18TH WORLD CONFERENCE ON EARTHQUAKE ENGINEERING
WCEE2024
MILAN

MILAN, ITALY
30TH JUNE - 5TH JULY 2024

www.wcee2024.it

PROPAGATING UNCERTAINTY IN TSUNAMIGENESIS THROUGH TO DAMAGE AND LOSS USING DATA REDUCTION

C.I. Nievas[1], A. Androsov[2], S. Harig[2], J. Karstens[3],
Y. Ren[3], M. Urlaub[3] & G. Weatherill[1]

[1] GFZ German Research Centre for Geosciences, Potsdam, Germany, cecilia.nievas@gfz-potsdam.de

[2] Alfred Wegener Institute (AWI) Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany

[3] GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany

Abstract: *The "risk workflow for CAScading and COmpounding hazards in COastal urban areas" (CASCO) project focuses on a scenario of cascading geo-hazards and compounding climatic events in Sicily, southern Italy, to facilitate the development of a workflow that allows us to propagate different sources of uncertainty along the full hazard, damage and loss chain. The scenario begins with a Mw 7.1 earthquake in the Strait of Messina, which triggers a co-seismic tsunami and a submarine landslide on the steep continental slope of East Sicily, already known to be unstable, which also contributes to the tsunami generation. The adoption of a fixed magnitude and rupture does not avoid the uncertainty associated with the distribution of slip within the rupture, which is already known to have a significant influence on the resulting tsunami. The location and size of the triggered landslides are unknown a priori as well. A classic logic tree approach could be used to propagate these uncertainties along the whole chain, but the possible combinations stemming from different ground motion fields, rupture slip patterns and potential landslides quickly leads to a very large number of branches. To address this challenge, a first step has been to generate 500 stochastic realisations of rupture slip and subsequent ground surface deformation to obtain realisations of tsunami inundations due to the earthquake trigger. At the same time, 54 physically possible submarine landslides have been defined along the eastern coast of Sicily and their occurrence and subsequent tsunamis modelled. The resulting distributions of maximum flow depth across a large sample of coastal points are rendered into low-dimensional 2D space using the Sammon's Mapping technique. The model space is partitioned into smaller segments, which allows us to identify a reduced number of realisations and assign them weights that would allow them to capture centre, body and range of the distributions of tsunami parameters implied by the full set of realisations. This data reduction allows us to use a more manageable number of branches to proceed to the next step in the chain. Results show that the method is successful in achieving such a goal, not only in terms of inundation parameters but also in terms of expected damage and loss ratios, thus enabling the combination and propagation of a greater variety of uncertainties into a multi-hazard risk analysis.*

1 Introduction

The extreme-event scenarios of the CASCO project combine a series of cascading geo-hazards and compounding climatic events in Sicily, southern Italy. A M_w 7.1 earthquake in the Strait of Messina results in strong ground motions that cause widespread damage, while at the same time triggers a co-seismic tsunami and a submarine landslide on the steep continental slope of East Sicily, already known to be unstable. The

submarine landslide further contributes to the tsunami generation, with the final tsunami waves resulting from the overlap of both tsunamigenic sources and finally reaching the coast and further stressing the population and the built environment. On the climatic side, one of the scenarios includes the occurrence of a heatwave in the whole island of Sicily, while another one considers the occurrence of extreme rainfall that leads to flash floods and debris floods in several catchments to the south of Messina; each of these two alternative climatic scenarios occur in a very quick succession with respect to the earthquake, submarine landslide and tsunami, causing an interaction in their consequences and effects on the built environment and the population. Subaerial landslides are triggered by both the earthquake action and the extreme rain.

The earthquake-followed-by-tsunami component of the scenario resembles the earthquake-tsunami sequence that occurred on 28 December 1908 in the exact same area, for which records exist of around 600 km^2 (including 92 localities) of macroseismic intensities of IX and X in the Mercalli-Cancani-Sieber (MCS) scale, and around 80,000 deaths, large numbers of injured, and over 120,000 people left homeless (Guidoboni et al., 2019). It has been hypothesised that the 1908 event also involved a submarine landslide contributing to the tsunami generation (e.g., Favalli et al., 2009). It is not the intention or purpose of the project to reproduce the 1908 event, and the use of the 1908 Messina earthquake rupture as a starting point is simply a strategy for the selection of the scenario. Though the scenario events will be based on projected future weather conditions, extreme rainfall in Messina has already caused floods and debris floods in the past, more specifically on 1 October 2009 in the catchments of Racinazzi and Giampilieri, which resulted in 38 casualties and significant damage to buildings and road infrastructure amounting to around 550 million Euro (Aronica et al., 2012).

Building a calculation chain including several cascading and/or compounding hazards and estimating damage to buildings and human casualties is challenging in many ways, one of which is the propagation of modelling uncertainties associated with each of the components. For example, the adoption of a fixed earthquake magnitude and rupture scenario *a priori* does not eliminate the uncertainty associated with the distribution of slip within the rupture, which is already known to have a significant influence on the resulting tsunami (e.g., Løvholt et al., 2012; Goda et al., 2014). Focusing on the tsunami generation and propagation, some of the uncertainties to investigate include:

- the distribution of slip within the earthquake rupture, which translates into sea-floor displacements that trigger the earthquake component of the tsunami;
- inland Manning friction coefficients (e.g., Harig et al., 2022);
- tidal scenario (spring/neap tides);
- location and volume of the submarine landslides;
- material properties (density, cohesion) of the continental slope of East Sicily where the submarine landslides would occur;
- time lags between earthquake trigger and landslide trigger of the tsunami.

The remaining components of the CASCO hazard-to-damage-and-loss calculation chain include, naturally, their own sources of epistemic and aleatory uncertainty. A classic logic tree approach could be used to propagate these uncertainties along the whole chain, but the number of possible combinations increases very fast with each branching level. While the calculation demand stemming from some of these might be manageable (for example, the uncertainty in earthquake ground motions will be treated in a stochastic way), the tsunami simulations are particularly resource-consuming. A strategy for propagating these uncertainties while keeping the computational demand constrained has thus been investigated and is hereby presented.

This paper presents in detail the steps taken for the earthquake-triggered component of the tsunami. Analogous steps will be followed for the landslide-triggered component. Of the several uncertainties listed above, this paper focuses on the impact of the distribution of slip within the earthquake rupture on the resulting tsunami and its consequences. A total of 500 stochastic realisations of rupture slip have been generated and lead to 500 seismically-driven sea-floor displacement fields, which are ultimately used to trigger 500 possible resulting tsunamis. A series of randomly selected points of the tsunami mesh are used to characterise each realisation in terms not only of maximum flow depth (inundation height) and maximum absolute velocity but also resulting damage and economic losses. The aim is to quantify how much each realisation differs from the rest and to be able to select a smaller number of samples to which weights can be assigned, so that the centre, body and range of the tsunami inundation depths and velocities, as well as resulting damage and economic losses, remains similar to those obtained with the original 500 realisations. This is achieved by reducing the

multi-dimensionality of the problem by means of Sammon's mapping, following the work done by Goulet et al. (2021) to capture epistemic uncertainty in earthquake ground motions for the NGA-East model in the USA. As the ultimate goal is to avoid needing to carry out computations for all possible combinations of all uncertainties to be included in the modelling, the evaluation of the appropriateness of the method is carried out considering damage and losses stemming from the tsunami alone, not yet in combination with the earthquake. The model distance between empirical cumulative density functions (CDF) of the calculated hazard, damage and loss parameters obtained using the selected 13 vs the original 500 realisations according to the Wasserstein metric is used as a measure of the resulting fit, considering geographic locations both used and not used for the selection of the realisations.

2 Earthquake-tsunami generation, damage and losses

2.1 Slip on the earthquake rupture and sea-floor displacements

The preferred rupture definition of the Italian Database of Individual Seismogenic Sources (DISS, 2018; rupture ID ITIS013) for the 1908 M_w 7.1 Messina earthquake has been selected for the CASCO scenario. However, the adoption of a fixed magnitude and rupture geometry for the modelling of a future earthquake does not eliminate the uncertainty associated with the distribution of slip within the rupture, which is largely determined by the stress conditions and the frictional properties of the fault before the occurrence of the earthquake. Previous studies have shown that the complexity resulting from these factors can be well represented by means of random field models (e.g., Goda et al., 2014). Here, the approach of Graves and Pitarka (2010) and Mai and Beroza (2002) is followed to generate 500 spatially-correlated normalised random fields, which, combined with the scaling relation of Wells and Coppersmith (1994), leads to 500 samples of slip distribution on the rupture, discretised into grid cells of 1 km x 1 km. All samples are scaled to match the target seismic moment and then translated into 500 realisations of ground surface deformation on land and at the bottom of the sea, using the model of Okada (1985, 1992).

2.2 Tsunami simulation

Tsunami simulations are run for each of the 500 realisations of sea-floor displacements using the TsunAWI software (Rakowsky et al., 2013), which implements the finite element method and non-linear shallow water equations augmented by parametrisations of bottom friction (0.025 Manning coefficient used) and viscosity (Smagorinsky, 0.3 used). The total duration of the tsunami simulation is 2 hours, run with a time step of 0.1 seconds. A variable-size triangular mesh with edge lengths between 6.2 m (on land) and 8.9 km (in the open sea) was defined using the 2008 General Bathymetric Chart of the Oceans grid (GEBCO_08, https://www.gebco.net/) combined with the TinItaly topography (Tarquini et al., 2023), and controlling the resolution as a function of bathymetric/topographic depth and slope.

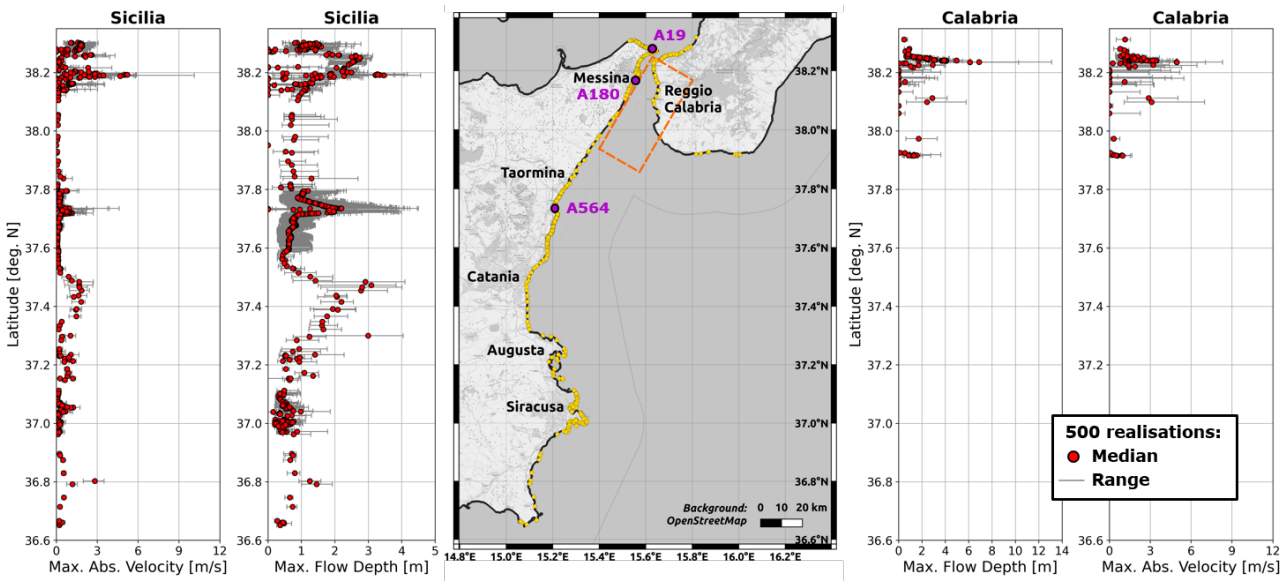

Figure 1. Maximum flow depths and maximum absolute velocities for randomly selected points along the coast of the regions of Sicily (left) and Calabria (right). Earthquake rupture shown in orange.

A series of 846 points from the mesh have been randomly selected along the coast to represent the output of each tsunami realisation. The random selection was done forcing the code to increase the sample density in areas with larger concentrations of population, which were observed to be correlated with areas where the mesh resolution was higher, while still sampling the more scarcely-populated areas. Sampling weights were assigned to the mesh points as a function of the inverse of their (squared) distance to the 0-meter contour line, which resulted in most sampled points lying on the contour line itself, or very close to it (i.e., at the boundary of the onshore-offshore domains). Figure 1 shows the maximum flow depths and maximum absolute velocities obtained for all 500 realisations (grey lines) as well as their median (red dots), for each of the geographic locations (the points on the Sicilian coast are depicted to the left and those on the Calabrian coast are depicted to the right of the map; some points have been omitted for clarity). As can be observed, the resulting variability can be large. Moreover, the range and shapes of the resulting distributions vary significantly for different geographic locations (e.g., Figure 2).

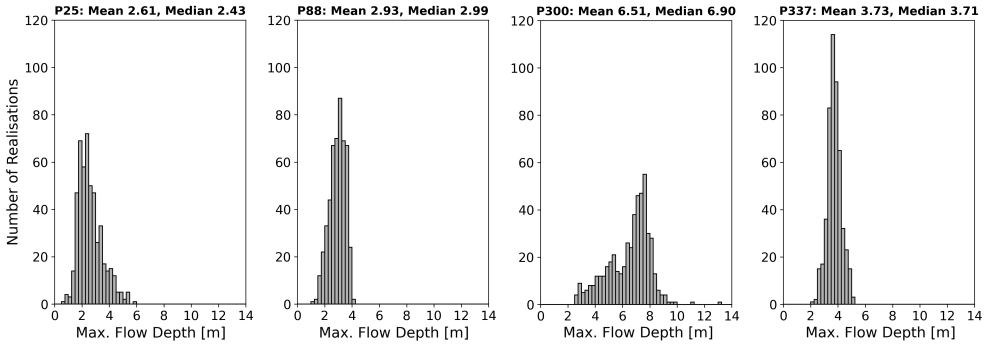

Figure 2. Distributions of maximum flow depths at four of the 846 randomly selected points.

2.3 Damage and economic losses

At the time of writing, the full exposure model for the Sicilian and Calabrian coast is not yet finalised within the CASCO project. For this reason, damage and economic losses are calculated at each of the 846 randomly selected coastal locations assuming 100% of the building stock is composed of residential masonry whose fragility to tsunami action can be represented by the model of Belliazzi et al. (2021). The model of Belliazzi et al. (2021) uses maximum inundation/flow depth as the intensity measure and considers six damage states: no damage (DS0), light damage to non-structural elements and negligible damage to structural elements (DS1), slight damage to structural elements (DS2), average damage to non-load-bearing structural elements (DS3), and high damage to load bearing structural elements without (DS4) and with (DS5) floor loads. Figure 3 depicts the fragility curves associated with each damage state, compared against the distribution of inundation depths (stemming from the 500 tsunami realisations) obtained for three of the selected geographic points. As can be observed, the variability of inundation depths will inevitably be reflected on the variability in the resulting damage and losses, and the connection between the two is non-linear, due to the shape of the fragility curves as well as the distributions and values of inundation depths. The earthquake action is ignored for this calculation, as considering cumulative damage from both hazards would require the combination of all the stochastic ground motion fields being generated for the earthquake (15,000) with the 500 tsunami realisations, which is the sort of computational demand to be minimised by means of this procedure.

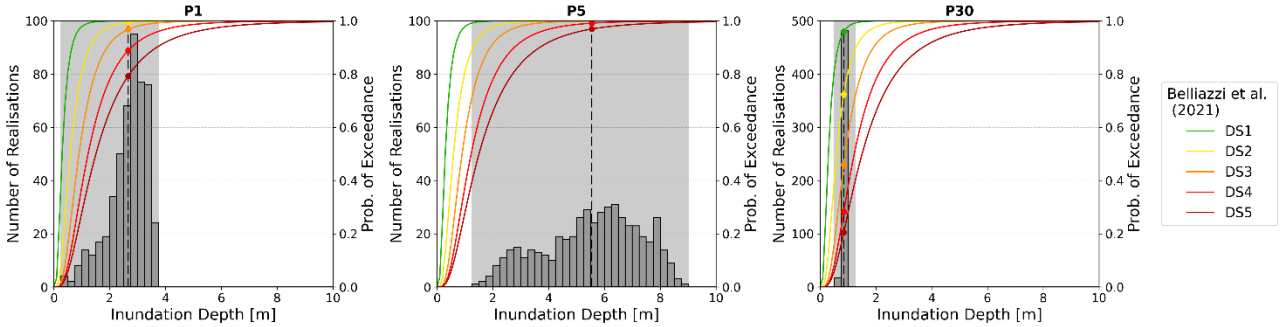

Figure 3. Fragility model of Belliazzi et al. (2021), superimposed to the distributions of maximum inundation depth for three selected geographic locations.

A preliminary economic consequence model is defined based on the European Seismic Risk Model 2020 (ESRM20; Crowley et al., 2021) and Gomez-Zapata et al. (2021) as follows: 5% loss ratio for damage state DS1, 15% for DS2, 45% for DS3, 85% for DS4 and 100% for DS5. A final economic consequence model will be decided upon later in the project. Economic loss ratios are defined as the quotient between the cost of repairing a building in a particular damage state and the total value of the building.

Figure 4 shows the loss ratios obtained for all 500 realisations (grey lines) as well as their median (red dots), for each of the 846 points. The summation of the probabilities of damage states 2 and 3 are shown as well. Plots showing the summation of the probabilities of damage states 4 and 5 as well as plots showing the mean damage state show a pattern very similar to those of the loss ratios.

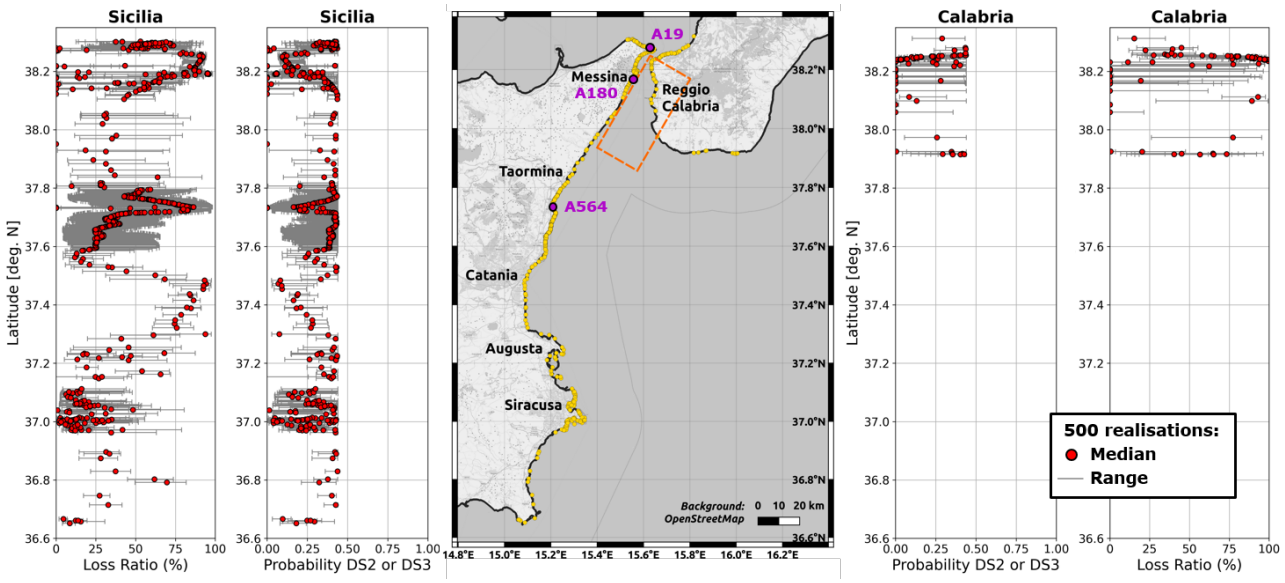

Figure 4. Loss ratios and probabilities of DS2 or DS3 for randomly selected points along the coast of the regions of Sicily (left) and Calabria (right). Earthquake rupture shown in orange.

3 Data reduction

3.1 Selection of realisations by means of Sammon's Mapping

Due to the spatial extent and variability of the tsunami waves and inundation, the comparison of one tsunami realisation against another is a high-dimensional problem. As a first step, a series of 846 points from the tsunami mesh are randomly selected along the coast to represent the output of each tsunami realisation, as explained earlier. Nevertheless, for each potential variable of interest (maximum flow depth, maximum absolute velocity, probability of different damage grades, expected loss ratio), there are 500 possible values for each of the 846 geographic points. The Sammon's Mapping (Sammon, 1969) technique is thus applied in order to be able to visualise the whole model space in two dimensions, with the purpose of, firstly, understanding the relative distribution of the 500 tsunami realisations within it and, subsequently, selecting a smaller number of realisations that could be used to represent the whole model space.

The procedure used conceptually follows the work of Goulet et al. (2021) and starts by normalising the maximum flow depths (h_{Pi}) for each geographic point P_i and tsunami realisation in terms of the mean (μ_{Pi}) and standard deviation (σ_{Pi}) of all 500 realisations at point P_i, as per Equation (1):

$$h_{norm\,Pi} = \frac{h_{Pi} - \mu_{Pi}}{\sigma_{Pi}} \quad (1)$$

The normalised maximum flow depths are then processed using the Sammon's Mapping algorithm[1], which iteratively calculates 2D coordinates for each of the 500 tsunami realisations so as to minimise the misfit function, which is named "Sammon's stress" (E) and is defined as the difference between the distance between

[1] Python implementation of the Sammon's Mapping algorithm by T.J. Pollard: https://github.com/tompollard/sammon.

two tsunami realisations *m* and *n* in the high-dimensional space ($\overline{\Delta}_{HD\,mn}$) and the distance between the same two tsunami realisations in the 2D space ($\overline{\Delta}_{map\,mn}$), as per Equation (2):

$$E = \frac{1}{\sum_{m<n} \overline{\Delta}_{HD\,mn}} \cdot \sum_{m<n} \frac{\left(\overline{\Delta}_{HD\,mn} - \overline{\Delta}_{map\,mn}\right)^2}{\overline{\Delta}_{HD\,mn}} \quad (2)$$

The resulting Sammon's map is shown in Figure 5a, where each point represents one of the 500 realisations (the 846 geographic locations can be thought of as being condensed in each point of the plot). As can be inferred, the value of the coordinates do not have any physical meaning, as it is the relative distance between two points (i.e., two tsunami realisations) that matters. As no evident clustering or sub-grouping of points can be observed, the model space is subdivided into 13 segments. The centroid of all points belonging to each segment is calculated and the point closest to the centroid (the so-called *medoid*) is identified. The 13 medoids are the 13 tsunami realisations selected to represent the tsunami inundation by means of a logic tree, with weights calculated proportionally to the number of points in each segment (shown in red in Figure 5a). Figure 5b and 5c show where these 13 realisations fall within the distribution of maximum flow depths: in (b), the spatial mean of the 846 geographic points for each realisation is displayed, while (c) focuses on the spatial maxima. The choice of the number of segments and their shape is arbitrary, and its adequacy for the selection of a reduced number of tsunami realisations and calculation of weights is evaluated in the next section.

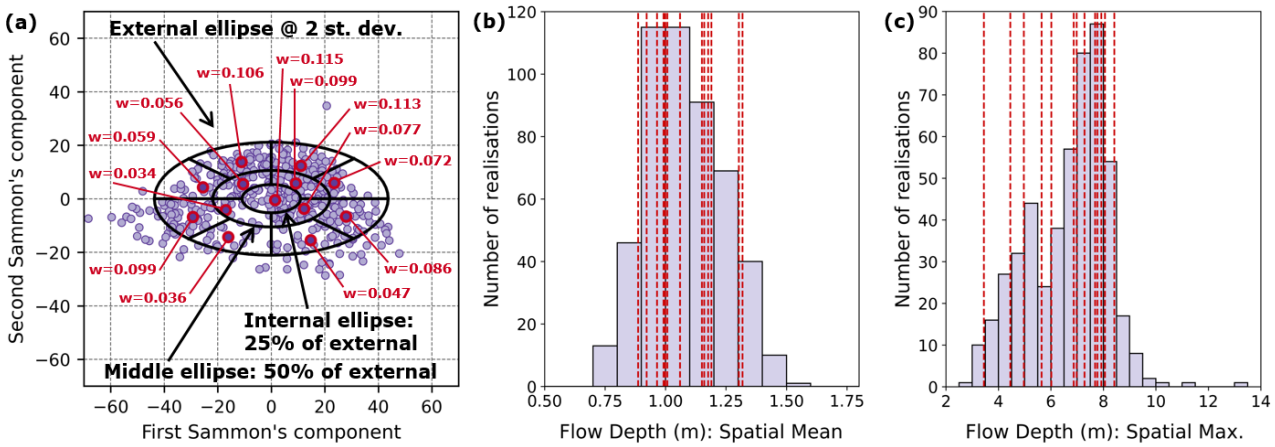

Figure 5. (a) Sammon's map representing the maximum flow depths of 500 tsunami realisations (purple points). Points marked in red are the medoids of the 13 segments defined, with their weights. (b) Distribution of mean (across all 846 geographic points) maximum (in time) flow depth of all 500 realisations; vertical lines correspond to the values of the 13 medoid realisations. (c) Same as (b), but for the maximum across all 846 geographic points.

3.2 Evaluation of the performance in the hazard, damage and loss domains

Having selected the 13 tsunami realisations based on the maximum flow depth observed for the 846 randomly-selected geographic points, the next step is to evaluate how well these 13 realisations (with their weights) can represent the original distributions of maximum flow depth, maximum absolute velocity, maximum flux, probabilities of damage and economic loss ratios, both for the same 846 geographic points as well as for a different set of 3,108 geographic points, randomly selected as well with the additional constraint that they don't include any of the 846 original points. In order to do this, empirical cumulative density functions (CDFs) are calculated and plotted for each geographic point using all 500 realisations (CDF_{500}), on the one hand, and the selected 13 realisations with their weights (CDF_{13}), on the other. Wasserstein distances between the two curves are calculated, previously normalising the CDFs by the maximum value of the parameter (e.g., flow depth, loss ratio) attained from the 500 realisations, so that the resulting values can be compared for different parameters (i.e., different units of measure). The Wasserstein metric is calculated as the area in between the two CDFs being compared (Equation 3):

$$WD = \int_{-\infty}^{+\infty} |CDF_{13} - CDF_{500}| \quad (3)$$

Original 846 geographic points, performance in the hazard domain

Figure 6 shows the results for one of the geographic points in which the 13 tsunami realisations with their weights (red lines) fit the whole distribution of maximum flow depth, maximum absolute velocity and maximum flux stemming from the 500 realisations (blue lines) best. For all three variables the normalised Wasserstein distance is 0.013. On the other end of the spectrum, Figure 7 presents the results associated with the worst-fitting geographic point for all three variables, with normalised Wasserstein distances of 0.122 (flow depth), 0.175 (absolute velocity) and 0.147 (flux). As shown in Figure 9 for the case of (normalised) maximum flow depth, such large values of the Wasserstein distances are very uncommon in the collection of 846 and 3,108 points. Figure 8 shows, as an example, the results for a geographic point with normalised Wasserstein distances around the mean/median values of the sets (see Table 1): 0.030, 0.035 and 0.035 for each of the three intensity measures. These combined results illustrate the overall good fit obtained with the selected 13 realisations in the hazard domain.

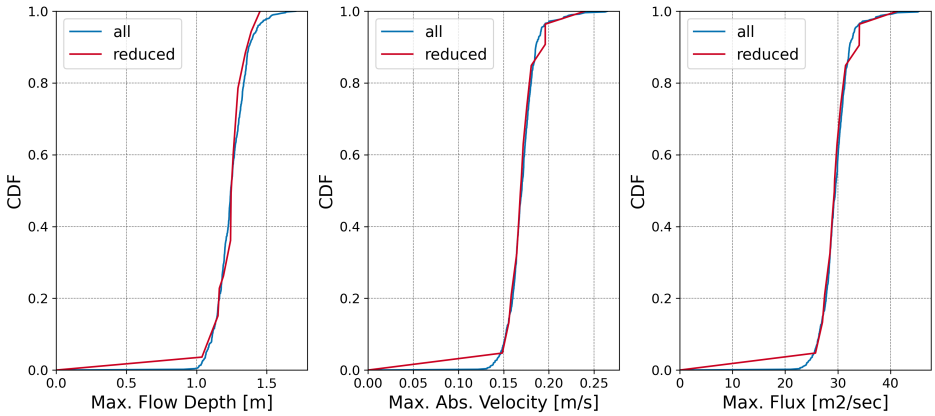

Figure 6. Example of a good fit of empirical CDFs of tsunami intensity measures from 500 tsunami realisations (blue) and 13 realisations with weights (red). Geographic point A180.

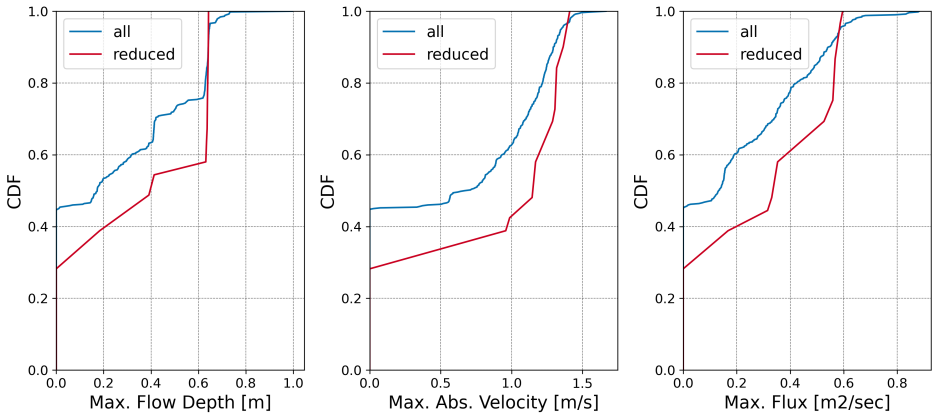

Figure 7. Worst fit of empirical CDFs of tsunami intensity measures from 500 tsunami realisations (blue) and 13 realisations with weights (red). Geographic point A19.

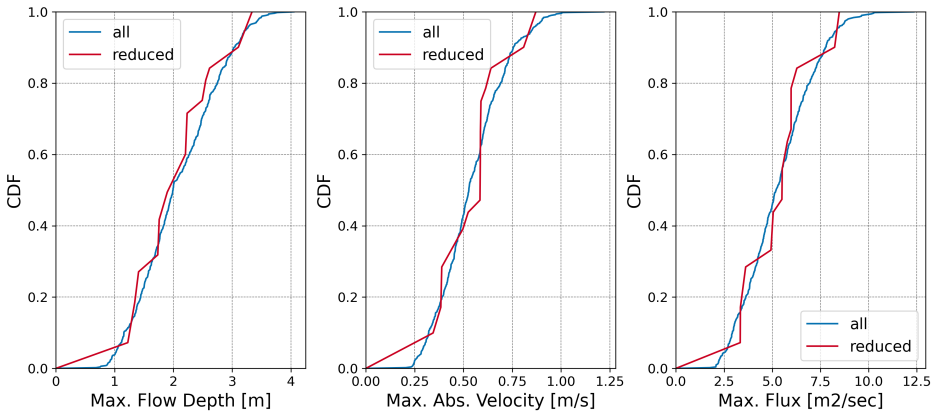

Figure 8. Example of fit of empirical CDFs of tsunami intensity measures from 500 tsunami realisations (blue) and 13 realisations with weights (red) for a case of normalised Wasserstein distances around the means and medians of the whole set (see Table 1). Geographic point A564.

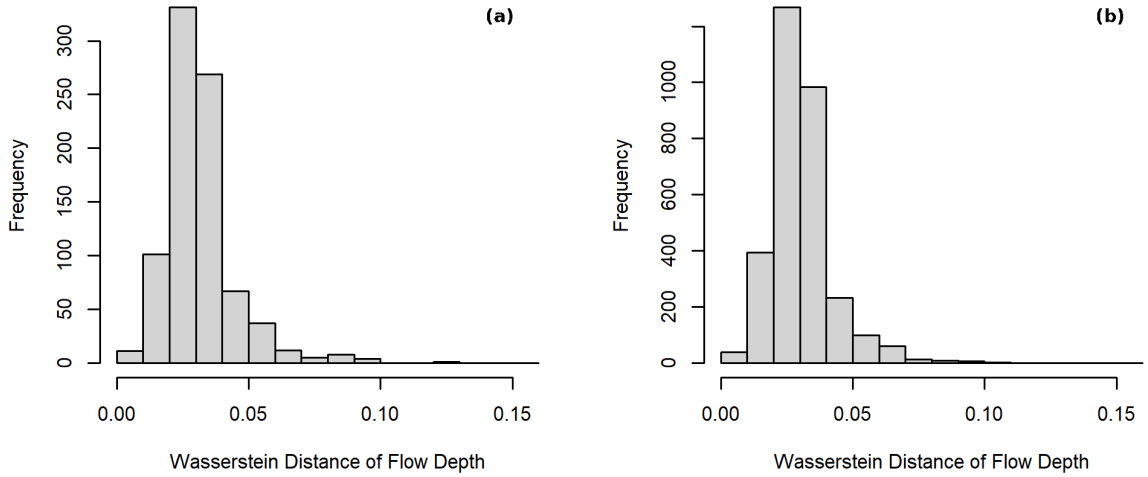

Figure 9. Wasserstein distances of the normalised empirical CDFs of maximum tsunami flow depth for (a) the original 846 and (b) the 3,108 newly-selected geographic points.

Original 846 geographic points, performance in the damage and loss domain

A high correlation can be observed between normalised Wasserstein distances of the maximum flow depths and those of mean damage state (Equation 4) and total loss ratio (Equation 5), indicating that geographic locations where the fit is good in the hazard domain also present good fits in the damage and loss domain. This is reflected in the mean, median and maximum values presented in Table 1 for these two variables. The normalised Wasserstein distances for the probabilities of occurrence of individual damage states tend to be slightly higher, particularly when compared against those for the maximum flow depth, which is the intensity measure used by the fragility model of Belliazzi et al. (2021) applied herein.

$$\overline{DS} = \sum_{k=0}^{5} k \cdot P[DS = DS_k] \tag{4}$$

$$LR = \sum_{k=0}^{5} LR_k \cdot P[DS = DS_k] \tag{5}$$

A great example of this correlation is point A19, the worst-fitting location in the hazard domain (Figure 7), which also presents the worst fit for the mean damage state and the ninth worst fit for the loss ratio (normalised Wasserstein distances of 0.128 and 0.106, respectively). Geographic point A180, which presented one of the best fits in the hazard domain (Figure 6), is ranked 102th (0.009) and 107th (0.013) for fitting mean damage

state and loss ratio, respectively. As depicted in Figure 10, these are still extremely satisfactory fits, and so are those for point A564, whose normalised Wasserstein distances in the hazard domain were around the means and medians of the whole set of points (Figure 8). In this last case, the Wasserstein distances for mean damage state and loss ratio are smaller than the means and medians of the set (0.021 and 0.027), while those for the probabilities of individual damage states have some slight variability (from DS0 through to DS5: 0.022, 0.034, 0.044, 0.036, 0.034, 0.035). As in the case of the hazard domain, the overall fit between the distributions obtained with the 13 against the 500 realisations is very satisfactory when looking at damage/losses as well.

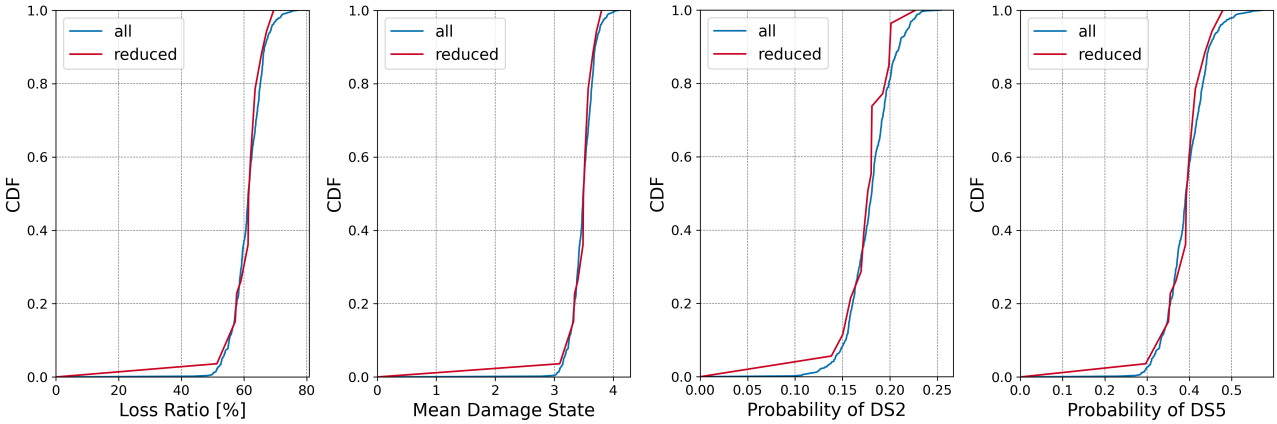

Figure 10. Example of a good fit of empirical CDFs of damage and loss ratios from 500 tsunami realisations (blue) and 13 realisations with weights (red). Geographic point A180 (see Figure 6).

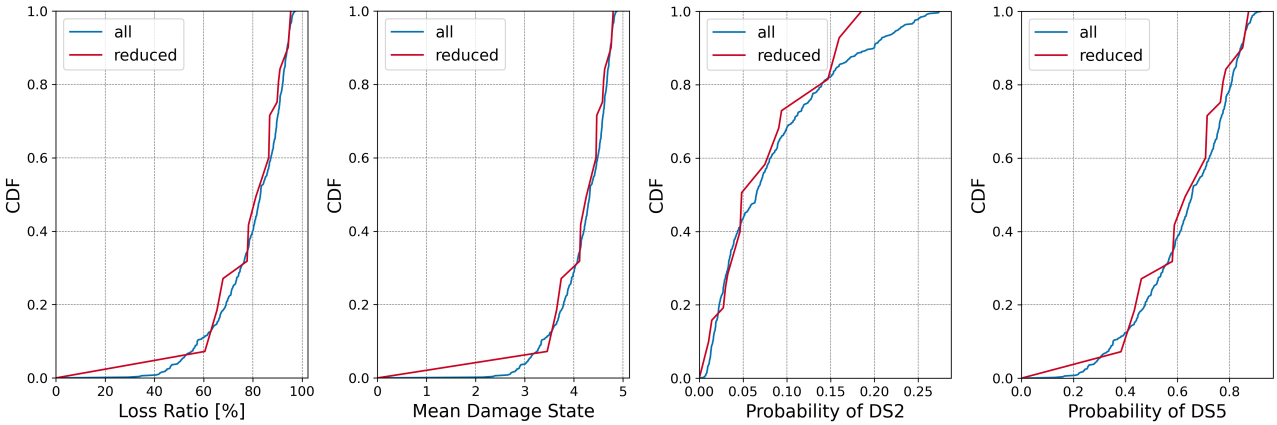

Figure 11. Example of fit of empirical CDFs of damage and loss ratios from 500 tsunami realisations (blue) and 13 realisations with weights (red) for geographic point A564 (see Figure 8).

Set of 3,108 newly-selected geographic points, performance in the hazard, damage and loss domains

The good fit observed when working with the 846 points originally used to select the 13 tsunami realisations and assign them weights remains when evaluating the performance of those 13 realisations using a set of 3,108 newly-selected geographic locations that exclude the original 846 (also located along or in the vicinity of the 0-meter contour line). Figure 12 through to Figure 15 compare the means, medians, standard deviations, 16th and 84th quantiles of the maximum flow depth, maximum absolute velocity, loss ratio (pattern very similar to mean damage state) and probability of occurrence of damage state DS4, and show a very good fit, with most points aligned around the 45-degree diagonal.

The distributions of the normalised Wasserstein distances for this set of 3,108 geographic points (see summary statistics in Table 1) are not significantly different with respect to those of the original 846 points. This suggests that evaluating the performance of the 13 tsunami realisations (with their weights) to represent the 500 realisations with a larger set of different geographic locations does not have a significant impact on the results

obtained, and the 13 realisations (with their weights) can be deemed as a satisfactory representation of the 500 realisations.

Table 1. Statistics of Wasserstein distances between the normalised empirical CDFs calculated with the 500 tsunami realisations and the 13 realisations with weights.

Parameter	846 geographic points					3,108 geographic points				
	Min	Max	Mean	Median	StDev	Min	Max	Mean	Median	StDev
Flow Depth	0.002	0.122	0.032	0.029	0.013	0.002	0.156	0.031	0.029	0.012
Abs. Velocity	0.002	0.175	0.037	0.034	0.016	0.002	0.187	0.036	0.034	0.015
Flux	0.002	0.147	0.037	0.035	0.015	0.002	0.126	0.036	0.034	0.014
Loss Ratio	0.002	0.131	0.036	0.034	0.019	0.002	0.156	0.034	0.034	0.018
Mean DS	0.001	0.128	0.030	0.028	0.018	0.002	0.156	0.029	0.028	0.017
Prob. of DS0	0.000	0.167	0.039	0.035	0.021	0.000	0.180	0.037	0.034	0.018
Prob. of DS1	0.002	0.189	0.040	0.037	0.021	0.002	0.263	0.040	0.037	0.021
Prob. of DS2	0.002	0.184	0.035	0.030	0.021	0.002	0.190	0.035	0.031	0.019
Prob. of DS3	0.002	0.171	0.039	0.035	0.021	0.002	0.191	0.037	0.035	0.019
Prob. of DS4	0.002	0.170	0.041	0.039	0.021	0.002	0.196	0.040	0.039	0.020
Prob. of DS5	0.002	0.134	0.040	0.038	0.018	0.002	0.156	0.039	0.037	0.017

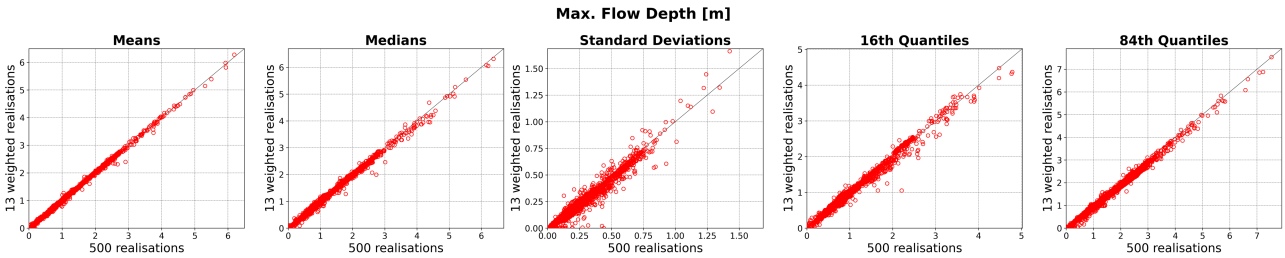

Figure 12. Means, medians, standard deviations, 16th and 84th quantiles of maximum flow depth for the 3,108 geographic points, calculated with 13 (vertical axis) vs 500 (horizontal axis) realisations.

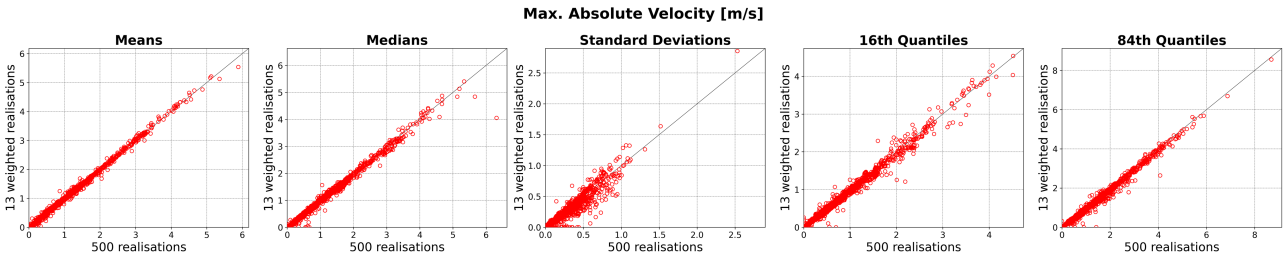

Figure 13. Means, medians, standard deviations, 16th and 84th quantiles of maximum absolute velocity for the 3,108 geographic points, calculated with 13 (vertical axis) vs 500 (horizontal axis) realisations.

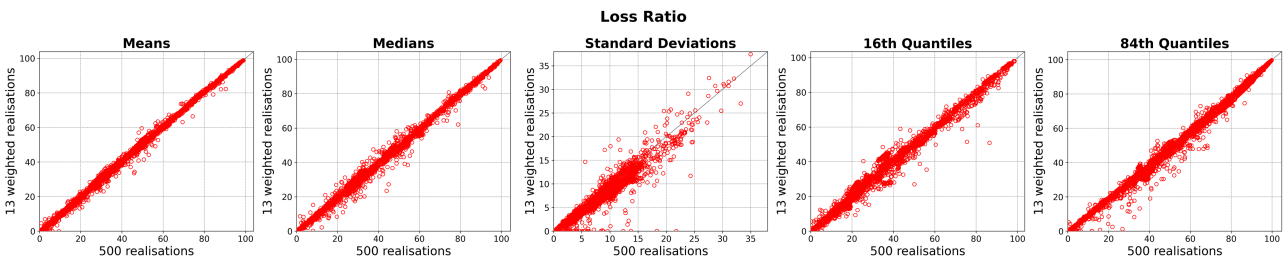

Figure 14. Means, medians, standard deviations, 16th and 84th quantiles of loss ratios for the 3,108 geographic points, calculated with 13 (vertical axis) vs 500 (horizontal axis) realisations.

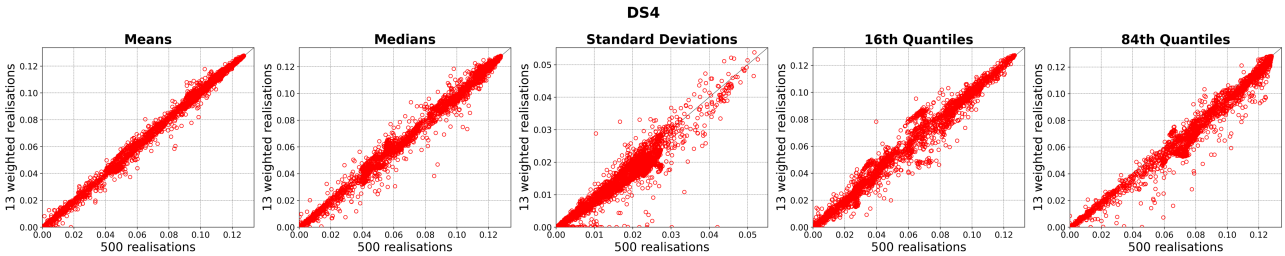

Figure 15. Means, medians, standard deviations, 16th and 84th quantiles of the probability of occurrence of DS4 for the 3,108 geographic points, calculated with 13 (vertical axis) vs 500 (horizontal axis) realisations.

4 Outlook

The results presented herein indicate that the selection of a subset of tsunami realisations by means of the Sammon's Mapping technique is adequate to reduce the computational demand associated with capturing the uncertainty stemming from the distribution of slip on the earthquake rupture in our case-study scenario. The selection of 13 out of 500 tsunami realisations leads to a technically-defensible range of tsunami intensity measures and consequences in terms of building damage and economic losses.

Whether such data reduction techniques are necessary for modelling extreme events ultimately depends on the kind of uncertainty, the computational demand associated with the modelling of the events, and the spatial domain (extent and resolution) of the calculations. Within the CASCO project, for example, earthquake ground motions are being modelled stochastically by means of ground motion prediction equations (GMPEs). Epistemic uncertainties are captured by means of the 15-branch logic tree of the European Seismic Hazard Model 2020 (ESHM20; Danciu et al., 2021) while aleatory uncertainties are considered by generating 1,000 random ground motion fields per branch, accounting for spatial correlation. Damage resulting from these 15 x 1,000 = 15,000 ground motion realisations does not take much time to run using OpenQuake (Pagani et al., 2014; Silva et al., 2014) on a server. However, physics-based simulations of tsunami generation and wave propagation do take much longer. For this reason, analogous steps to those presented herein will be followed for the landslide-triggered component of the tsunami, for which 54 physically possible submarine landslides have been defined along the eastern coast of Sicily and their occurrence and subsequent tsunamis modelled. In this case, a smaller number of segments might be defined to subdivide the Sammon's map, given the smaller number of alternative realisations (54 vs 500). A different grouping strategy altogether might even be considered if the 54 landslide-triggered tsunamis show some degree of clustering on the map. The realisations of the landslide-triggered tsunami selected in this way will be combined with the 13 realisations of the earthquake-triggered tsunami for the joint computation of the final tsunami propagation resulting from these two sources. The alternative route of considering all possible combinations of 500 earthquake triggers with 54 landslide triggers (500 x 54 = 27,000), potentially combined with some of the other uncertainties listed in the introduction would result in quite a prohibitive number of combinations. The decision of using data reduction techniques of full enumeration of logic tree branches ultimately depends on pondering these factors as a function of the particular goals of the calculations.

5 References

Aronica G.T., Brigandí G., Morey N. (2012). Flash floods and debris flow in the city area of Messina, north-east part of Sicily, Italy in October 2009: the case of the Giampilieri catchment. *Natural Hazards and Earth System Sciences*, 12(5): 1295-1309.

Belliazzi S., Lignola G. P., Di Ludovico M., Prota A. (2021). Preliminary tsunami analytical fragility functions proposal for Italian coastal residential masonry buildings, *Structures*, 31: 68-79.

Crowley H., Dabbeek J., Despotaki V., Rodrigues D., Martins L., Silva V., Romão X., Pereira N., Weatherill G., Danciu L. (2021). *European Seismic Risk Model (ESRM20), EFEHR Technical Report 002, V1.0.1*, 84 pp, DOI: 10.7414/EUC-EFEHR-TR002-ESRM20.

Danciu L., Nandan S., Reyes C., Basili R., Weatherill G., Beauval C., Rovida A., Vilanova S., Sesetyan K., Bard P.-Y., Cotton F., Wiemer S., Giardini D. (2021). *The 2020 update of the European Seismic Hazard Model – ESHM20: Model Overview, EFEHR Technical Report 001 v1.0.0*. https://doi.org/10.12686/a15

DISS Working Group (2018). *Database of Individual Seismogenic Sources (DISS), Version 3.2.1: A compilation of potential sources for earthquakes larger than M 5.5 in Italy and surrounding areas*, http://diss.rm.ingv.it/diss/, Istituto Nazionale di Geofisica e Vulcanologia (INGV) - DOI:10.6092/INGV.IT-DISS3.2.1.

Favalli M., Boschi E., Mazzarini F., Pareschi M.T. (2009). Seismic and landslide source of the 1908 Straits of Messina tsunami (Sicily, Italy), *Geophysical Research Letters*, 36: L16304.

Goda K., Mai P.M., Yasuda T., Mori N. (2014). Sensitivity of tsunami wave profiles and inundation simulations to earthquake slip and fault geometry for the 2011 Tohoku earthquake. *Earth, Planets and Space*, 66: 105.

Gomez-Zapata J.C., Brinckmann N., Harig S., Zafrir R., Pittore M., Cotton F., Babeyko A. (2021). Variable-resolution building exposure modelling for earthquake and tsunami scenario-based risk assessment: an application case in Lima, Peru, *Natural Hazards and Earth System Sciences*, 21(11): 3599-3628.

Goulet C.A., Bozorgnia Y., Kuehn N., Al Atik L., Youngs R.R., Graves R.W., Atkinson G.M. (2021). NGA-East ground-motion characterization model Part I: Summary of products and model development, *Earthquake Spectra*, 37(1_suppl): 1231-1282.

Graves R.W., Pitarka A. (2010). Broadband ground-motion simulation using a hybrid approach, *Bulletin of the Seismological Society of America*, 100(5A): 2095-2123.

Guidoboni E., Ferrari G., Tarabusi G., Sgattoni G., Comastri A., Mariotti D., Ciuccarelli C., Bianchi M.G., Valensise G. (2019). CFTI5Med, the new release of the catalogue of strong earthquakes in Italy and in the Mediterranean area, *Scientific Data*, 6: 80.

Harig S., Zamora N., Gubler A., Rakowsky, N. (2022). Systematic Comparison of Tsunami Simulations on the Chilean Coast Based on Different Numerical Approaches, *GeoHazards*, 3: 345–370.

Løvholt F., Pedersen G., Bazin S., Kühn D., Bredesen R.E., Harbitz C. (2012). Stochastic analysis of tsunami runup due to heterogeneous coseismic slip and dispersion, *Journal of Geophysical Research: Oceans*, 117: C03047.

Mai P.M., Beroza, G.C. (2002). A spatial random field model to characterize complexity in earthquake slip, *Journal of Geophysical Research: Solid Earth*, 107(B11): ESE 10-1-ESE 10-21.

Okada Y. (1985). Surface deformation due to shear and tensile faults in a half-space, *Bulletin of the Seismological Society of America*, 75(4): 1135-1154.

Okada Y. (1992). Internal deformation due to shear and tensile faults in a half-space, *Bulletin of the Seismological Society of America*, 82(2): 1018-1040.

Pagani M., Monelli D., Weatherill G., Danciu L., Crowley H., Silva V., Henshaw P., Butler L., Nastasi M., Panzeri L., Simionato M., Vigano D. (2014). OpenQuake engine: An open hazard (and risk) software for the global earthquake model, *Seismological Research Letters*, 85(3): 692-702.

Rakowsky N., Androsov A., Fuchs A., Harig S., Immerz A., Danilov S., Hiller W., Schröter J. (2013). Operational tsunami modelling with TsunAWI–recent developments and applications, *Natural Hazards and Earth System Sciences*, 13(6): 1629-1642.

Sammon J.W. Jr. (1969). A nonlinear mapping for data structure analysis, *IEEE Transactions on Computers*, 100(5): 401-409.

Silva V., Crowley H., Pagani M., Monelli D., Pinho R (2014). Development of the OpenQuake engine, the Global Earthquake Model's open-source software for seismic risk assessment, *Natural Hazards*, 72: 1409-1427.

Tarquini S., Isola I., Favalli M., Battistini A., Dotta G. (2023). *TINITALY, a digital elevation model of Italy with a 10 meters cell size (Version 1.1)*, Istituto Nazionale di Geofisica e Vulcanologia (INGV). https://doi.org/10.13127/tinitaly/1.1.

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*, 84(4): 974-1002.