POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

EXECUTIVE SUMMARY OF THE THESIS

Advanced Sensitivity Analysis Methods for Seismic-Induced Tsunami NaTech Risk Assessment

TESI MAGISTRALE IN SAFETY AND PREVENTION ENGINEERING IN THE PROCESS INDUSTRY – INGEGNERIA DELLA PREVENZIONE E DELLA SICUREZZA NELL'INDUSTRIA DI PROCESSO

AUTHOR: NICOLA GALLO

ADVISOR: ENRICO ZIO

ACADEMIC YEAR: 2020-2021

Introduction

NaTech events (Natural Hazard Triggering Technological Disasters) are industrial accidents triggered by natural hazards which may lead to losses of hazardous materials with potentially tremendous impact on the environment and the surrounding population. Seismic-induced tsunami NaTech risk assessment entails the seismic sources to be characterised and modelled in support of the seismic-induced tsunamis modelling and simulation needed for a Seismic Probabilistic Tsunami Hazard Analysis (SPTHA). In this thesis, we propose two Sensitivity Analysis (SA) methods to deal with the computational issues related with:

1. The identification of the model parameters most affecting the Peak Ground Acceleration (PGA) which ultimately determines the height of the tsunami wave height;
2. The identification most relevant features of the seismic model, for deciding a priori the seismic scenarios to be simulated.

With respect to the first issue, we propose a novel Bootstrapped Modularised Global Sensitivity Analysis (BMGSA) method. The method is tested on a benchmark case study. The results are compared with a standard variance-based Global SA method. The strength of the proposed method is that its application only requires input-output data and not the direct accessibility to the code.

With respect to the second 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 a Multi-Objective Differential Evolution Algorithm (MODEA) and is developed with reference to a case study whose objective of the analysis is calculating the annual rate of a 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 target site on the eastern coast of Sicily (Siracusa). The comparison between the mean values of annual rate of exceedance of the tsunami wave height estimated considering only the selected scenarios and the full set of scenarios shows that the proposed approach allows a significant reduction of the number of scenarios with half of the features to be considered, and with no appreciable loss of accuracy.

To manage tsunami threat, tsunami hazard and risks methodologies have been developed through time to quantify the tsunami hazard and the potential consequent risks [1]. Early on, "worst credible"/ "worst case" scenarios approaches have been adopted [2] that have proven to be limited in modelling seismic sources as well as tsunamis, due to the large uncertainty, both epistemic and aleatory, given by the scarcity of tsunami observations [3]. To overcome "worst credible"/ "worst case" scenarios analyses, SPTHA is aimed at estimating, for a certain location, the annual rate of exceedance of a seismic-induced tsunami wave with respect to a predefined threshold. SPTHA relies on computationally demanding numerical simulations of seismic-induced tsunami generation and propagation, and coastal areas inundation.

SPTHA entails performing:

1. Seismic sources characterisation and modelling,
2. Seismic-induced tsunamis modelling and simulation.

Probabilistic Seismic Hazard Analysis (PSHA) consists of assessing at a given target location and for a given exposure time window ΔT, the probability that a given intensity measure (IM) of the ground motion, typically the PGA, exceeds a threshold value γ [4]. The output of the PSHA are hazard curves, defined by quantifying the mean annual rates of exceedance of a set of IM values. Considering PGA as the IM and assuming a Poisson process, as the model of earthquake occurrence, with parameter λ_H denoting the mean annual rate of exceedance of the γ-th PGA level, the probability of interest is calculated:

$$P(PGA > \gamma, \Delta T) = 1 - \exp[-\lambda_H(PGA > \gamma)\Delta T] \tag{1}$$

Since the propagation of the earthquake wave in the soil is typically evaluated by empirical relationships, called Ground Motion Prediction Equations (GMPEs), λ_H is quantified by means of the total probability theorem as [5]:

$$\lambda_H(PGA > \gamma) = \lambda \int_{m_{min}}^{m_{max}} \int_0^r P(PGA > \gamma | m, r) f_m(m) f_r(r) dm dr \tag{2}$$

where λ is the mean annual rate of earthquake occurrence at a given source location (i.e., the number of occurrence of earthquakes with intensity of PGA above a given threshold per year); the distribution $f_m(m)$ describes the probability distribution of different earthquake magnitudes, typically assumed to follow a truncated Gutenberg-Richter distribution within the interval of values $[m_{min}; m_{max}]$ and slope parameter b [5]; $f_r(r)$ describes the probability distribution of the source-to-target distance r, assuming a spatial distribution for earthquakes [4].

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 $\bar{a}$ [1]. Each tsunami is assumed to be generated by a seismic scenario σ_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:

$$P_e = Pr(\psi_{\bar{a}} \geq \tilde{\psi}; \Delta T) \approx 1 - exp(-\Lambda(\psi_{\bar{a}} \geq \tilde{\psi})\,\Delta T) \tag{3}$$

where $\Lambda(\psi_{\bar{a}} \geq \tilde{\psi})$ 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 $Pr(\psi_{\bar{a}} \geq \tilde{\psi}|\sigma_{\bar{x}})$ that the tsunami wave generated by the scenario exceeds $\tilde{\psi}$:

$$\Lambda(\psi_{\bar{a}} \geq \tilde{\psi}) = \int_\Sigma \lambda(\sigma_{\bar{x}}) Pr(\psi_{\bar{a}} \geq \tilde{\psi}|\sigma_{\bar{x}}) d\sigma_{\bar{x}} \tag{4}$$

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. (4) can be approximated as:

$$\Lambda(\psi_{\bar{a}} \geq \tilde{\psi}) \approx \sum_{q=1}^{Q} \lambda\left(\sigma_{\bar{x}_q}\right) Pr\left(\psi_{\bar{a}} \geq \tilde{\psi}|\sigma_{\bar{x}_q}\right) \tag{5}$$

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. (5). The mean hazard rate can, then, be evaluated as:

$$\Lambda(\psi_{\bar{a}} \geq \tilde{\psi}) \approx \frac{1}{M} \sum_{m=1}^{M} \sum_{q=1}^{Q} \lambda\left(\sigma_{\bar{x}_q}\right)_m Pr\left(\psi_{\bar{a}} \geq \tilde{\psi}|\sigma_{\bar{x}_q}\right)_m \tag{6}$$

In this thesis, we propose two novel Sensitivity Analysis methods to address the aforementioned computational issues related with SPTHA, namely:

1. A Bootstrapped Modularised method of Global Sensitivity Analysis (GSA) for Probabilistic Seismic Hazard Assessment;
2. A heuristic features selection approach for scenario analysis of a Regional Seismic Probabilistic Tsunami Hazard Assessment.

1. A Bootstrapped Modularised Method of Global Sensitivity Analysis for Probabilistic Seismic Hazard Assessment

To identify the input variables which the output of a seismic model is most sensitive to, assuming that only an input-output dataset is given and with no need of repeating hazard computations, we propose a novel modularised GSA method based on bootstrapping and ensemble strategies (BMGSA), consisting of:

1) Generating D alternative bootstrapped artificial datasets from the available input-output dataset $\bar{\bar{Z}}$ [6];
2) From each d-th alternative dataset and for each input variable X_n, calculating a sensitivity index (here the first-order Sobol index) S_n with the modularised method [7];
3) Aggregating the D individual rankings (one for each alternative dataset) with Bottom-Up/All-Out strategies [8].

The proposed method has been tested on a hypothetical PSHA case study with a point seismic source and a nearby target point, where the hazard intensity corresponding to 10% probability of being exceeded in 50y is to be calculated. The epistemic uncertainty of the PSHA is evaluated with respect to six input parameters, accounting for a total of 16384 alternative computational settings, resulting in $\bar{\bar{Z}} = [16384 \times 7]$. The input parameters are $\bar{X} = (\sigma_{GMPE}, \lambda, m_{max}, m_{min}, b, r)$ where: σ_{GMPE} is the standard deviation of the GMPE, λ is the mean annual rate of seismic activity at the source location (i.e., the number of earthquakes per year of intensity magnitude m) m_{min} and m_{max} are the minimum and the maximum magnitude parameters of the truncated Gutenberg-Richter distribution, whose slope is b [5], and r is the source-to-target distance [5]. The output variable is the IM PGA, i.e., the reference peak ground acceleration at the target location that has annual rate of exceedance λ_H assumed to be equal to 1/475y.

The results have been compared to those obtained by a standard variance-based GSA method [9], which is the state-of-practice approach when the simulation model is available. Notably, the main drivers of the epistemic uncertainty on the reference PGA are m_{min} and σ_{GMPE} (see Table 1). Both ensemble strategies and the standard GSA identify the σ_{GMPE} and m_{min} as the most important variables. The disagreement regarding the ranking for the positions 4-6 may be due to hidden dependences and/or correlations between the input variables, as well as to the quantity of data upon which the rankings are drawn.

Table 1 – Input variables rankings (sample size S=16384).

Rank	Standard GSA	Bottom-Up (BMGSA)	All-Out (BMGSA)	No bootstrap (MGSA)
1	m_{min}	m_{min}	m_{min}	m_{min}
2	σ_{GMPE}	σ_{GMPE}	σ_{GMPE}	σ_{GMPE}
3	λ	λ	λ	r
4	r	m_{max}	m_{max}	m_{max}
5	b	r	r	λ
6	m_{max}	b	b	b

To highlight the important role played by the bootstrapping in obtaining such results, we show the results that would have been obtained with a given input-output dataset $\bar{\bar{Z}}$ of decreasing size ($S = 16384, 4096, 1024$), employing the more transparent AO ensemble strategy. When $\bar{\bar{Z}} = [16384 \times 7]$, as shown in Table 1, the Standard GSA, the BMGSA and the MGSA agree on the identification of m_{min} and σ_{GMPE} as the most important variables, whereas for the third most important variable only Standard GSA and BMGSA agree on λ. Then, the approaches provide different rankings for lower ranking positions.

When $\bar{\bar{Z}} = [4096 \times 7]$ the Standard GSA and the BMGSA agree on the identification of m_{min} and σ_{GMPE} as the most important variables, as well as on third (λ) and fourth (r) most important variables. Then, the approaches provide different rankings for lower ranking positions. The MGSA instead yields a completely different ranking (except for position 4).

Table 2 – Input variables rankings (sample size S=4096).

Rank	Standard GSA	All-Out (BMGSA)	No bootstrap (MGSA)
1	m_{min}	m_{min}	σ_{GMPE}
2	σ_{GMPE}	σ_{GMPE}	m_{min}
3	λ	λ	m_{max}
4	r	r	r
5	b	m_{max}	λ

6	m_{max}	b	b

When $\bar{\bar{Z}} = [1024 \times 7]$ the Standard GSA, the BMGSA and the MGSA agree on the identification of m_{min} as the most important variable, whereas for the second (σ_{GMPE}) and third (λ) most important variables only Standard GSA and BMGSA agree. Then, the approaches provide different rankings for lower ranking positions. Nevertheless, the numerical values of the Sobol indices obtained with the proposed BMGSA may not be considered satisfactory: when the dimension of $\bar{\bar{Z}}$ decreases, the distributions of $S_{n,d}$ become wider (i.e., bootstrap replicates are subject to noise and, as a result, the Sobol indices estimate are not precise) even if the most important variables are correctly identified, a less accurate estimation of the Sobol indices is provided.

Table 3 – Input variables rankings (sample size S=1024).

Rank	Standard GSA	All-Out (BMGSA)	No bootstrap (MGSA)
1	m_{min}	m_{min}	m_{min}
2	σ_{GMPE}	σ_{GMPE}	λ
3	λ	λ	σ_{GMPE}
4	r	b	b
5	b	r	m_{max}
6	m_{max}	m_{max}	r

As general conclusion, we can state that bootstrapping allows relying on a very small dataset. Indeed, $S = 4096$ yields already a very satisfactory estimate of the Sobol indices values (compared with the GSA estimates). Thus, as a general recommendation, we may conclude that a ratio of 4:1 of $S:D$ (dataset size vs number of bootstrap replicates) is enough to guarantee satisfactory results, without resorting further to demanding computations.

For the case study at hand, we can conclude that, m_{min}, σ_{GMPE}, and λ have been identified as the input variables which most influence the reference PGA, whereas r, b, and m_{max} influence is negligible.

We underline that, obviously, the numerical results obtained are relative to the specific case and cannot be generalised to other PSHA case studies.

2. A Heuristic Features Selection Approach for Scenario Analysis of a Regional Seismic Probabilistic Tsunami Hazard Analysis

To select the relevant features of the seismic scenarios to be simulated for an accurate SPTHA, we propose a wrapper-based feature selection heuristic approach based on MODEA [10]. The proposed approach is developed with reference to a case study whose objective of the analysis is calculating 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. The target site $\bar{a}$ for the propagation of the wave is at Siracusa, on the eastern coast of Sicily. The case study considers the crustal seismicity generated in the Kefalonia-Lefkada region. This source area comprises a total of $Q_{tot} = 23272$ seismic scenarios and $M = 1000$ alternative models for the calculation of $\Lambda(\psi_{\bar{a}} \geq 1m)$.

Earthquakes are assumed to be generated at specific epicentral locations with different magnitudes, depths, and faulting mechanisms. Without loss of generality, the following assumptions are made:

i. Threshold is of $\tilde{\psi} = 1m$ at 50m from the coastline.
ii. One epicentral location is considered, since a large number $Q = 721$ of seismic scenarios $\sigma_{\bar{x}}$ is available.

Each $\sigma_{\bar{x}}$ is characterised by the set of parameters $\bar{x} = (x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8)$ [11]. To alleviate the computational burden of the SPTHA, the procedure sketched in Figure 1 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. (6). 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 DE engine [10] iteratively searches for candidates sets of scenarios, among the original dataset of $Q = 721$ scenarios, whose performance is evaluated with respect to a given cost function. Once the optimal set of scenarios is identified, their common features are retrieved by statistical analysis.

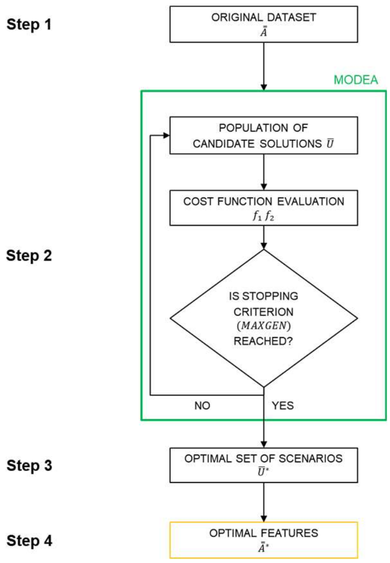

Figure 1: Wrapper approach for optimal set of scenarios selection based on MODEA

The procedure is explained in detail here below.

Step 1: Consider the original dataset

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

$$\bar{\bar{A}} = \begin{pmatrix} x_{1,1} & \cdots & x_{8,1} & \Lambda(\psi_{\bar{a}} \geq 1m|\sigma_{\bar{x}_1}) \\ \vdots & \vdots & \vdots & \vdots \\ x_{1,q} & \cdots & x_{8,q} & \Lambda\left(\psi_{\bar{a}} \geq 1m|\sigma_{\bar{x}_q}\right) \\ \vdots & \vdots & \vdots & \vdots \\ x_{1,Q} & \cdots & x_{8,Q} & \Lambda\left(\psi_{\bar{a}} \geq 1m|\sigma_{\bar{x}_Q}\right) \end{pmatrix} \quad (7)$$

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) [12]. 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. 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. In our case, the objective functions considered are:

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

$$f_1 = \sum_{q=1}^{Q} U_q \quad (8)$$

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

$$f_2 = \left(\Lambda(\psi_{\bar{a}} \geq 1m) - \Lambda^*(\psi_{\bar{a}} \geq 1m)\right)^2 \quad (9)$$

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

$$\Lambda^*(\psi_{\bar{a}} \geq 1m) = \sum_{q=1}^{Q} \lambda\left(\sigma_{\bar{x}_q}\right) Pr\left(\psi_{\bar{a}} \geq 1m|\sigma_{\bar{x}_q}\right) U_q \quad (10)$$

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) is selected from the Pareto optimal front, 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):

$$\bar{\bar{A}}^* = \bar{\bar{A}} \circ \bar{U}^* \quad (11)$$

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).

In this work, the preferred solution $\bar{U}_1^*$ is the one that 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 $\Lambda(\psi_{\bar{a}} \geq 1m|H_{14})$ i.e., most of the $Q = 721$ seismic scenarios bring a negligible contribution to the estimation of $\Lambda(\psi_{\bar{a}} \geq 1m)$ but increase the computational burden. As a result of the MODE selection, 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 [13] 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.

3. Conclusions

In this thesis, we have proposed two SA methods to deal with the computational issues of the SPTHA related with:

1. The identification of the model parameters most affecting the PGA;
2. The identification most relevant features of the seismic model, for deciding a priori the seismic scenarios to be simulated.

We have proposed a novel Bootstrapped Modularised Sensitivity Analysis (BMGSA) method based on bootstrapping, MGSA and ensemble strategies to identify the input parameters which the output of a PSHA model is most sensitive to, assuming that only an input-output dataset is given whereas the model is not available. The novelty and strength of the proposed BMGSA method is that to be applied it only needs data and not the source simulation code. The capability of the proposed method is tested on a benchmark case study. The results have been compared with a standard variance-based GSA method of literature, showing that the proposed method and the standard GSA agree on the identification of the three by-far most important input variables. Furthermore, the BMGSA has proved to be reliable even when applied to very small datasets. The application of the developed technique to PSHA demonstrates its capability of scoring correctly the importance of existing epistemic uncertainty factor, needing only the input and the output data. This allows applying the technique to any hazard model in which epistemic uncertainty is to be evaluated. Its systematic application to hazard studies to detect the most influential parameters, would allow hazard practitioners to both improve the sanity checks during the assessment and to focus future research toward the reduction of epistemic uncertainty by further characterisation of the important factors.

Then, a novel approach for reducing the number of seismic scenarios to be analysed 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 with reference to 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 target site $\bar{a}$ on the eastern coast of Sicily (Italy). The proposed approach is shown to be able to significantly reduce the number of features describing the seismic source variability and, thus, 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 been provided.

References

[1] A. Grezio *et al.*, "Probabilistic Tsunami Hazard Analysis: Multiple Sources and Global Applications," *Reviews of Geophysics*, vol. 55, no. 4. pp. 1158–1198, 2017.

[2] S. Tinti and A. Armigliato, "The use of scenarios to evaluate the tsunami impact in southern Italy," *Mar. Geol.*, vol. 199, no. 3–4, pp. 221–243, Sep. 2003.

[3] E. L. Geist and T. Parsons, "Undersampling power-law size distributions: effect on the assessment of extreme natural hazards," vol. 72, pp. 565–595, 2014.

[4] R. K. Mcguire, "Probabilistic seismic hazard analysis: Early history," *Earthq. Eng.*

Struct. Dyn. Earthq. Engng Struct. Dyn, vol. 37, pp. 329–338, 2008.

[5] J. Selva and L. Sandri, “Probabilistic seismic hazard assessment: Combining Cornell-like approaches and data at sites through Bayesian inference,” *Bull. Seismol. Soc. Am.*, vol. 103, no. 3, pp. 1709–1722, Jun. 2013.

[6] P. Baraldi, G. Gola, E. Zio, D. Roverso, and M. Hoffmann, “A randomized model ensemble approach for reconstructing signals from faulty sensors,” *Expert Syst. Appl.*, vol. 38, no. 8, pp. 9211–9224, Aug. 2011.

[7] E. Plischke, E. Borgonovo, and C. L. Smith, “Global sensitivity measures from given data,” *Eur. J. Oper. Res.*, vol. 226, no. 3, pp. 536–550, May 2013.

[8] F. Di Maio, A. Bandini, E. Zio, S. C. Alberola, F. Sanchez-Saez, and S. Martorell, “Bootstrapped-ensemble-based Sensitivity Analysis of a trace thermal-hydraulic model based on a limited number of PWR large break loca simulations,” *Reliab. Eng. Syst. Saf.*, vol. 153, pp. 122–134, Sep. 2016.

[9] C. Molkenthin, F. Scherbaum, A. Griewank, H. Leovey, S. Kucherenko, and F. Cotton, “Derivative-based global sensitivity analysis: Upper bounding of sensitivities in seismic-hazard assessment using automatic differentiation,” *Bull. Seismol. Soc. Am.*, vol. 107, no. 2, pp. 984–1004, Apr. 2017.

[10] R. Storn and K. Price, “Differential Evolution - A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces,” *J. Glob. Optim.*, vol. 11, no. 4, pp. 341–359, 1997.

[11] J. Selva *et al.*, “Quantification of source uncertainties in Seismic Probabilistic Tsunami Hazard Analysis (SPTHA),” *Geophys. J. Int.*, vol. 205, no. 3, pp. 1780–1803, Jun. 2016.

[12] E. Zio, P. Baraldi, and N. Pedroni, “Selecting features for nuclear transients classification by means of genetic algorithms,” *IEEE Trans. Nucl. Sci.*, vol. 53, no. 3, pp. 1479–1493, 2006.

[13] A. Grezio, L. Sandri, W. Marzocchi, A. Argnani, P. Gasparini, and J. Selva, “Probabilistic tsunami hazard assessment for Messina Strait Area (Sicily, Italy),” *Nat. Hazards*, vol. 64, no. 1, pp. 329–358, Sep. 2012.

4. Acknowledgements

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).