A Heuristic Features Selection Approach for Scenario Analysis in a Regional Seismic Probabilistic Tsunami Hazard Assessment

Francesco Di Maio[a], Nicola Gallo[a], Matteo Taroni[c], Enrico Baglione[d], Jacopo Selva[d], Enrico Zio[a,b]

[a] Energy department, Politecnico di Milano, Milan, Italy

[b] MINES ParisTech, PSL Research University, CRC, Sophia Antipolis, France

[c] Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Roma 1, Rome, Italy

[d] Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Bologna, Bologna, Italy

ABSTRACT

Seismic Probabilistic Tsunami Hazard Analysis (SPTHA) is aimed at estimating the annual rate of exceedance of an earthquake-induced tsunami wave of a certain location with reference to a predefined height threshold. The analysis relies on computationally demanding numerical simulations of seismic-induced tsunami wave generation and propagation. A large number of scenarios needs to be simulated to account for uncertainties. However, the exceedance of tsunami wave threshold height is a rare event so that most of the simulated scenarios bring little statistical contribution to the estimation of the annual rate yet increasing the computational burden. To efficiently address this issue, we propose a wrapper-based heuristic approach to select the set of most relevant features of the seismic model, for deciding a priori the seismic scenarios to be simulated. The proposed approach is based on a Multi-Objective Differential Evolution Algorithm (MODEA) and is developed with reference to a case study whose objective is calculating the annual rate of threshold exceedance of the height of tsunami waves caused by subduction earthquakes that might be generated on a section of the Hellenic Arc, and propagated to a set of target sites: Siracusa, on the eastern coast of Sicily,

Crotone, on the southern coast of Calabria, and Santa Maria di Leuca, on the southern coast of Puglia. The results show that, in all cases, the proposed approach allows a reduction of 95% of the number of scenarios with half of the features to be considered, and with no appreciable loss of accuracy.

Keywords: Seismic Probabilistic Tsunami Hazard Analysis (SPTHA); Scenario selection; Feature selection; Wrapper approach; Multi-Objective Differential Evolution Algorithm (MODEA).

Acronyms

DE	Differential Evolution
GA	Genetic Algorithm
MODEA	Multi Objective Differential Evolution Algorithm
NEAM	North-eastern Atlantic, the Mediterranean, and connected seas
SPTHA	Seismic Probabilistic Tsunami Hazard

Symbols

$\bar{a}$	Target site coordinates
$\psi_{\bar{a}}$	Tsunami intensity in $\bar{a}$
ΔT	Exposure time
$\bar{x}$	Seismic scenario parameters vector
$\sigma_{\bar{x}}$	Seismic scenario of parameters $\bar{x}$
Σ	Space of possible seismic scenarios
$\lambda(\sigma_{\bar{x}})$	Annual frequency of the seismic scenario $\sigma_{\bar{x}}$
Pr	Probability
P_e	Probability of exceedance
$\tilde{\psi}$	Tsunami intensity threshold
$\Lambda(\psi_{\bar{a}} \geq \tilde{\psi})$	Annual frequency of occurrence of a tsunami of intensity $\psi_{\bar{a}} \geq \tilde{\psi}$ at location $\bar{a}$
q	Generic simulated seismic scenario
Q	Total number of seismic scenarios to be simulated
Q^*	Optimised (i.e., minimum) number of seismic scenarios to be simulated
H_i	Generic seismic zone

H_n	Total number of seismic zones
Θ	Set of alternative ET models
ϑ	Generic alternative ET model
M	Number of alternative models for the calculation of $\Lambda\left(\psi_{\bar{a}} \geq \tilde{\psi}\right)$
x_1	Magnitude of an earthquake
x_2	Depth of the fault
x_3	Strike of the focal mechanism
x_4	Dip of the focal mechanism
x_5	Rake of the focal mechanism
x_6	Area of the fault
x_7	Length of the fault
x_8	Slip of the fault
F	Set of the objective functions of the MODEA
f	Objective function of the MODEA
$\bar{U}$	Decision variable vector
$\bar{U}^*$	Solution vector
$\bar{\bar{A}}$	Features matrix
$\bar{\bar{A}}^*$	Optimal features matrix
SE	Squared error
NP	Population size of the DE
F	Scaling factor of the DE
CR	Crossover probability of the DE
$MAXGEN$	Maximum number of DE generations
D	Number of objectives of the MODEA

1. INTRODUCTION

Tsunami hazard is classically assessed by simulation of either one "worst credible" or few representative scenarios (Lorito et al. 2015; Rikitake and Aida 1988; Geist et al., 2006; Heidarzadeh et al., 2011; Gopinathan et al., 2021). This can be an effective approach when (i) the effects of frequent, small magnitude earthquakes are expected to be negligible compared to those less frequent large magnitude earthquakes, and (ii) the analysis is conducted in a relatively simple geophysical context where tsunami hazard is dominated by the large magnitude earthquakes occurring in subduction zones, whose geometries are reasonably well constrained (Lorito et al. 2015). On the other hand, when tsunamis are generated in complex and fragmented tectonic environments (e.g., the Caribbean Sea and the Mediterranean Sea) or when relatively short return periods need to be considered, the tsunami hazard might be severely biased (Lorito et al. 2015) and the consequences underestimated, especially if the extent of damage is not well modelled and caught (Suppasri et al., 2021). To explicitly account for the whole spectrum of seismic triggering events and their related uncertainty, a probabilistic analysis of a large set of potential tsunamis can be performed (Seismic Probabilistic Tsunami Hazard Analysis, SPTHA) (Rikitake and Aida 1988; Grezio et al. 2017; Fakhruddin et al. 2021; Park et al. 2021). Specifically, SPTHA aims to estimate the probability that the height ψ of an earthquake-induced tsunami wave exceeds a threshold $\tilde{\psi}$, within in an exposure time ΔT, at a location of coordinates $\bar{a}$ (Grezio et al. 2017). Each tsunami is assumed to be generated by a seismic scenario $\sigma_{\bar{x}}$ belonging to the space of possible seismic scenarios Σ ($\sigma_{\bar{x}} \in \Sigma$), characterized by parameters $\bar{x}$ and occurring with annual frequency $\lambda(\sigma_{\bar{x}})$. Considering a Poisson process for the wave exceedance event occurrence in time, the probability of exceedance P_e can be written as:

$$P_e = Pr\left(\psi_{\bar{a}} \geq \tilde{\psi}; \Delta T\right) = 1 - exp\left(-\Lambda\left(\psi_{\bar{a}} \geq \tilde{\psi}\right) \Delta T\right) \tag{1}$$

where $\Lambda\left(\psi_{\bar{a}} \geq \tilde{\psi}\right)$ is the annual rate of occurrence of a tsunami of intensity $\psi_{\bar{a}} \geq \tilde{\psi}$ at location $\bar{a}$. This rate is calculated by integrating, over the space Σ, the annual frequency $\lambda(\sigma_{\bar{x}})$ of occurrence of the seismic scenario $\sigma_{\bar{x}}$ times the probability $Pr\left(\psi_{\bar{a}} \geq \tilde{\psi} | \sigma_{\bar{x}}\right)$ that the tsunami wave generated by the scenario exceeds $\tilde{\psi}$:

$$\Lambda(\psi_{\bar{a}} \geq \tilde{\psi}) = \int_{\Sigma} \lambda(\sigma_{\bar{x}}) Pr(\psi_{\bar{a}} \geq \tilde{\psi} | \sigma_{\bar{x}}) d\sigma_{\bar{x}} \tag{2}$$

Considering, without loss of generality and for the sake of simplicity, a set of Q discretized seismic scenarios $\sigma_{\bar{x}_q}$ ($q = 1, \dots, Q$) with $\lambda\left(\sigma_{\bar{x}_q}\right)$ and $Pr\left(\psi_{\bar{a}} \geq \tilde{\psi} | \sigma_{\bar{x}_q}\right)$, Eq. (2) can be approximated as:

$$\Lambda(\psi_{\bar{a}} \geq \tilde{\psi}) \approx \sum_{q=1}^{Q} \lambda\left(\sigma_{\bar{x}_q}\right) Pr\left(\psi_{\bar{a}} \geq \tilde{\psi} | \sigma_{\bar{x}_q}\right) \tag{3}$$

To account for epistemic uncertainty, M alternative formulations of $\lambda\left(\sigma_{\bar{x}_q}\right)$ and $Pr\left(\psi_{\bar{a}} \geq \tilde{\psi} | \sigma_{\bar{x}_q}\right)$ can be considered, producing M alternative quantifications of both factors in Eq. (3). The mean hazard rate can, then, be evaluated as:

$$\Lambda(\psi_{\bar{a}} \geq \tilde{\psi}) \approx \frac{1}{M} \sum_{m=1}^{M} \sum_{q=1}^{Q} \lambda\left(\sigma_{\bar{x}_q}\right)_m Pr\left(\psi_{\bar{a}} \geq \tilde{\psi} | \sigma_{\bar{x}_q}\right)_m \tag{4}$$

where $\lambda\left(\sigma_{\bar{x}_q}\right)_m$ is the generic entry of the matrix $\overline{\overline{\lambda(\sigma_{\bar{x}})}}$:

$$\overline{\overline{\lambda(\sigma_{\bar{x}})}} = \begin{pmatrix} \lambda(\sigma_{\bar{x}_1})_1 & \cdots & \lambda(\sigma_{\bar{x}_1})_m & \cdots & \lambda(\sigma_{\bar{x}_1})_M \\ \vdots & & \vdots & & \vdots \\ \lambda\left(\sigma_{\bar{x}_q}\right)_1 & \cdots & \lambda\left(\sigma_{\bar{x}_q}\right)_m & \cdots & \lambda\left(\sigma_{\bar{x}_q}\right)_M \\ \vdots & & \vdots & & \vdots \\ \lambda\left(\sigma_{\bar{x}_Q}\right)_1 & \cdots & \lambda\left(\sigma_{\bar{x}_Q}\right)_m & \cdots & \lambda\left(\sigma_{\bar{x}_Q}\right)_M \end{pmatrix} \tag{5}$$

and $Pr\left(\psi_{\bar{a}} \geq \tilde{\psi} | \sigma_{\bar{x}_q}\right)_m$ is the generic entry of the matrix $\overline{\overline{Pr(\psi_{\bar{a}} \geq \tilde{\psi} | \sigma_{\bar{x}})}}$:

$$\overline{\overline{Pr(\psi_{\bar{a}} \geq \tilde{\psi} | \sigma_{\bar{x}})}} = \begin{pmatrix} Pr(\psi_{\bar{a}} \geq \tilde{\psi} | \sigma_{\bar{x}_1})_1 & \cdots & Pr(\psi_{\bar{a}} \geq \tilde{\psi} | \sigma_{\bar{x}_1})_m & \cdots & Pr(\psi_{\bar{a}} \geq \tilde{\psi} | \sigma_{\bar{x}_1})_M \\ \vdots & & \vdots & & \vdots \\ Pr\left(\psi_{\bar{a}} \geq \tilde{\psi} | \sigma_{\bar{x}_q}\right)_1 & \cdots & Pr\left(\psi_{\bar{a}} \geq \tilde{\psi} | \sigma_{\bar{x}_q}\right)_m & \cdots & Pr\left(\psi_{\bar{a}} \geq \tilde{\psi} | \sigma_{\bar{x}_q}\right)_M \\ \vdots & & \vdots & & \vdots \\ Pr\left(\psi_{\bar{a}} \geq \tilde{\psi} | \sigma_{\bar{x}_Q}\right)_1 & \cdots & Pr\left(\psi_{\bar{a}} \geq \tilde{\psi} | \sigma_{\bar{x}_Q}\right)_m & \cdots & Pr\left(\psi_{\bar{a}} \geq \tilde{\psi} | \sigma_{\bar{x}_Q}\right)_M \end{pmatrix} \tag{6}$$

Note that the calculation of the entries of $\overline{\overline{Pr(\psi_{\bar{a}} \geq \tilde{\psi} | \sigma_{\bar{x}})}}$ may result computationally burdensome, for example when using highly non-linear tsunami simulation models.

In the particular case of a local SPTHA for the estimation of inundation hazard curves for a small target site, e.g., a refinery, high-resolution inundation simulations are needed. This requires either large High Performance Computing (HPC) resources (Gibbons et al. 2020; González et al. 2009), or statistical emulation (Samanidou et al., 2019),or a reduction of the number of simulations by, for example a two-stage filtering procedure (González et al. 2009; Lorito et al. 2015; Volpe et al. 2019), or training a metamodel, for example an Adaptive Kriging that mimics the behaviour of the computationally demanding tsunami inundation simulator, e.g., (Bacchi et al. 2020; Di Maio et al. 2021).

On the other hand, when the interest is in a regional SPTHA for the estimation of inundation hazard curves for large areas, such as countries or continents, the computation is usually performed by using simplified relationships between the water elevation at the shoreline and the maximum inundation height (Gailler et al. 2018; Glimsdal et al. 2019).

The present work considers regional SPTHA and illustrates a novel approach for identifying the relevant features of the seismic scenarios and the selection of a limited number of them needed for performing the annual rate estimation with sufficient accuracy. Specifically, a Multi-Objective Differential Evolution Algorithm (MODEA) is used to select the features (Storn and Price 1997).

As example of practical use, we consider the estimation of the annual rate of exceedance of a threshold $\tilde{\psi} = 1m$ of tsunami wave height, resulting from subduction earthquakes in a section of the Hellenic Arc. To show the applicability of the approach to different target sites, we consider the propagation of the waves towards Siracusa, on the eastern coast of Sicily, Crotone, on the southern coast of Calabria, and Santa Maria di Leuca, on the southern coast of Puglia. We consider the crustal seismicity generated in the Kefalonia-Lefkada region, thus developing outside the subduction interface of the Hellenic Arc (Basili et al. 2021; Selva et al. 2016); this is one of the regions considered in the tsunami hazard model recently released for the North-eastern Atlantic, the Mediterranean, and connected seas (i.e., NEAM region) (Basili et al. 2018, 2021). In NEAMTHM18, a predefined source discretization method is applied for magnitude, hypocentres, strike, dip and rake angles. Based on this, a combinatorial setting of the parameters values makes the target source area exposed to a total of $Q_{tot} = 23272$ seismic scenarios where $M = 1000$ alternative models can be used for the calculation of $\Lambda(\psi_{\bar{a}} \geq 1m)$ (Basili et al., 2021).

For all the target sites, a comparison is provided between the value of the mean annual rate of exceedance estimated considering only the selected scenarios SPTHA and the full set of scenarios SPTHA. The outcome of the comparison shows that in all cases the proposed approach allows a significant reduction of the number of scenarios needed to be processed, without affecting the accuracy of the estimate.

The paper is organised as follows. Section 2 presents the case study. Section 3 explains the approach developed. Section 4 shows the results of the application of the proposed approach to the case study. Conclusions are drawn in Section 5.

2. CASE STUDY

We consider the regional SPTHA for three target sites $\bar{a}$ located on the eastern coast of Sicily (Siracusa), the southern coast of Calabria (Crotone), and Puglia (Santa Maria di Leuca), as reported in Figure 1. All targets are exposed to tsunamis triggered by crustal earthquakes occurring outside the subduction interface of the Hellenic Arc in the Kefalonia-Lefkada region (Basili et al. 2021). Earthquakes are assumed to be generated at specific epicentral locations H_i (i=1,…, 42, blue points in Figure 1) with different magnitudes, depths, and faulting mechanisms. Without loss of generality, the following assumptions are made:

i. The threshold is of $\tilde{\psi} = 1m$ at 50m from the coastline.

ii. One epicentral location (black diamond in location 14,in Figure 1) is considered, since the largest number $Q = 721$ of seismic scenarios $\sigma_{\bar{x}}$ is available among all other locations, making, $\Lambda(\psi_{\bar{a}} \geq 1m|H_{14})$ equal to:

$$\Lambda(\psi_{\bar{a}} \geq 1m|H_{14}) \approx \frac{1}{M}\sum_{m=1}^{M}\sum_{q=1}^{Q}\lambda\left(\sigma_{\bar{x}_q}|H_{14}\right)_m Pr\left(\psi_{\bar{a}} \geq 1m|\sigma_{\bar{x}_q}, H_{14}\right)_m \quad (7)$$

(Herein after, for the sake of readability, H_{14} will be omitted).

The scenarios associated with the other neighbouring areas have not been included in the study due to their similarity to those of point 14 and would not have brought any further significant information.

iii. Each $\sigma_{\bar{x}}$ is characterised by the set of parameters $\bar{x} = (x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8)$ (Basili et al. 2021; Selva et al. 2016), whose support and values are listed in Table 1. These parameters (see Figure 2 for a schematic representation) are:

1. x_1 Magnitude
2. x_2 Depth (top of the fault)
3. x_3 Strike (of the focal mechanism)
4. x_4 Dip (of the focal mechanism)
5. x_5 Rake (of the focal mechanism)
6. x_6 Area (of the fault), i.e., the product of its width by its length
7. x_7 Length (of the fault)
8. x_8 Slip (of the fault)

It is worth mentioning that all parameters are discretized to build a finite list of scenarios, in line with the recommendations provided by NEAMTHM18 (Basili et. 2021): bins are defined linearly for all the parameters except for the magnitude for which the larger the magnitude events the finer the discretization needed, motivated by a tsunamigenic reasons (Lorito et al. 2015, Selva et al. 2016). Also, since the slip parameter value is dependent on the value of the magnitude (and, therefore, and the discretization adopted), two different scaling laws (see Basili et al., 2021) are adopted for strike and dip-slip events.

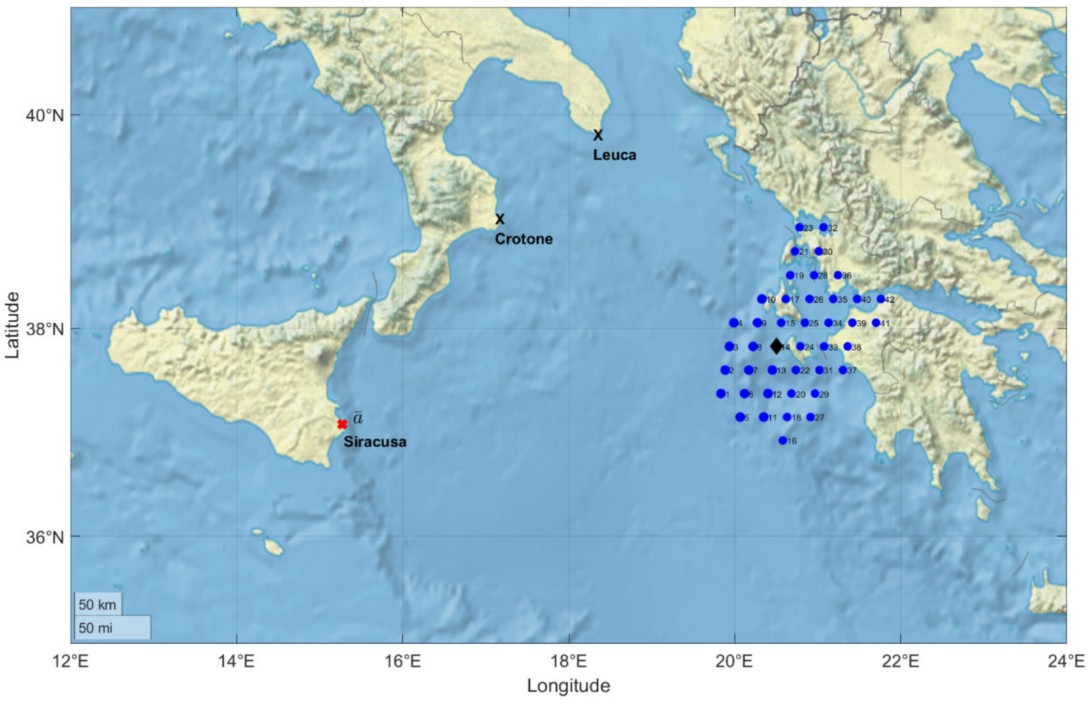

Figure 1 – Seismic zones H_i (i=1,…,42) of the Hellenic Arc (blue dots), the source considered (black diamond), the target sites $\bar{a}$ Siracusa (red cross), and Crotone and Santa Maria di Leuca (black crosses).

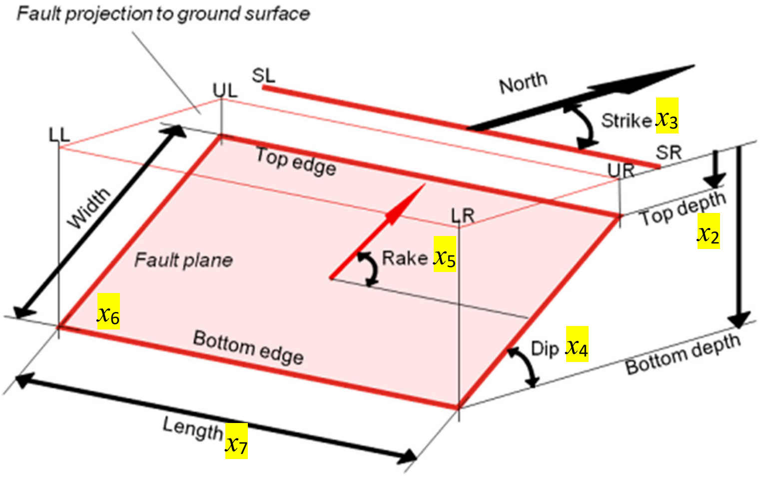

Figure 2 – Schematic representation of an earthquake and its parameters (Basili et al. 2008).

Table 1 – Parameters of the seismic scenarios (Each parameter has its specific discretization, resulting in a different number of reference values and bins).

Parameter	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
Support	[6.5000-8.0933]	[1.00-17.86]	[22.5-337.5]	[10-90]	[0-270]	[318.95-12648.92]	[22.68-665.73]	[0.67-4.21]
Reference Values	6.5000	1.00	22.5	10	0	318.95	22.68	0.67
	6.8012	5.97	67.5	30	90	558.32	34.39	0.95
	7.0737	7.56	112.5	50	180	638.11	37.88	1.09
	7.3203	9.43	157.5	70	270	1194.98	50.10	1.29
	7.5435	10.94	202.5	90		1205.54	63.64	1.30
	7.7453	1158	247.5			2108.29	70.44	1.71
	7.9280	14.00	292.5			2133.31	95.87	1.73
	8.0933	14.12	337.5			3566.59	112.28	2.21
		16.65				3524.55	126.69	2.24
		17.86				5608.92	163.06	2.79
						5676.15	187.72	2.82
						8541.96	204.89	3.44
						8644.77	298.74	3.48
						12497.92	454.99	4.16
						12648.92	665.73	4.21

3. METHODOLOGY

To alleviate the computational burden of the SPTHA, the procedure sketched in Figure 3 is developed. Firstly, an optimisation problem is solved to identify the optimal set of seismic scenarios that contribute most to $\Lambda(\psi_{\bar{a}} \geq 1m)$ of Eq. (7). Then, their features values are identified. The optimisation is performed by a wrapper-based heuristic approach: based on a Multi-Objective Differential Evolution Algorithm (MODEA) wherein the Differential Evolution (DE) engine (Reynoso-Meza 2021; Storn and Price 1997) iteratively searches for candidate sets of scenarios, among the original dataset of $Q = 721$ scenarios, whose performance is evaluated with respect to a given cost function (the interested reader may refer to Appendix A). Once the optimal set of scenarios is identified, their common features are retrieved by statistical analysis.

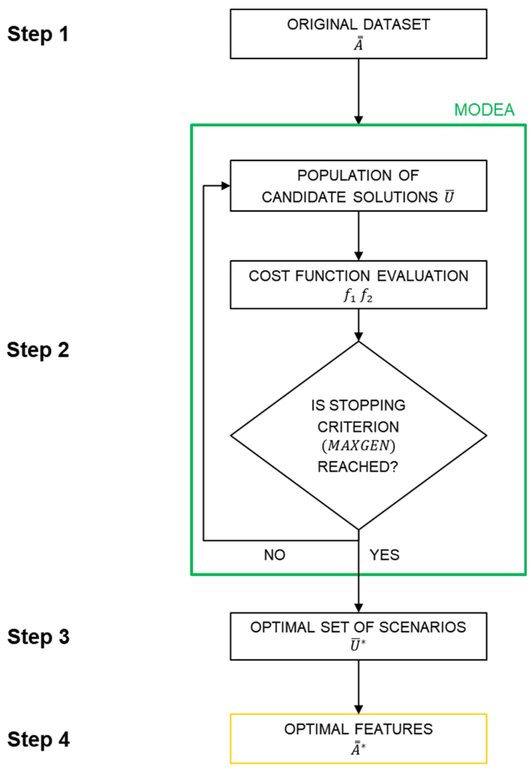

Figure 3 – Wrapper approach for optimal set of scenarios selection based on MODEA

The procedure is explained in detail here below.

Step 1: Consider the original dataset

The original dataset $\bar{\bar{A}} = [Q \times 9]$ is:

$$\bar{\bar{\mathrm{A}}} = \begin{pmatrix} x_{1,1} & \cdots & x_{8,1} & \Lambda\left(\psi_{\bar{a}} \geq 1m|\sigma_{\bar{x}_1}\right) \\ \vdots & \vdots & \vdots & \vdots \\ x_{1,q} & \cdots & x_{8,q} & \Lambda\left(\psi_{\bar{a}} \geq 1m|\sigma_{\bar{x}_q}\right) \\ \vdots & \vdots & \vdots & \vdots \\ x_{1,Q} & \cdots & x_{8,Q} & \Lambda\left(\psi_{\bar{a}} \geq 1m|\sigma_{\bar{x}_Q}\right) \end{pmatrix} \quad (8)$$

where $x_{1,q}$ is the value of the parameter x_1 in the q-th scenario, $x_{2,q}$ is the value of the parameter x_2 in the q-th scenario, etc., and $\Lambda\left(\psi_{\bar{a}} \geq 1m|\sigma_{\bar{x}_q}\right) = \frac{1}{M}\sum_{m=1}^{M} \lambda\left(\sigma_{\bar{x}_q}\right)_m Pr\left(\psi_{\bar{a}} \geq 1m|\sigma_{\bar{x}_q}\right)_m$ is the annual rate of exceedance of the q-th scenario.

Step 2: Apply MODEA to identify the most relevant scenarios

The MODEA searches the global minimum of a set of objective (cost) functions $F = \{f(\cdot)\}$, of one (or more) decision vector(s) $\bar{U}$ (typically a string of binary digits) (Zio, Baraldi, and Gola 2007; Zio, Baraldi, and Pedroni 2006). In the case of interest for this work, $\bar{U}$ indicates whether the q-th seismic scenario is considered in the candidate solution (q-th bit equal to 1) or not (q-th bit equal to 0).

The MODEA search is performed by initially randomly sampling the bits of the NP vectors that compose the initial population strings (Zio et al. 2006). Then, iteratively, the population is enriched by the solution $\bar{U}$ that best fits the objective functions, through a selection process driven by a set of parameters, i.e., the scaling factor F and the crossover probability CR (Storn and Price 1997). For a thorough description of the process based on DE and its controlling parameters, the interested reader may refer to the Appendix A or to (Storn and Price 1997).

The two objective functions considered are:

1. Minimisation of Q (i.e., the number of scenarios $\sigma_{\bar{x}_q}$ considered in the solution):

$$f_1 = \sum_{q=1}^{Q} U_q \quad (9)$$

2. Minimisation of the squared error SE between the annual rate of exceedance $\Lambda(\psi_{\bar{a}} \geq 1m)$ and the annual rate of exceedance calculated considering exclusively the $Q^* = min\left(\sum_{q=1}^{Q} U_q\right)$ selected scenarios $\Lambda^*(\psi_{\bar{a}} \geq 1m)$:

$$f_2 = \left(\Lambda(\psi_{\bar{a}} \geq 1m) - \Lambda^*(\psi_{\bar{a}} \geq 1m)\right)^2 \quad (10)$$

where $\Lambda^*(\psi_{\bar{a}} \geq 1m)$ is calculated as:

$$\Lambda^*(\psi_{\bar{a}} \geq 1m) = \sum_{q=1}^{Q} \lambda\left(\sigma_{\bar{x}_q}\right) Pr\left(\psi_{\bar{a}} \geq 1m | \sigma_{\bar{x}_q}\right) U_q \quad (11)$$

The search procedure ends when the stopping criterion (e.g., the maximum number of generations $MAXGEN$) is reached.

Step 3: Optimal set of scenarios

The optimal solution vector $\bar{U}^*$ (i.e., the optimal set of scenarios) that optimizes the multi-objective function of Eqs. (9) and (10) is selected from the Pareto optimal front (Zio et al. 2006), as the solution with the minimum number Q^* of entries equal to 1 (i.e., the scenarios considered in the candidate solution).

Step 4: optimal features identification

To identify the most relevant features to be considered for the SPTHA, we first calculate the optimal features matrix $\bar{\bar{A}}^* = [Q^* \times 9]$, as the Hadamard product of the original dataset $\bar{\bar{A}}$ with $\bar{U}^*$ (with $(Q - Q^*)$ null vector rows):

$$\bar{\bar{A}}^* = \bar{\bar{A}} \circ \bar{U}^* \quad (12)$$

$$\bar{\bar{A}}^* = \begin{pmatrix} x_{1,1} & \cdots & x_{8,1} & \Lambda\left(\psi_{\bar{a}} \geq 1m | \sigma_{\bar{x}_1}\right) \\ \vdots & \vdots & \vdots & \vdots \\ x_{1,q^*} & \cdots & x_{8,q^*} & \Lambda\left(\psi_{\bar{a}} \geq 1m | \sigma_{\bar{x}_{q^*}}\right) \\ \vdots & \vdots & \vdots & \vdots \\ x_{1,Q^*} & \cdots & x_{8,Q^*} & \Lambda\left(\psi_{\bar{a}} \geq 1m | \sigma_{\bar{x}_{Q^*}}\right) \end{pmatrix} \tag{13}$$

Then, the matrix $\bar{\bar{A}}^*$ is columnwise compared with the original dataset $\bar{\bar{A}}$ to assess their commonality (i.e., the optimal features subset).

4. RESULTS

The approach described in Section 3 has been applied to the case study presented in Section 2. The search for optimal scenarios among the $Q = 721$ of the original dataset is performed by a MODEA (DE/rand/1/bin strategy, see Appendix A for further details), with objective functions f_1 and f_2 (respectively Eq. (9) and Eq. (10)), where f_2 is calculated referring to the benchmark value of the annual rate of exceedance $\Lambda(\psi_{\bar{a}} \geq 1m) = 3.3193 \cdot 10^{-12} yr^{-1}$ calculated from the full set of scenarios. In practice, each candidate solution $\bar{U}$ is a binary string of $Q = 721$ bits. The population size NP, the scaling factor F, the crossover probability CR and the generation bound $MAXGEN$, have been expertly set equal to 20, 0.5, 0.9 and 10000, respectively: specifically, NP has been set equal to $10 \cdot (\#of\ objectives = 2) = 20$ in line with (Storn and Price 1997); CR has been set equal to 0.9 for a fast convergence (Storn and Price 1997); F has been set equal to 0.5 in line with (Ahmed 2005; Storn and Price 1997); the stopping criterion $MAXGEN = 10000$ has been set following a trial-and-error procedure (Reynoso-Meza 2021).

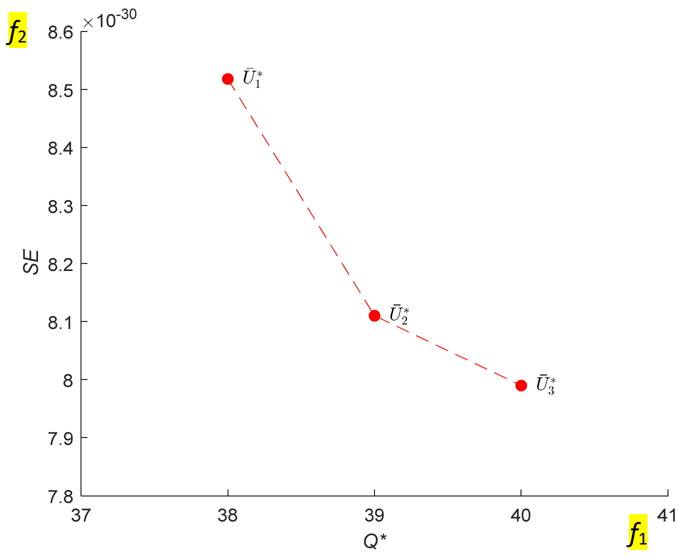

Figure 4 – Pareto optimal front after MAXGEN iterations

When the stopping criterion is reached, the Pareto front shown in Figure 4 is obtained:

1. $\overline{U}_1^*$ yields $Q^* = 38$ scenarios with a $SE = 8.5^{-30} yr^{-2}$ and a percentage error of 0.085%
2. $\overline{U}_2^*$ yields $Q^* = 39$ scenarios with a $SE = 8.1^{-30} yr^{-2}$ and a percentage error of 0.066%
3. $\overline{U}_3^*$ yields $Q^* = 40$ scenarios with a $SE = 8.0^{-30} yr^{-2}$ and a percentage error of 0.063%

In this work, the solution $\overline{U}_1^*$ is preferred because it yields the minimum number of $Q^* = 38$ scenarios (i.e., a 95% reduction with respect to Q) with a reasonably small $SE = 8.5^{-30} years^{-2}$ (i.e., a percentage error of 0.085%) in the estimation of $\Lambda(\psi_{\bar{a}} \geq 1m|H_{14})$. In Table 2, all the features and Q^* scenarios selected by the MODEA are listed (for the Siracusa target), without discarding low-frequency scenarios (Di Maio et al. 2021). All Q^* selected scenarios contribute to $\Lambda(\psi_{\bar{a}} \geq 1m|H_{14})$ with a relatively large probability of threshold exceedance $\Lambda\left(\psi_{\bar{a}} \geq 1m|H_{14}, \sigma_{\bar{x}_q}\right)$. On the contrary, most of the $Q = 721$ seismic scenarios in the original dataset have a $\Lambda\left(\psi_{\bar{a}} \geq 1m, \sigma_{\bar{x}_q}\right) < 10^{-20}$, i.e., bring a negligible contribution to the estimation of $\Lambda(\psi_{\bar{a}} \geq 1m|H_{14})$ but increase the computational burden. Furthermore, regarding the features selected to characterise the $Q^* = 38$

scenarios, these are reduced with respect to those that characterise the Q scenarios as shown in Table 3 and Figure 4-Figure 12.

For geophysical interpretation of the results obtained, we consider the independent features, i.e., the Magnitude and the angles of the faulting mechanism (Strike, Dip, and Rake).

For the remaining features, Depth, Area, Length and Slip that are dependent on the independent features, a discussion is provided in the Appendix B.

Table 2 – Features and $\Lambda\left(\psi_{\bar{a}} \geq 1m|H_{14}, \sigma_{\bar{x}_q}\right)$ of the $Q^* = 38$ selected scenarios (Siracusa target site)

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	$\Lambda\left(\psi_{\bar{a}} \geq 1m|H_{14}, \sigma_{\bar{x}_q}\right)$
6.5000	1.00	337.5	50	90	318.95	22.68	0.67	3.98E-17
6.8012	1.00	157.5	50	90	638.11	34.39	0.95	4.17E-16
6.8012	1.00	157.5	70	90	638.11	34.39	0.95	2.79E-16
6.8012	1.00	157.5	90	270	638.11	34.39	0.95	3.28E-16
6.8012	1.00	157.5	90	90	638.11	34.39	0.95	1.35E-16
6.8012	1.00	337.5	70	90	638.11	34.39	0.95	4.16E-17
6.8012	1.00	337.5	50	270	638.11	34.39	0.95	2.12E-17
6.8012	1.00	337.5	50	90	638.11	34.39	0.95	4.17E-16
6.8012	1.00	337.5	30	90	638.11	34.39	0.95	4.04E-13
6.8012	7.56	337.5	50	90	638.11	34.39	0.95	2.95E-16
6.8012	7.56	337.5	30	90	638.11	34.39	0.95	4.38E-13
6.8012	14.12	337.5	30	90	638.11	34.39	0.95	8.72E-15
7.0737	1.00	22.5	50	90	1194.98	50.10	1.30	2.41E-17
7.0737	1.00	157.5	50	270	1194.98	50.10	1.30	1.33E-16
7.0737	1.00	157.5	50	90	1194.98	50.10	1.30	1.50E-15
7.0737	1.00	157.5	70	90	1194.98	50.10	1.30	1.59E-15
7.0737	1.00	157.5	90	270	1194.98	50.10	1.30	1.11E-15
7.0737	1.00	337.5	70	90	1194.98	50.10	1.30	1.64E-16
7.0737	1.00	337.5	50	270	1194.98	50.10	1.30	4.02E-16
7.0737	1.00	337.5	50	90	1194.98	50.10	1.30	2.05E-15

7.0737	1.00	337.5	30	270	1194.98	50.10	1.30	3.26E-17
7.0737	1.00	337.5	30	90	1194.98	50.10	1.30	9.94E-13
7.0737	9.43	157.5	70	90	1194.98	50.10	1.30	2.32E-17
7.0737	9.43	337.5	30	90	1194.98	50.10	1.30	7.13E-13
7.0737	9.43	337.5	10	90	1194.98	50.10	1.30	4.25E-17
7.3203	1.00	157.5	50	90	2108.29	70.44	1.73	3.30E-16
7.3203	1.00	157.5	70	90	2108.29	70.44	1.73	1.05E-16
7.3203	1.00	157.5	90	270	2108.29	70.44	1.73	1.94E-16
7.3203	1.00	337.5	30	90	2108.29	70.44	1.73	3.45E-13
7.3203	11.58	157.5	50	90	2108.29	70.44	1.73	5.44E-17
7.3203	11.58	337.5	50	90	2108.29	70.44	1.73	7.61E-17
7.3203	11.58	337.5	30	90	2108.29	70.44	1.73	1.48E-13
7.5435	1.00	157.5	50	90	3524.55	95.87	2.24	1.96E-16
7.5435	1.00	157.5	70	90	3524.55	95.87	2.24	7.34E-17
7.5435	1.00	337.5	50	90	3524.55	95.87	2.24	1.78E-16
7.5435	1.00	337.5	30	90	3524.55	95.87	2.24	2.51E-13
7.7453	1.00	337.5	30	90	5608.92	126.69	2.82	3.83E-15

Table 3 – Support values of the selected scenarios (Siracusa target site)

Parameter	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
Values	6.5000	1.00	22.5	10	90	318.95	22.68	0.67
	6.8012	7.56	157.5	30	270	638.11	34.39	0.95
	7.0737	9.43	337.5	50		1194.98	50.10	1.30
	7.3203	1158		70		2108.29	70.44	1.73
	7.5435	14.12		90		2133.31	95.87	2.24
	7.7453					3524.55	126.69	2.82
						5608.92		

MAGNITUDE

The DE search engine for the Siracusa site has not selected (red slices in Figure 5a)) those scenarios characterized by large magnitudes ($x_1 = 7.7453, 8.0933$). Indeed, for such scenarios, the annual rates are negligible and do not bring any significant contribution to the hazard curve estimation (see Eq. (7)), even if a relatively large threshold of $\tilde{\psi} = 1m$ at 50m is assumed. Similarly, also for the target sites of Crotone and Santa Maria di Leuca (Figure 5b) and 5c), respectively), large magnitudes are neglected. Regarding Santa Maria di Leuca, also the scenarios with magnitude $x_1 = 6.5$ are neglected, probably due to its unfavorable position with respect to the source.

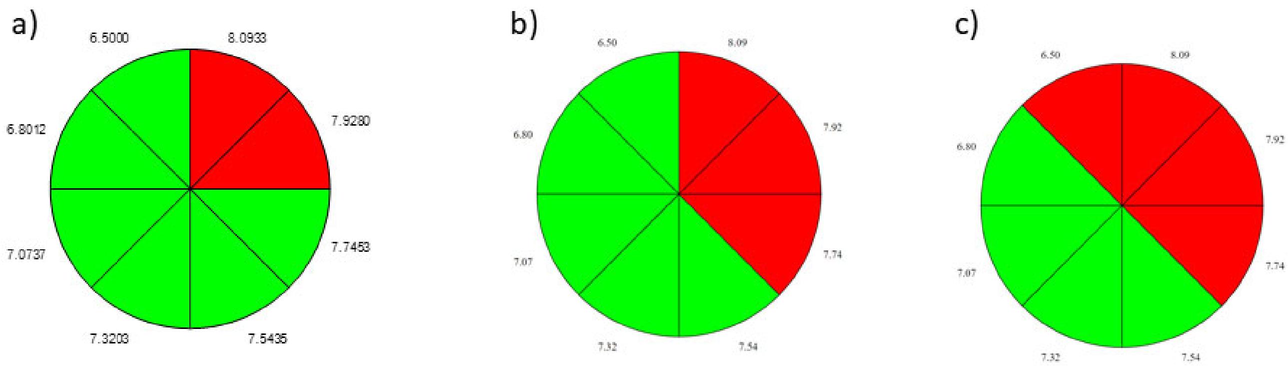

Figure 5 – Values of the magnitude (green) and non-selected ones (red) in the selected scenarios; a) Siracusa, b) Crotone, and c) Santa Maria di Leuca.

STRIKE

Regardless of the regional tectonic setting, some strikes are privileged over others due to their greater tsunamigenic impact. That is, the algorithm chooses scenarios based on the combination of their probability of occurrence and the effect in terms of tsunami. In fact, some strikes, those perpendicular to the source-target propagation direction, are associated with a greater probability of generating a considerable impact tsunami.

For Siracusa, the DE search engine has identified as relevant (green in Figure 6a)) those scenarios characterised by strike angle values of $22.5°$, $157.5°$, $337.5°$. These strikes are aligned with the

Hellenic subduction zone. Hence, they are both highly probable and maximize the tsunami in Siracusa, as strike directions are approximately perpendicular to the source-to-site tsunami propagation path (see Figure 1). Moving toward East, for Crotone and, even more, for Leuca (see Figure 6b) and 6c), respectively), other strikes are added to the most probable and perpendicular to the source-target direction ones: in particular, being Santa Maria di Leuca located northern to the source, all strikes are selected.

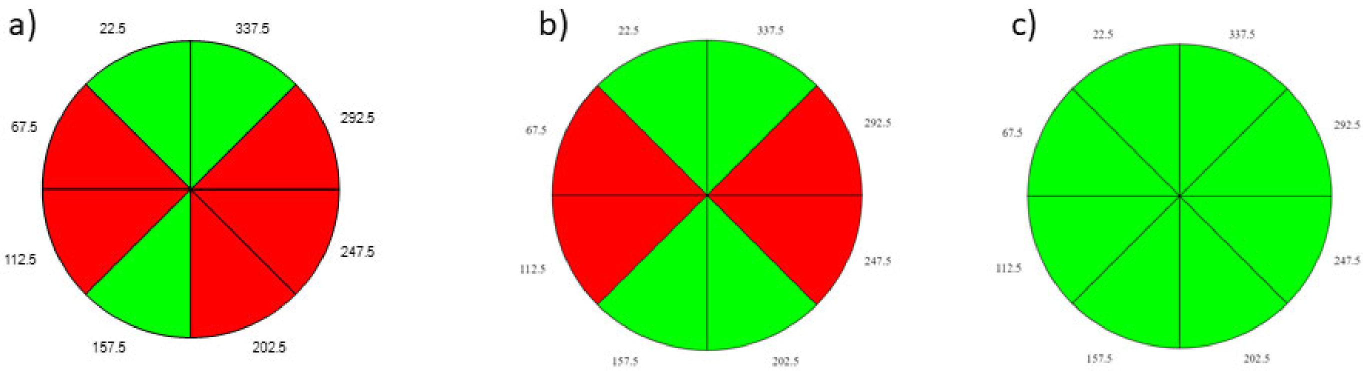

Figure 6 – Values of the strike (green) and non-selected ones (red) in the selected scenarios; a) Siracusa, b) Crotone, and c) Santa Maria di Leuca.

DIP

For Siracusa and Crotone (Figure 7a) and 7b), respectively), the DE search engine has identified as relevant all dip angles in the selected scenarios: thus, dip is not a distinguishing characteristic of the scenarios.

On the other hand, for Santa Maria di Leuca, the DE excludes the less steep dip angles, because the orientation is associated with a less impactful hazard than Siracusa and Crotone. Indeed, a steeper dip angle tends to produce a vertical deformation with a higher tsunamigenic potential.

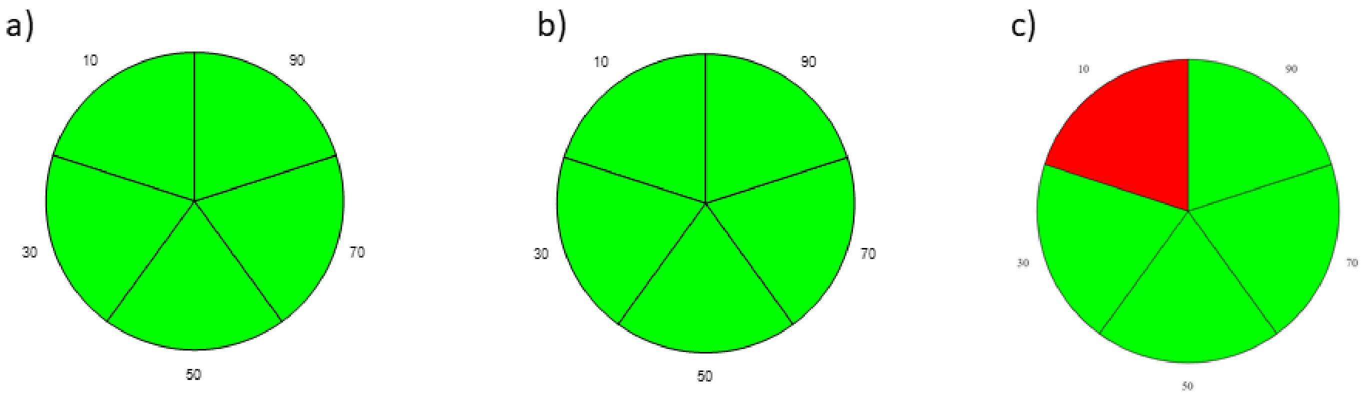

Figure 7 – Values of the dip (green) and non-selected ones (red) in the selected scenarios; a) Siracusa, b) Crotone, and c) Santa Maria di Leuca.

RAKE

For all the target sites considered, the normal and transverse mechanisms (rake values of 90° and 270°) are privileged, as they are associated with a greater coseismic deformation in the component linked to the deformation of the seabed useful for triggering more dangerous tsunamis.

Indeed, only scenarios with rake values of 90° and 270° have been selected (green in Figure 8). This result is expected, since dip-slip earthquakes usually generate a significant deformation of the sea bottom, thus generating higher tsunami waves.

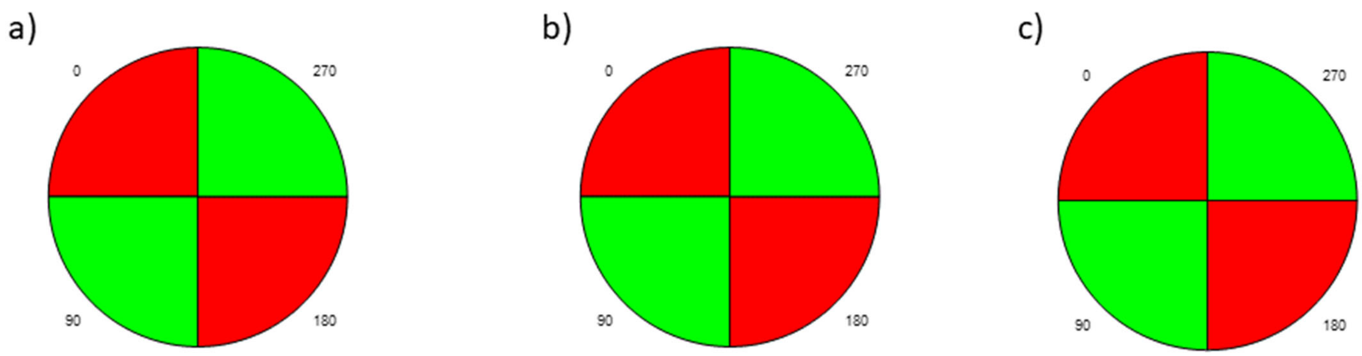

Figure 8 – Values of the rake (green) and non-selected ones (red) in the selected scenarios; a) Siracusa, b) Crotone, and c) Santa Maria di Leuca.

Without loss of generality, as a result of the MODE selection for the target site of Siracusa, the analyst may simulate the scenarios characterised by:

- magnitude $x_1 \in (6.5000, 6.8012, 7.0737, 7.3203, 7.5435, 7.7453)$
- depth $x_2 \in (1, 7.56, 9.43, 11.58, 14.12)$;
- strike $x_3 \in (22.5, 157.5, 337.5)$;
- dip $x_4 \in (10, 30, 50, 70, 90)$;
- rake $x_5 \in (90, 270)$;
- area $x_6 \in (318.5, 638.11, 1194.98, 2108.29, 3524.55, 5608.92)$;
- length $x_7 \in (22.68, 34.39, 50.10, 70.44, 95.87, 126.69)$;
- slip $x_8 \in (0.67, 0.95, 1.30, 1.73, 2.24, 2.82)$.

These results are expected, based on the tsunamigenic capability of earthquakes (see (Grezio et al. 2017) and references therein). They depend both on the particular case study analysed and on the specific tsunami threshold of $\psi_{\bar{a}} \geq 1m$ chosen. Larger tsunami intensities, e.g., $\psi_{\bar{a}} \geq 10m$, would have involved different (probably larger) magnitudes. On the other hand, the results for the Strike, Dip, and Rake angles are probably more general, and they are possibly still valid for larger tsunami intensities. Similar conclusions may be drawn for the other target sites Crotone and Santa Maria di Leuca.

5. CONCLUSIONS

In this work, a novel approach for reducing the number of seismic scenarios to be considered for SPTHA has been presented. The approach is a wrapper-based feature selection heuristic approach based on MODEA. It selects the relevant features of the seismic scenarios to be simulated.

The proposed approach has been applied to a case study regarding the estimation of the annual rate of exceedance of a height threshold $\tilde{\psi} = 1m$ of tsunami waves caused by crustal earthquakes that might be generated on the Kefalonia-Lefkada region in North-western Greece and propagated to a set of target sites on the southern coast of Italy.

The proposed approach is shown capable of significantly reducing the number of features to describe the seismic source variability and the number of scenarios to be considered in the analysis, without affecting the accuracy of the estimate of the annual rate of exceedance. A geophysical interpretation of the results has also been provided.

Further research work will be devoted to the comparison of the proposed approach to other existing methods that may be applied with similar goals, e.g., a standard disaggregation procedure (Bazzurro and Cornell 1999; Gibbons et al. 2020; Selva et al. 2016).

Appendix A. Differential Evolution

Differential Evolution (DE) is a parallel direct search method which utilizes NP D-dimensional parameter vectors $x_{i,G}, i = 1,2, \dots, NP$ as a population for each generation G. NP does not change during the minimisation process. The initial vector population is chosen randomly and should cover the entire parameter space. DE generates new vectors by adding the weighted difference between two population vectors to a third vector in an operation called "mutation". The mutated vector's parameters are then mixed with the elements of another predetermined vector, the target vector, to yield the so-called trial vector, in an operation referred to as "crossover". If the trial vector yields a lower cost function value than the target vector, the trial vector replaces the target vector in the following generation. This last operation is called selection. Each population vector has to serve once as the target vector so that NP competitions take place in one generation. DE's basic strategy can be described as follows.

Mutation

For each target vector $x_{i,G}, i = 1,2, \dots, NP$, a mutant vector is generated according to:

$$v_{i,G+1} = x_{r_1,G} + F \cdot \left(x_{r_2,G} - x_{r_3,G}\right) \tag{A.1}$$

with random indexes $r_1, r_2, r_3 \in \{1,2, \dots, NP\}$, integer, mutually different and $F > 0$. The randomly chosen integers r_1, r_2, r_3 are also chosen to be different from the running index i, so that NP must be greater or equal to four to allow for this condition. F is a real and constant factor $\in [0, 2]$ which controls the amplification of the differential variation $\left(x_{r_2,G} - x_{r_3,G}\right)$.

Crossover

In order to increase the diversity of the perturbed parameter vectors, crossover is introduced. To this end, the trial vector:

$$u_{i,G+1} = \left(u_{1i,G+1}, u_{2i,G+1}, \dots, u_{Di,G+1}\right) \tag{A.2}$$

is formed, where

$$u_{ji,G+1} = \begin{cases} v_{ji,G+1} & if\ (randb(j) \le CR)\ or\ j = rnbr(i) \\ x_{ji,G} & if\ (randb(j) > CR)\ and\ j \ne rnbr(i) \end{cases} \tag{A.3}$$

$$j = 1,2, \dots, D$$

In (A.3), $randb(j)$ is the j-th evaluation of a uniform random number generator with outcome ∈ $[0;\ 1]$. CR is the crossover constant ∈ $[0;\ 1]$ and has to be determined by the user. $rnbr(i)$ is a randomly chosen index ∈ $1, 2, \dots, D$ which ensures that $u_{i,G+1}$ gets at least one parameter from $v_{i,G+1}$.

Selection

To decide whether or not it should become a member of generation $G + 1$, the trial vector $u_{i,G+1}$ is compared to the target vector $x_{i,G}$ using the greedy criterion. If vector $u_{i,G+1}$ yields a smaller cost function value than $x_{i,G}$, then $x_{i,G+}$ is set to $u_{i,G+1}$; otherwise, the old value $x_{i,G}$ is retained.

The above scheme is not the only variant of DE which has proven to be useful. In order to classify the different variants, the notation: DE/x/y/z is introduced where: x specifies the vector to be mutated which currently can be “rand” (a randomly chosen population vector) or “best” (the vector of lowest cost from the current population); y is the number of difference vectors used; z denotes the crossover scheme. The current variant is “bin” (Crossover due to independent binomial experiments). Using this notation, the basic DE-strategy described can be written as DE/rand/1/bin.

This whole section is extracted from (Storn and Price 1997).

Appendix B

DEPTH

The DE search engine has identified as relevant mainly those scenarios characterised by a depth value of $1km$, along with a few scenarios characterised by depth values of $7.56km$, $9.43km$, $11.58km$, and $14.12km$ calculated in line with NEAMTHM18 documentation (Basili et al. 2018, 2021) (green in Figure 9). This result is justified by the dependence of the depth on the magnitude: a depth equal to $1km$ is considered for all magnitudes whereas larger depths, instead, are modelled for smaller magnitudes only, that have been found as important (see Figure 9).

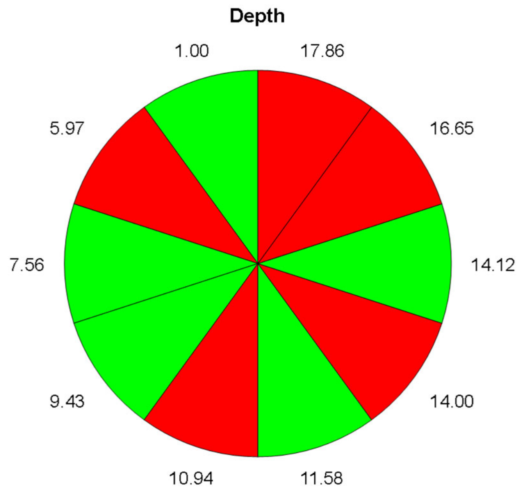

Figure 9 – Values of the depth (green) and non-selected ones (red) in the selected scenarios

AREA

Area values are computed relying on empirical scaling relationships from magnitude (e.g. $\log Area = A + B \times Magnitude$), using different relationships for dip-slip ($x_5 = 90°, 270°$) and strike-slip ($x_5 = 0°, 180°$) earthquakes (Basili et al. 2008). Only the scenarios with area values of $318.5km^2$, $638.11km^2$, $1194.98km^2$, $2108.29km^2$, $3524.55km^2$, $5608.92km^2$ (green in Figure 10) have been selected, i.e., those scenarios corresponding to small magnitude dip-slip earthquakes.

In other words, larger area values, corresponding to larger magnitudes, have not been coherently selected as well as smaller area values, corresponding to smaller magnitudes and strike-slip earthquakes.

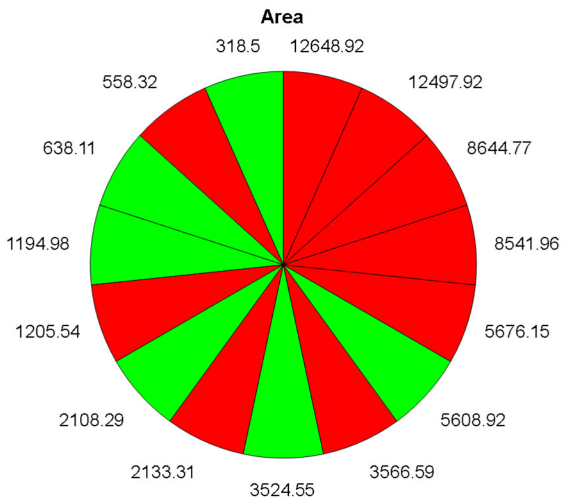

Figure 10 – Values of the area (green) and non-selected ones (red) in the selected scenarios

LENGTH

Length values are computed relying on empirical scaling relationships from magnitude (e.g. $\log Length = A + B \times Magnitude$), using different relationships for dip-slip ($x_5 = 90°, 270°$) and strike-slip ($x_5 = 0°, 180°$) earthquakes (Basili et al. 2008). Only the scenarios with length values of $22.68km$, $34.39km$, $50.10km$, $70.44km$, $95.87km$, $126.69km$ (green in Figure 11) have been selected, i.e., those scenarios corresponding to small magnitude dip-slip earthquakes. In other words, larger length values, corresponding to larger magnitudes, have not been coherently selected as well as smaller length values, corresponding to smaller magnitudes and strike-slip earthquakes.

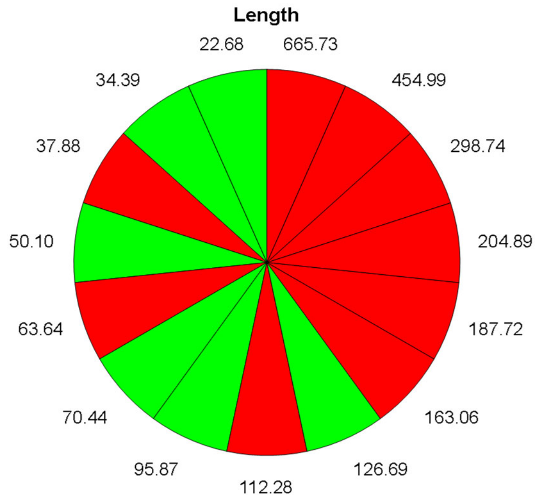

Figure 11 – Values of the length (green) and non-selected ones (red) in the selected scenarios

SLIP

Slip values are computed relying on empirical scaling relationships from magnitude (e.g. $Slip \propto Magnitude/Area$), using different relationships for dip-slip ($x_5 = 90°, 270°$) and strike-slip ($x_5 = 0°, 180°$) earthquakes (Basili et al. 2008). Only the scenarios with slip values of 0.67, 0.95, 1.30, 1.73, 2.24, 2.82 (green in Figure 12) have been selected, i.e., those scenarios corresponding to small magnitude dip-slip earthquakes. In other words, larger slip values, corresponding to larger magnitudes, have been coherently selected as well as smaller slip values, corresponding to smaller magnitudes and strike-slip earthquakes.

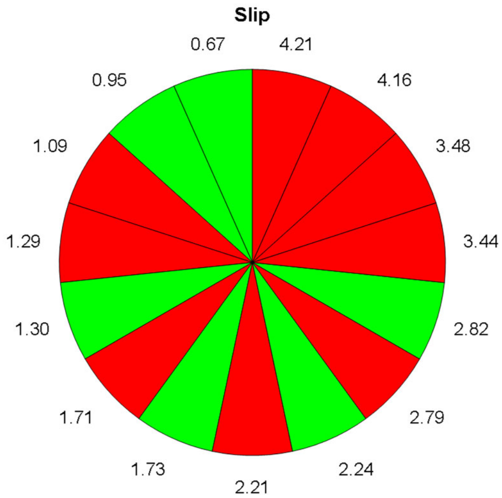

Figure 12 – Values of the slip (green) and non-selected ones (red) in the selected scenarios

ACKNOWLEDGEMENT

This study was developed within the research project “Assessment of Cascading Events triggered by the Interaction of Natural Hazards and Technological Scenarios involving the release of Hazardous Substances” funded by MIUR - Italian Ministry for Scientific Research under the PRIN 2017 program (grant 2017CEYPS8).

REFERENCES

Geist, E. L., & Parsons, T. (2006). Probabilistic analysis of tsunami hazards. Natural Hazards, 37(3), 277-314.

Heidarzadeh, M., Kijko, A. (2011). A probabilistic tsunami hazard assessment for the Makran subduction zone at the northwestern Indian Ocean. Natural Hazards, 56 (3), 577-593.

Gopinathan, D., Heidarzadeh, M., Guillas, S. (2021). Probabilistic Quantification of tsunami current hazard using statistical emulation. Philosophical Transactions of the Royal Society A, 477, 20210180. https://doi.org/10.1098/rspa.2021.0180.

Salmanidou, D.M., Heidarzadeh, M., Guillas, S. (2019). Probabilistic landslide-generated tsunamis in the Indus Canyon, NW Indian Ocean, using statistical emulation. Pure and Applied Geophysics, 176, 3099-3114, https://doi.org/10.1007/s00024-019-02187-3.

Ahmed, Al-Ani. 2005. "Feature Subset Selection Using Ant Colony Optimization.", *PROCEEDINGS OF WORLD ACADEMY OF SCIENCE, ENGINEERING AND TECHNOLOGY VOLUME 4 FEBRUARY 2005 ISSN 1307-6884*.

Bacchi, Vito, Hervé Jomard, Oona Scotti, Ekaterina Antoshchenkova, Lise Bardet, Claire-Marie Duluc, and Hélène Hebert. 2020. "Using Meta-Models for Tsunami Hazard Analysis: An Example of Application for the French Atlantic Coast.", *Frontiers in Earth Science* 8:41.

Basili, R., B. Brizuela, A. Herrero, S. Iqbal, S. Lorito, F. E. Maesano, S. Murphy, P. Perfetti, F. Romano, A. Scala, J. Selva, M. Taroni, H. K. Thio, M. M. Tiberti, R. Tonini, M. Volpe, S. Glimsdal, C. B. Harbitz, F. Løvholt, and F. Oueslati. 2018. "NEAM Tsunami Hazard Model 2018 (NEAMTHM18): Online Data of the Probabilistic Tsunami Hazard Model for the NEAM Region from the TSUMAPS-NEAM Project."

Basili, R., B. Brizuela, A. Herrero, S. Iqbal, S. Lorito, F. E. Maesano, S. Murphy, P. Perfetti, F. Romano, A. Scala, J. Selva, M. Taroni, H. K. Thio, M. M. Tiberti, R. Tonini, M. Volpe, S. Glimsdal, C. B. Harbitz, F. Løvholt, and F. … Oueslati. 2021. "The Making of the NEAM Tsunami Hazard Model 2018 (NEAMTHM18)." *Frontiers in Earth Science* 8:753.

Basili, Roberto, Gianluca Valensise, Paola Vannoli, Pierfrancesco Burrato, Umberto Fracassi, Sofia Mariano, Mara Monica Tiberti, and Enzo Boschi. 2008. "The Database of Individual Seismogenic Sources (DISS), Version 3: Summarizing 20 Years of Research on Italy's Earthquake Geology." *Tectonophysics* 453(1–4):20–43.

Bazzurro, Paolo and C. Allin Cornell. 1999. "Disaggregation of Seismic Hazard." *Bulletin of the Seismological Society of America* 89(2):501–20.

Gailler, A., H. Hébert, F. Schindelé, and D. Reymond. 2018. "Coastal Amplification Laws for the French Tsunami Warning Center: Numerical Modeling and Fast Estimate of Tsunami Wave Heights Along the French Riviera." *Pure and Applied Geophysics* 175(4):1429–44.

Gibbons, Steven J., Stefano Lorito, Jorge Macías, Finn Løvholt, Jacopo Selva, Manuela Volpe, Carlos Sánchez-Linares, Andrey Babeyko, Beatriz Brizuela, Antonella Cirella, Manuel J. Castro, Marc de la Asunción, Piero Lanucara, Sylfest Glimsdal, Maria Concetta Lorenzino, Massimo Nazaria, Luca Pizzimenti, Fabrizio Romano, Antonio Scala, Roberto Tonini, José Manuel

González Vida, Malte Vöge, Joanna Faure Walker, Alberto Armigliato, de M. la Asunción, and Manuel J. González Vida. 2020. “Probabilistic Tsunami Hazard Analysis: High Performance Computing for Massive Scale Inundation Simulations.” *Front. Earth Sci* 8:1.

Glimsdal, S., F. Løvholt, C. B. Harbitz, F. Romano, S. Lorito, S. Orefice, B. Brizuela, J. Selva, A. Hoechner, M. Volpe, A. Babeyko, R. Tonini, M. Wronna, and R. Omira. 2019. “A New Approximate Method for Quantifying Tsunami Maximum Inundation Height Probability.” *Pure and Applied Geophysics* 176(7):3227–46.

González, F. I., E. L. Geist, B. Jaffe, U. Kânog, H. Mofjeld, C. E. Synolakis, V. V Titov, D. Arcas, D. Bellomo, D. Carlton, T. Horning, J. Johnson, J. Newman, T. Parsons, R. Peters, C. Peterson, G. Priest, A. Ve, J. Weber, F. Wong, and A. Yalciner. 2009. “Probabilistic Tsunami Hazard Assessment at Seaside, Oregon, for near- and Far-Field Seismic Sources.” *J. Geophys. Res* 114:11023.

Grezio, Anita, Andrey Babeyko, Maria Ana Baptista, Jörn Behrens, Antonio Costa, Gareth Davies, Eric L. Geist, Sylfest Glimsdal, Frank I. González, Jonathan Griffin, Carl B. Harbitz, Randall J. LeVeque, Stefano Lorito, Finn Løvholt, Rachid Omira, Christof Mueller, Raphaël Paris, Tom Parsons, Jascha Polet, William Power, Jacopo Selva, Mathilde B. Sørensen, and Hong Kie Thio. 2017. “Probabilistic Tsunami Hazard Analysis: Multiple Sources and Global Applications.” *Reviews of Geophysics* 55(4):1158–98.

Grezio, Anita, Laura Sandri, Warner Marzocchi, Andrea Argnani, Paolo Gasparini, and Jacopo Selva. 2012. “Probabilistic Tsunami Hazard Assessment for Messina Strait Area (Sicily, Italy).” *Natural Hazards* 64(1):329–58.

Di Maio, Francesco, Mario Belotti, Manuela Volpe, Jacopo Selva, and Enrico Zio. 2021. “Seismic Probabilistic Tsunami Hazard Assessment: A Novel Approach Based on Parallel Density Scanned Adaptive Kriging.”

Fakhruddin, B., Kintada, K., Tilley, L. 2021. "Probabilistic tsunami hazard and exposure assessment for the pacific islands- Fiji". *International Journal of Disaster Risk Reduction*, 64, art. no. 102458, .

Lorito, S., J. Selva, R. Basili, F. Romano, M. M. Tiberti, and A. Piatanesi. 2015. “Probabilistic Hazard for Seismically Induced Tsunamis: Accuracy and Feasibility of Inundation Maps.” *Geophysical Journal International* 200(1):574–88.

Park, H., Alam, M.S., Cox, D.T., Barbosa, A.R., van de Lindt, J.W. 2021. "Probabilistic seismic and tsunami damage analysis (PSTDA) of the Cascadia Subduction Zone applied to Seaside, Oregon". *International Journal of Disaster Risk Reduction*, 35, art. no. 101076.

Reynoso-Meza, Gilberto. 2021. “Multi-Objective Optimization Differential Evolution Algorithm.”

Rikitake, T. and I. Aida. 1988. “Tsunami Hazard Probability in Japan.” *Bulletin of the Seismological Society of America* 78(3):1268–78.

Selva, J., R. Tonini, I. Molinari, M. M. Tiberti, F. Romano, A. Grezio, D. Melini, A. Piatanesi, R. Basili, and S. Lorito. 2016. “Quantification of Source Uncertainties in Seismic Probabilistic Tsunami Hazard Analysis (SPTHA).” *Geophysical Journal*

International 205(3):1780–1803.

Storn, Rainer and Kenneth Price. 1997. “Differential Evolution - A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces.” *Journal of Global Optimization* 11(4):341–59.

Suppasri, A., Maly, E., Kitamura, M., Syamsidik, Pescaroli, G., Alexander, D., Imamura, F. 2021. "Cascading disasters triggered by tsunami hazards: A perspective for critical infrastructure resilience and disaster risk reduction". *International Journal of Disaster Risk Reduction*, 66, art. no. 102597,

Volpe, Manuela, Stefano Lorito, Jacopo Selva, Roberto Tonini, Fabrizio Romano, and Beatriz Brizuela. 2019. “From Regional to Local SPTHA: Efficient Computation of Probabilistic Tsunami Inundation Maps Addressing near-Field Sources.” *Natural Hazards and Earth System Sciences* 19(3):455–69.

Zio, E., P. Baraldi, and G. Gola. 2007. “Ensemble Feature Selection for Diagnosing Multiple Faults in Rotating Machinery.” *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability* 221(1):29–41.

Zio, E., P. Baraldi, and N. Pedroni. 2006. “Selecting Features for Nuclear Transients Classification by Means of Genetic Algorithms.” *IEEE Transactions on Nuclear Science* 53(3):1479–93.