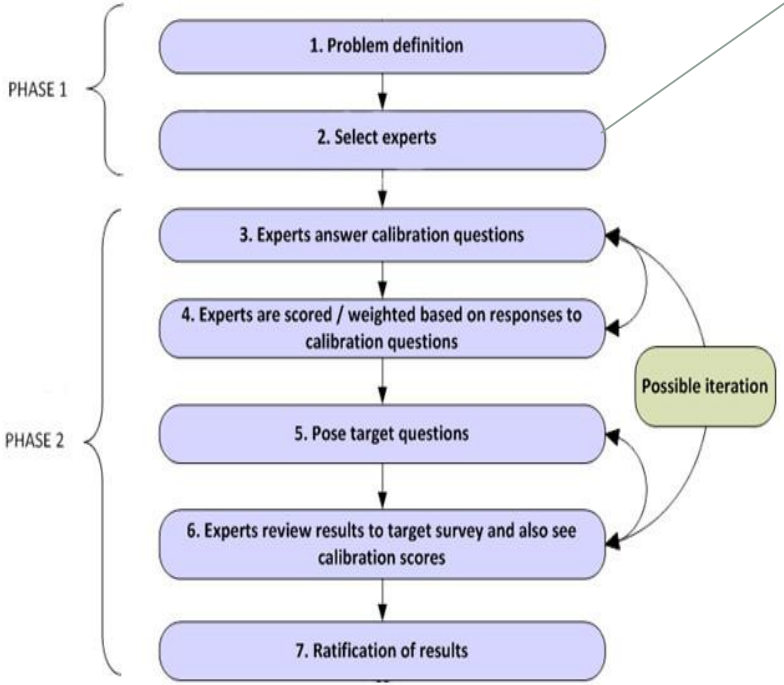
- degrees of belief on alternative conceptual models
- unknown/uncertain material quantities.

Performance-based EE include an empirical step of expert ranking, aimed at measuring their uncertainty quantification capabilities.

The EXCALIBUR software (http://www.expertsin-uncertainty.net/) was a pioneering tool to assess such performance weights following the so called 'Classical method'.



Performance-based EE



Phases of a structured elicitation session (courtesy of W. Aspinall).

At least >6 (better >10)
Basic background on at least one aspect of the problem
Worked in the study area (most)

Questionnaires:

- The **seed questions,** with known answers.
- The **target questions,** i.e. the questions of interest**.**

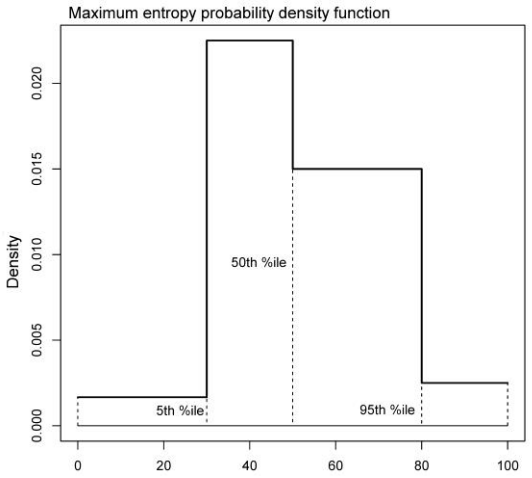
For each question, the experts express their views as the values of the **5th, 50th and 95th percentiles** of a probability distribution representing their uncertainty.

The **seed questions** should be similar as much as possible to the target questions.
They are adopted to score the experts' uncertainty estimation performances.

The diverse answers to the **target questions** are then pooled using the obtained scores, and their combination defines a new virtual expert:
the **global Decision Maker (DM)**.

Uncertainty distributions examples

Examples of probability distributions.



We generally use CM as a reference and also compare it to EW and other scoring rules.

The Cooke classical method (CM) uses uniform PDF in each inter-percentile range, i.e. maximum entropy distributions. (Cooke, 1991)

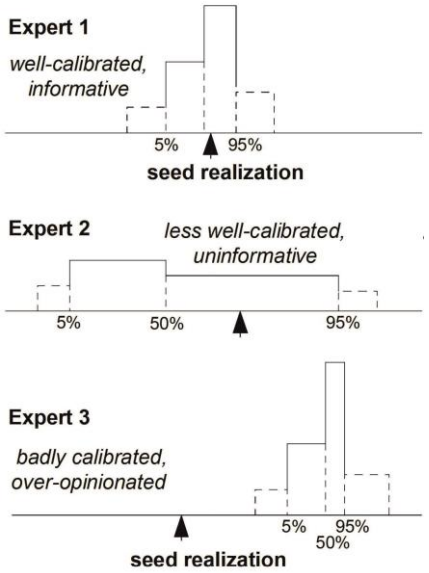
The performance score in CM is the product of two values:

Calibration score: likelihood that the true results correspond to the expert distributions. It is a statistical accuracy.

Informativeness score: average relative information w.r.t. a uniform distribution. It penalizes too large uncertainty ranges.

Alternative methods implement different **scoring rules** and **uncertainty distributions**

Maximum entropy distributions associated to different performances.



Experts' pooling: the Decision Maker (DM)

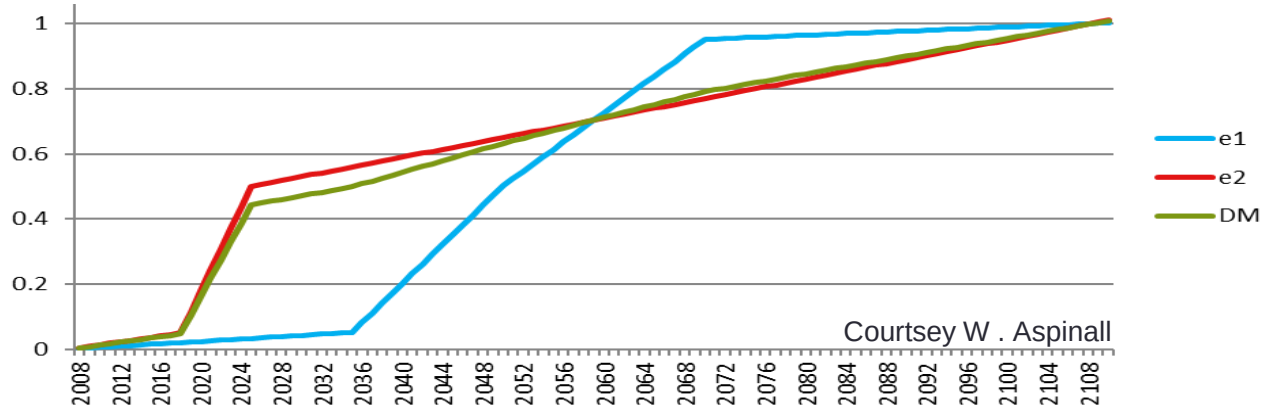
Experts' answers are **pooled together** according to the weight w_i of each expert e_i

The DM is typically defined by the **weighted linear combination** of the probability distributions of all the experts, i.e. by a probability mixture.

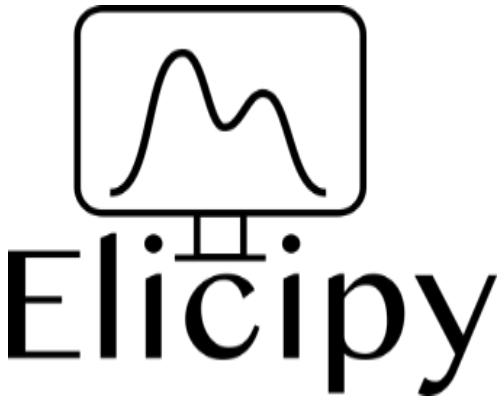
In fact, the **statistical sampling of the DM** is performed by randomly choosing one of the experts, with a chance proportional to their weight, and then by sampling their distribution.

Example of expert pooling and DM's definition by a probability mixture.

$$w_1 = 0.12$$
$$w_2 = 0.88$$

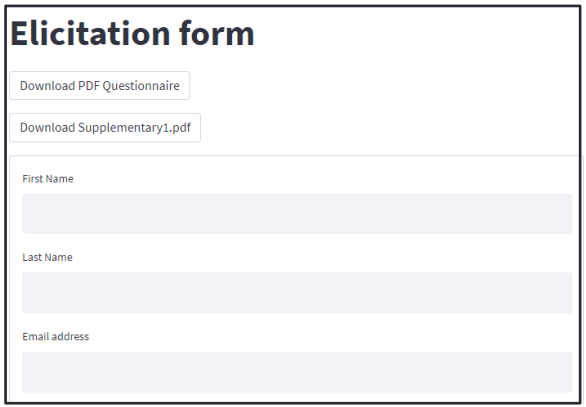


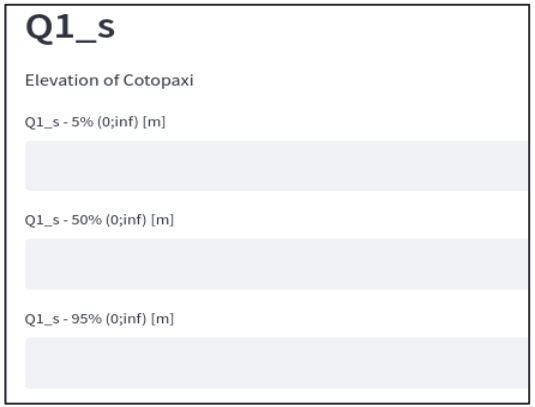
NEW TOOL FOR EE: ELICIPY

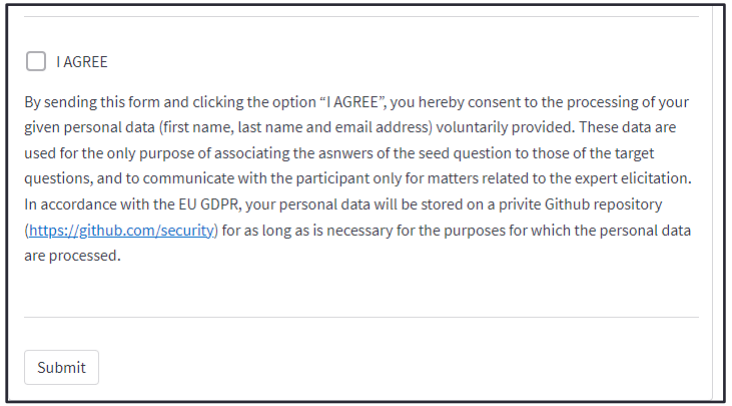


Questionnaire design

Example of the online forms





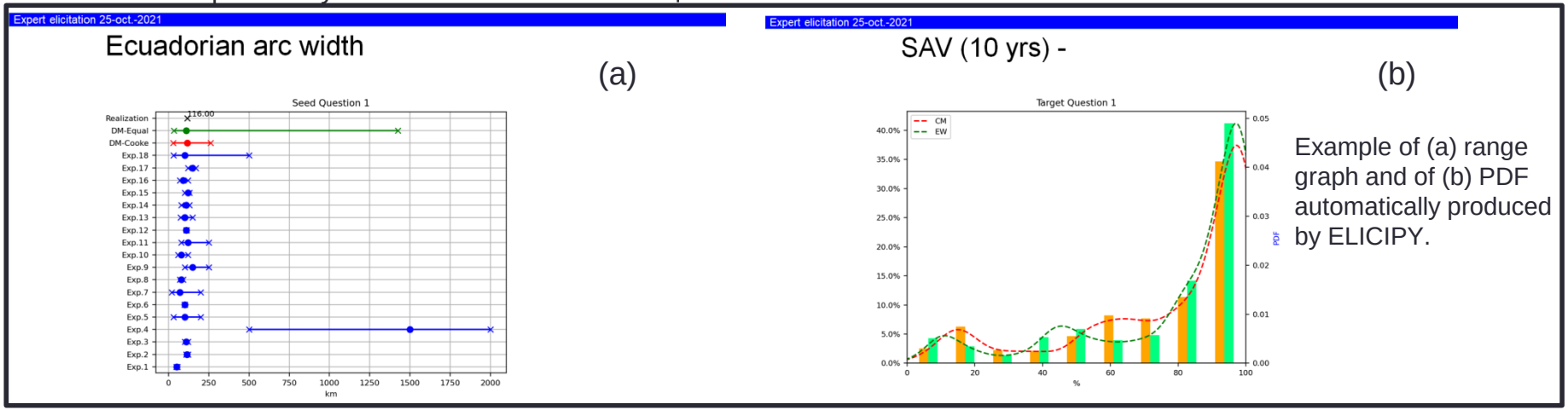


- Online questionnaires (seed/target) answered by each expert
  - Saved into online encripted repository into standard csv files
- Controls on experts' answers (numeric values, within bounds, sum to 100, increasing percentile values)
- Images added for some of the question
- Download pdf with questions/supporting information
- csv with answers sent to expert's email

Results analysis

- Csv outputs from webforms → assembled by the analysis tool into a single csv file.
- Elicipy based on the combination of the Cooke's method scripts of the open source package Anduril (CM,EW) and the R scripts of INGV Pisa (that we typically use in data processing). All translated in Python language.
- Outputs:
  - itemwise range graphs,
  - statistical sampling of the DM responses
  - PDF and histograms for each question
  - text files retro-consistent with previously existent EE software

AUTOMATIC PRODUCTION → Power point presentation



Example of (a) range graph and of (b) PDF automatically produced by ELICIPY.

STROMBOLI ELICITATION

with Mattia de' Michieli Vitturi, Andrea Bevilacqua, Alessandro Tadini, Tomaso Esposti Ongaro, Augusto Neri, Matteo Cerminara, Emmie Bonilauri, Andrew Harris, Raphaël Paris, Marco Pistolesi, Willy Aspinall

Expert elicitation at Stromboli: target questions/1a

Part I – number of tsunamigenic landslides

- VIII century CE – 1878
- 1879 – Present day
- Within the next 50 years

→ *DM answers translated into annual rates by analysts*

Annual rate (documented events) for VIII century CE-1878

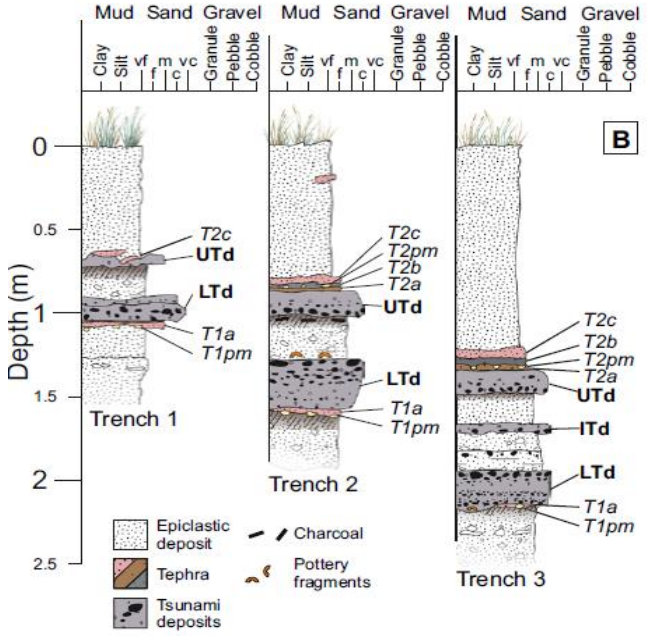
3 tsunamigenic landslides/1177 years = 0.002

If we apply the same annuale rate (0.04) of the period 1879-Present day (see following slide) we obtain, for VIII century CE-1878:

≈ 50 tsunamigenic landslides

XIV-XV century CE

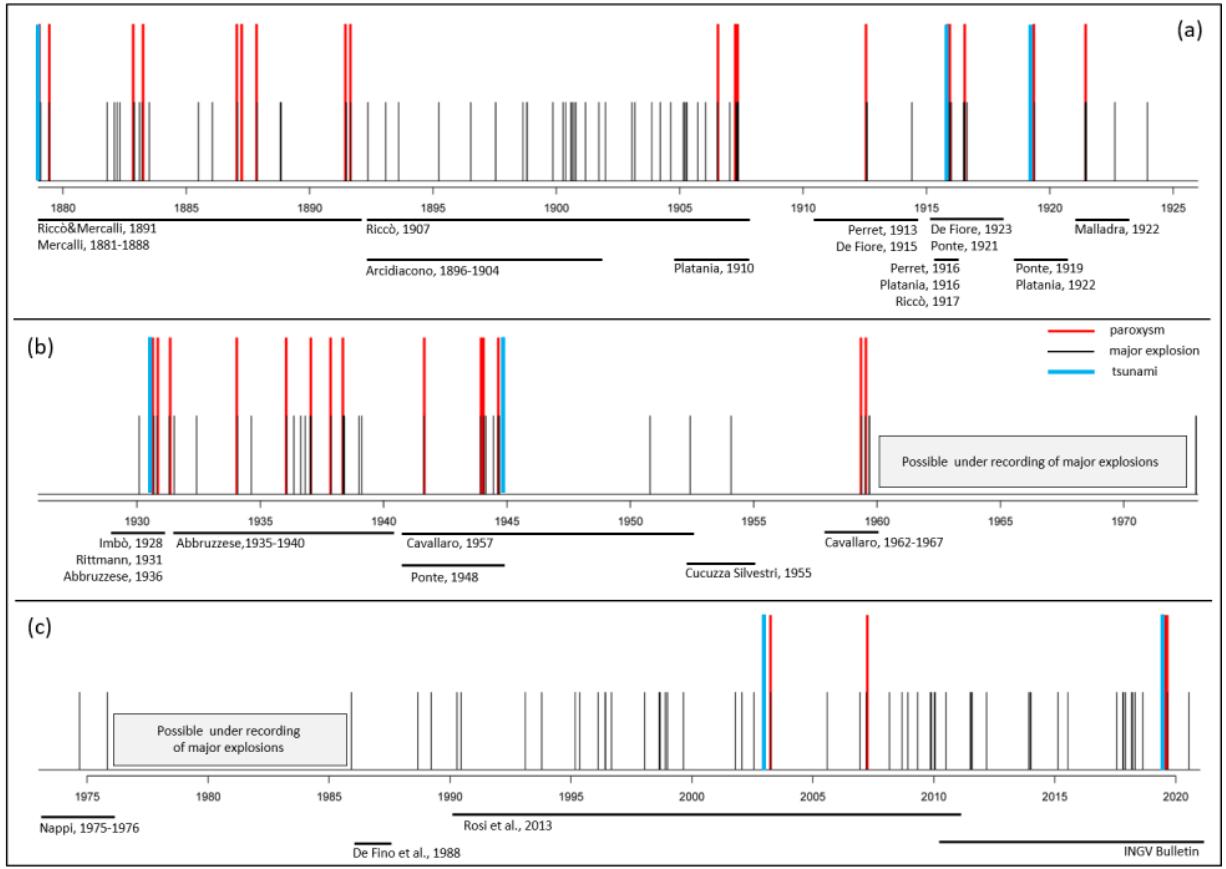
Volume (LTd+ITd+UTd): ≈ 180 x 10^6 m^3



Tsunami and tephra sequences in trenches (LTD—lower tsunami deposit; ITD—intermediate tsunami deposit; UTD—upper tsunami deposit). [Pistolesi et al., 2020]

Expert elicitation at Stromboli: target questions/1b

Historical catalog of major explosions and paroxysms and tsunami at Stromboli in [1879, 2020]



Annual rate (documented events) 1879-Present (nov-2022):

7 tsunamigenic landslides/143 years = 0.04

↑

7 TL considered:
1879,1916,1919,1930,1944,2002,2019

↓

< 1 Mm^3?

Uncertain TL (not considered):

- 1887?
- 1954?

LAST TL: dec-2022

- Tsunami wave of 1.5 m (front of the SdF)

Major explosions (black lines), Paroxysms (red lines) and tsunamis (blue lines) at Stromboli, 1879-present day [modified from Bevilacqua et al., 2020]

Expert elicitation at Stromboli: target questions/2

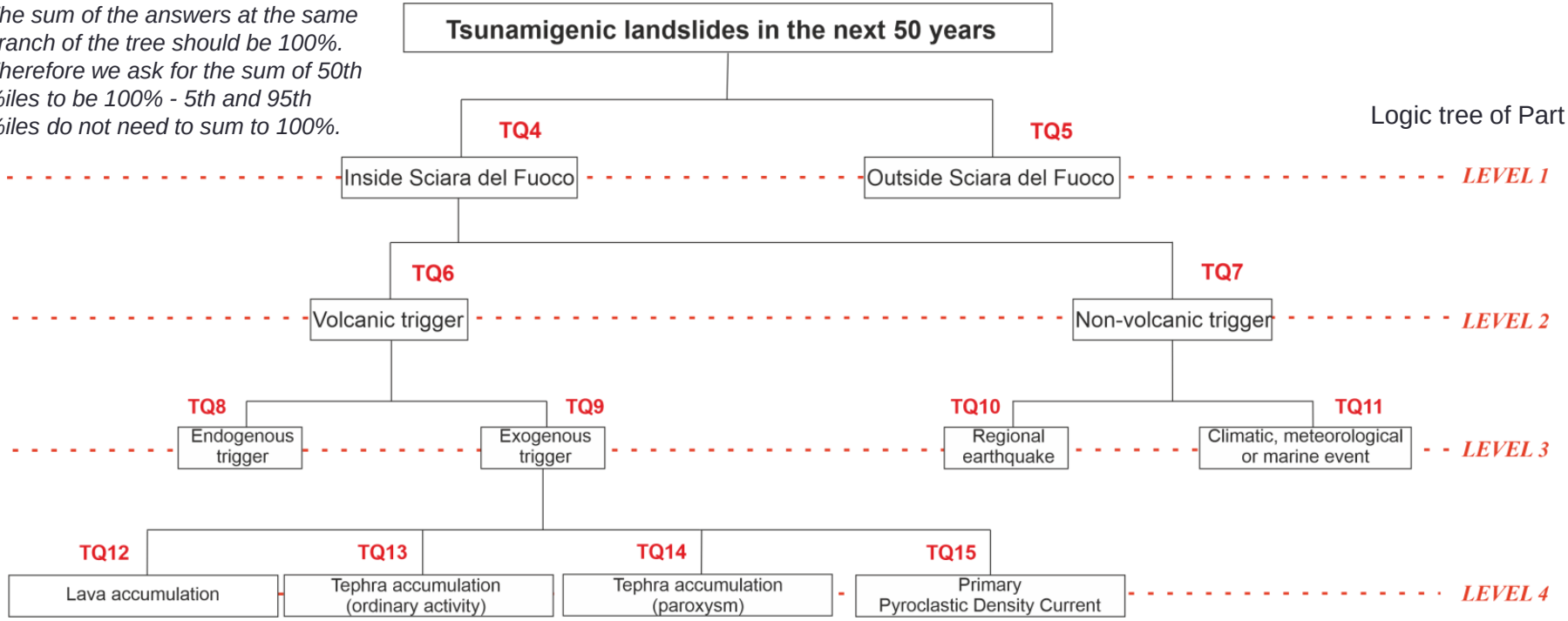
Part II – triggering conditions of the tsunamigenic landslide

These questions define *conditional probabilities*, e.g. P12= P(Lava Accumulation | Exogenous trigger).

In postprocessing we will calculate *absolute probabilities*, e.g. P(Lava accumulation) = P12 * P9 * P6 * P4.

The sum of the answers at the same branch of the tree should be 100%. Therefore we ask for the sum of 50th %iles to be 100% - 5th and 95th %iles do not need to sum to 100%.

Logic tree of Part II



Expert elicitation at Stromboli: target questions/3

Part III - spatial location and volume of the tsunamigenic landslide
- Simulations performed at INGV Pisa
- Only along the Sciara del Fuoco (SdF)

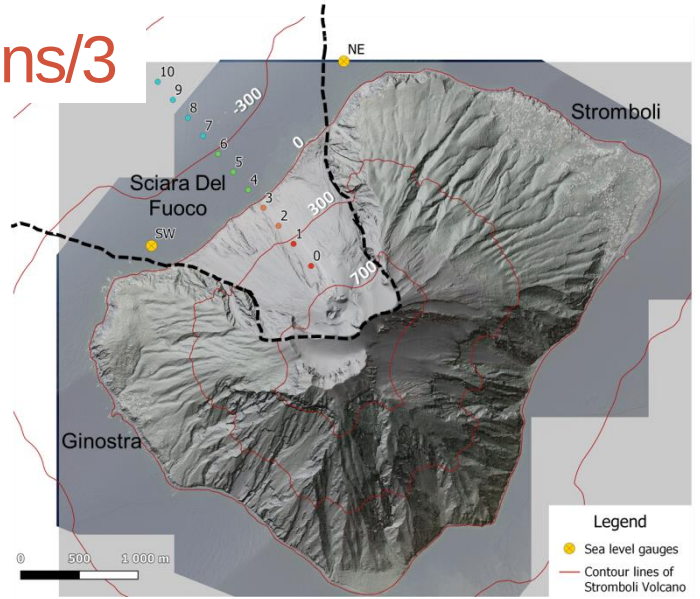
Four volume classes

V1 = {1 < Volume ≤ 5}
V2 = {5 < Volume ≤ 14}
V3 = {14 < Volume ≤ 30}
V4 = { Volume ≥ 30}
x 10^6 m^3

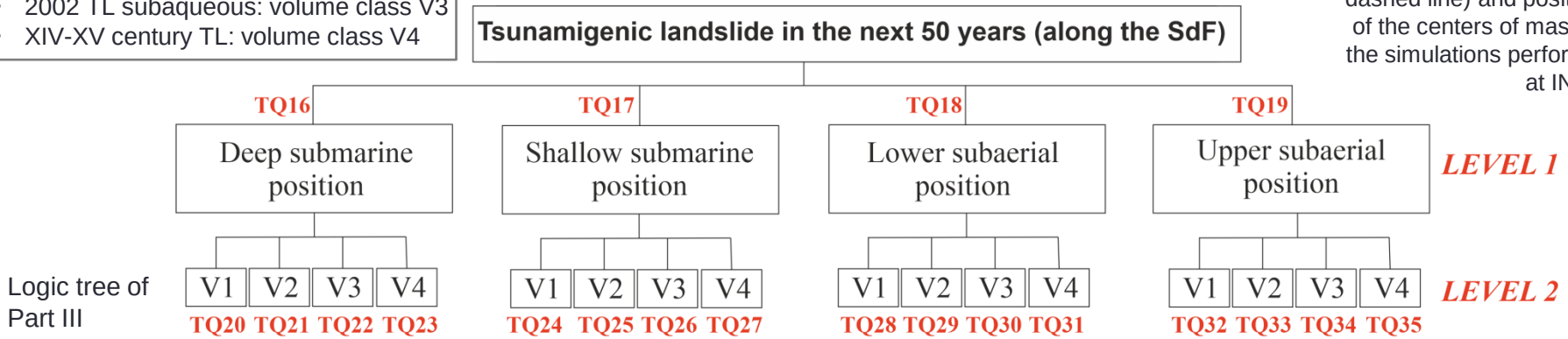
Four spatial classes

Deep submarine	700-300 m BSL
Shallow submarine	300-0 m BSL
Lower subaerial	0-300 m ASL
Upper subaerial	300-700 m ASL

Possible volume class examples:
- 2019 (?) & 2022 TL: volume class V1
- 2002 TL subaerial: volume class V2
- 2002 TL subaqueous: volume class V3
- XIV-XV century TL: volume class V4



Sciara del Fuoco (black dashed line) and positions of the centers of mass for the simulations performed at INGV.

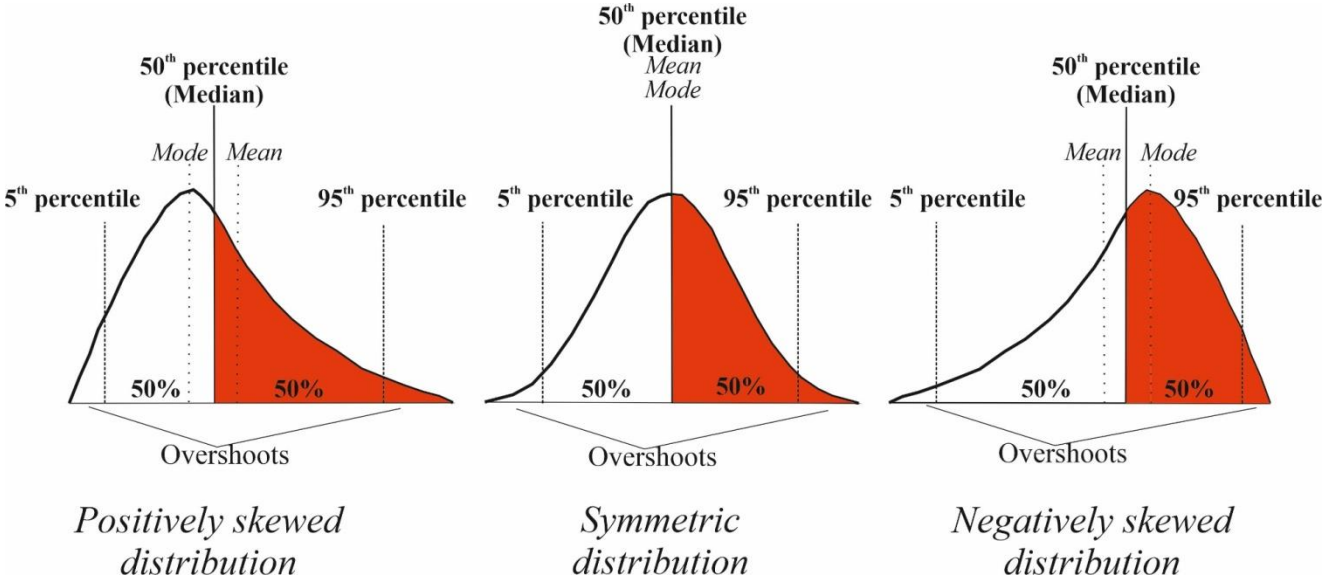


Logic tree of Part III

Tips for answering at the questions

Remember that the three percentiles that you provide defines a probability distribution:

- Option 1: define first the 50th percentile (distribution divided in half) and then add the two upper and lower bounds (5th/95th percentiles) - *this is the best strategy to adopt when groups of medians must sum to 100%*.
- Option 2: define first your uncertainty range (5th/95th percentiles) and then divide the whole distribution in half



Provide increasing values of percentiles (CONTROL)

Look carefully at the physical unit (and the reference for seed qs.)

Write first on a hard copy and then on the online form

Consider overshoots, i.e. the 5% chances that the response is greater than the 95th%ile, or smaller than the 5th%ile.

Do not provide same values for some (or all) percentiles (CONTROL)

THANKS FOR YOUR ATTENTION!

Questions?