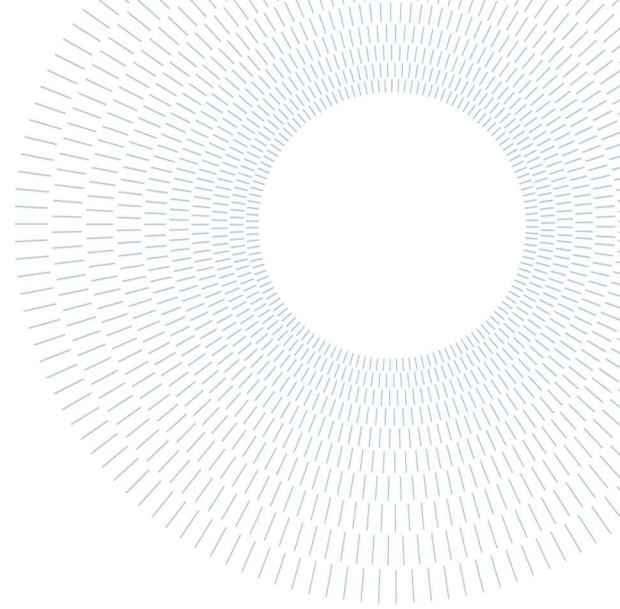




**POLITECNICO
MILANO 1863**

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

Gli eventi NaTech (Natural Hazard Triggering Technological Disasters) sono incidenti industriali innescati da pericoli naturali che possono portare a perdite di materiali pericolosi con un impatto potenzialmente tremendo sull'ambiente e sulla popolazione circostante. La valutazione del rischio di tsunami NaTech indotto dal sisma comporta la caratterizzazione e la modellazione delle sorgenti sismiche a supporto della modellazione e della simulazione degli tsunami indotti dal sisma necessaria per una Seismic Probabilistic Tsunami Hazard Analysis (SPTHA). In questa tesi, proponiamo due metodi di Sensitivity Analysis (SA) per affrontare i problemi computazionali relativi a:

1. L'individuazione dei parametri del modello che influenzano maggiormente la *Peak Ground Acceleration* (PGA) che in ultima analisi determina l'altezza dell'onda di tsunami;
2. L'individuazione delle *feature* del modello sismico più rilevanti per decidere a priori gli scenari sismici da simulare.

Per quanto riguarda il primo problema, proponiamo un nuovo metodo di *Bootstrapped Modularised Global Sensitivity Analysis* (BMGSA). Il metodo è testato su un caso di studio di riferimento. I risultati sono confrontati con un metodo standard di *Global SA* basato sulla varianza. La forza del metodo proposto è che la sua applicazione richiede solo dati di input-output e non l'accessibilità diretta al codice.

Per quanto riguarda il secondo problema, proponiamo un approccio euristico *wrapper-based*, per selezionare l'insieme delle caratteristiche più rilevanti del modello sismico, per decidere a priori gli scenari sismici da simulare. L'approccio proposto si basa su un Algoritmo di Evoluzione Differenziale Multi-Obiettivo (MODEA) ed è sviluppato con riferimento ad un caso di studio il cui obiettivo dell'analisi è il calcolo del tasso annuale di superamento di una soglia di altezza delle onde di tsunami causate da terremoti di subduzione che potrebbero essere generate su un tratto dell'Arco Ellenico e propagate ad un sito target sulla costa orientale della Sicilia (Siracusa). Il confronto tra i valori medi del tasso annuale di superamento dell'altezza dell'onda di tsunami stimato considerando solo gli scenari selezionati e

l'insieme completo degli scenari mostra che l'approccio proposto permette una significativa riduzione del numero di scenari con la metà delle caratteristiche da considerare, e senza apprezzabili perdite di accuratezza.

Per gestire la minaccia degli tsunami, sono state sviluppate nel tempo metodologie per quantificare il pericolo di tsunami e i potenziali rischi conseguenti [1]. Inizialmente, sono stati adottati approcci "worst credible"/"worst case" scenario [2] che si sono dimostrati limitati nella modellizzazione delle sorgenti sismiche e degli tsunami, a causa della grande incertezza, sia epistemica che aleatoria, data dalla scarsità di osservazioni di tsunami [3]. Per superare le analisi "worst credible"/"worst case" scenario, SPTHA stima, per una certa località, il tasso annuale di superamento di soglia predefinita di un'onda di tsunami sismogenato. SPTHA si basa su simulazioni numeriche computazionalmente impegnative della generazione e della propagazione dello tsunami sismico e dell'inondazione delle aree costiere.

SPTHA comporta l'esecuzione di:

1. Caratterizzazione e modellazione delle sorgenti sismiche,
2. Modellazione e simulazione degli tsunami sismogenati.

Probabilistic Seismic Hazard Analysis (PSHA) consiste nel valutare, in una data località e per una data finestra temporale di esposizione ΔT , la probabilità che una data Misura di Intensità (IM) del moto del suolo, tipicamente PGA, superi un valore soglia γ [4]. L'output di PSHA consiste in curve di pericolo, definite quantificando i tassi medi annuali di superamento di un insieme di valori di IM. Considerando PGA come IM e ipotizzando un processo di Poisson, come modello di frequenza del terremoto, con il parametro λ_H che denota il tasso medio annuo di superamento del livello γ -esimo di PGA, si calcola la probabilità di interesse:

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

Poiché la propagazione dell'onda sismica nel suolo è tipicamente valutata tramite relazioni empiriche, chiamate *Ground Motion Prediction Equations* (GMPEs), λ_H è quantificato per mezzo del teorema della probabilità totale come [4]:

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

dove λ è il tasso medio annuo di occorrenza dei terremoti in una data località sorgente (i.e., il numero di occorrenze di terremoti con intensità di PGA superiore a una data soglia all'anno); la distribuzione $f_m(m)$ descrive la distribuzione di probabilità delle diverse magnitudini dei terremoti, tipicamente assunta per seguire una distribuzione tronca di Gutenberg-Richter entro l'intervallo di valori $[m_{min}; m_{max}]$ e il parametro di pendenza b [4]; $f_r(r)$ descrive la distribuzione di probabilità della distanza sorgente-target r , ipotizzando una distribuzione spaziale per i terremoti [5].

SPTHA consiste nello stimare la probabilità che l'altezza ψ di un'onda di tsunami sismogenato superi una soglia $\tilde{\psi}$, in un tempo di esposizione ΔT , in una località con coordinate \bar{a} [1]. Si assume che ogni tsunami sia generato da uno scenario sismico $\sigma_{\bar{x}}$ appartenente allo spazio dei possibili scenari sismici Σ ($\sigma_{\bar{x}} \in \Sigma$), caratterizzato dai parametri \bar{x} e che si verifica con frequenza annuale $\lambda(\sigma_{\bar{x}})$ considerando un processo di Poisson per il verificarsi dell'evento di superamento dell'altezza soglia dell'onda nel tempo, la probabilità di superamento P_e può essere scritta come:

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

dove $\Lambda(\psi_{\bar{a}} \geq \tilde{\psi})$ è il rateo annuale di occorrenza di uno tsunami di intensità $\psi_{\bar{a}} \geq \tilde{\psi}$ nella località \bar{a} . Questo tasso è calcolato integrando, sullo spazio Σ , il prodotto tra la frequenza annuale $\lambda(\sigma_{\bar{x}})$ di occorrenza dello scenario sismico $\sigma_{\bar{x}}$ e la probabilità $Pr(\psi_{\bar{a}} \geq \tilde{\psi} | \sigma_{\bar{x}})$ che l'onda di tsunami da generata dallo scenario superi $\tilde{\psi}$:

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

Considerando, senza perdere generalità e per semplicità, un insieme di Q scenari sismici discretizzati $\sigma_{\bar{x}_q}$ ($q = 1, \dots, Q$) con $\lambda(\sigma_{\bar{x}_q})$ e $Pr(\psi_{\bar{a}} \geq \tilde{\psi} | \sigma_{\bar{x}_q})$, l'Eq. (4) può essere approssimata come:

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

Per tenere conto dell'incertezza epistemica, si possono considerare M formulazioni alternative di $\lambda(\sigma_{\bar{x}_q})$ e $Pr(\psi_{\bar{a}} \geq \tilde{\psi} | \sigma_{\bar{x}_q})$, da cui derivano M quantificazioni alternative di entrambi i fattori in Eq. (5). Il tasso di rischio medio può, quindi, essere calcolato come:

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

In questa tesi, proponiamo due nuovi metodi di *Sensitivity Analysis* per affrontare le suddette problematiche computazionali relative a SPTHA, cioè:

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

Per individuare le variabili di input che influenzano maggiormente l'output di un modello sismico, assumendo che sia disponibile solo un set di dati di input-output e senza bisogno di ripetere i calcoli di rischio, proponiamo un nuovo metodo di GSA modularizzato, basato su *bootstrapping* e strategie di *ensemble* (BMGSA), che consiste in:

- 1) Generare artificialmente D set alternativi di dati tramite *bootstrapping* dal set di dati input-output disponibile \bar{Z} [6];
- 2) Calcolare, da ogni d -esimo dataset alternativo e per ogni variabile di input X_n , un indice di sensibilità (in questo caso l'indice di Sobol del primo ordine) S_n con il metodo modularizzato [7];
- 3) Aggregare i D ranking individuali (uno per ogni dataset alternativo) con le strategie *Bottom-Up/All-Out* [8].

Il metodo proposto è stato testato su un ipotetico caso studio di PSHA con una sorgente sismica puntiforme e un punto target vicino, dove l'intensità di pericolo corrispondente al 10% di probabilità di essere superato in 50 anni deve essere calcolato. L'incertezza epistemica di PSHA è valutata rispetto a sei parametri di input, per un totale di 16384 impostazioni di calcolo alternative, risultando in $\bar{Z} = [16384 \times 7]$. I parametri di input sono $\bar{X} = (\sigma_{GMPE}, \lambda, m_{max}, m_{min}, b, r)$ dove: σ_{GMPE} è la deviazione standard di GMPE, λ è il tasso medio annuo di attività sismica nel luogo della sorgente (i.e., il numero di terremoti all'anno di magnitudo

m), m_{min} e m_{max} sono i parametri di magnitudo minima e massima della distribuzione tronca di Gutenberg-Richter, la cui pendenza è b [4], r è la distanza tra sorgente e target [4]. La variabile di output è la IM *PGA*, i.e., la PGA di riferimento nella località target che ha un tasso annuo di superamento λ_H ipotizzato uguale a 1/475y.

I risultati sono stati confrontati con quelli ottenuti da un metodo GSA standard basato sulla varianza [9], che è l'approccio allo stato dell'arte quando il modello di simulazione è disponibile. In particolare, i principali *driver* dell'incertezza epistemica sul riferimento PGA sono m_{min} e σ_{GMPE} (Table 1). Entrambe le strategie di *ensemble* e la standard GSA identificano σ_{GMPE} e m_{min} come le variabili più importanti. Il disaccordo sulla classifica per le posizioni 4-6 può essere dovuto a dipendenze e/o correlazioni nascoste tra le variabili di input, così come alla quantità di dati su cui le classifiche sono stilate.

Table 1 – Classifica variabili di input (dimensione sample $S=16384$).

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

Per evidenziare l'importanza del ruolo del *bootstrapping* per l'ottenimento di questi risultati, mostriamo i risultati che sarebbero stati ottenuti con un dataset \bar{Z} di input-output di dimensioni decrescenti ($S = 16384, 4096, 1024$), impiegando la più trasparente strategia di *ensemble* AO.

Quando $\bar{Z} = [16384 \times 7]$ Standard GSA, BMGSA e MGSA concordano sull'identificare m_{min} e σ_{GMPE} come le variabili più importanti, mentre per la terza più importante solo Standard GSA e BMGSA concordano su λ (Table 1). Gli approcci poi divergono sulle posizioni più basse della classifica.

Quando $\bar{Z} = [4096 \times 7]$ Standard GSA e BMGSA concordano sull'identificare m_{min} e σ_{GMPE} come le variabili più importanti, così come per la terza (λ) e la quarta posizione (r), per poi divergere sulle

posizioni più basse della classifica (Table 2). MGSA, invece, fornisce una classifica completamente diversa (fatta eccezione per la posizione 4).

Table 2 – Classifica variabili di input (dimensione sample $S=4096$).

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

Quando $\bar{Z} = [1024 \times 7]$ Standard GSA, BMGSA e MGSA concordano sull'identificare m_{min} come la variabile più importanti, mentre per la seconda (σ_{GMPE}) e la terza (λ) solo Standard GSA e BMGSA concordano (Table 3). Gli approcci poi divergono sulle posizioni più basse della classifica. Ciononostante, i valori numerici degli indici di Sobol ottenuti con BMGSA non possono essere considerati soddisfacenti: quanto la dimensione di \bar{Z} diminuisce, la distribuzione di $S_{n,d}$ si allarga (in altre parole le repliche di bootstrap sono soggette a rumore e di conseguenza la stima degli indici di Sobol è imprecisa) anche se le variabili più importanti sono correttamente identificate, ne deriva una stima meno precisa degli indici di Sobol.

Table 3 – Classifica variabili di input (dimensione sample $S=1024$).

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

Come conclusione generale, possiamo affermare che il *bootstrapping* permette di basarsi su un set di dati molto piccolo. Infatti, $S = 4096$ produce una stima molto soddisfacente dei valori degli indici di Sobol (rispetto alle stime di GSA). Quindi, come raccomandazione generale, possiamo concludere che un rapporto di 4:1 tra S:D (dimensione del

dataset rispetto al numero di repliche del bootstrap) è sufficiente a garantire risultati soddisfacenti, senza ricorrere a ulteriori calcoli.

Per il caso studio in esame, invece, m_{min} , σ_{GMPE} , e λ risultano essere le variabili di input che influenzano maggiormente la PGA di riferimento, mentre l'influenza di r , b e m_{max} è trascurabile.

Sottolineiamo che, ovviamente, i risultati numerici ottenuti sono relativi al caso specifico e non possono essere generalizzati ad altri casi studio di PSHA.

2. A heuristic features selection approach for the scenario analysis of a regional seismic probabilistic tsunami hazard analysis

Per selezionare le *feature* rilevanti degli scenari sismici da simulare per una SPTHA accurata, proponiamo un approccio euristico *wrapper-based* basato MODEA [10].

L'approccio proposto è sviluppato con riferimento a un caso di studio il cui obiettivo dell'analisi è il calcolo del tasso annuale di superamento di una altezza soglia $\tilde{\psi} = 1m$ di un'onda di tsunami, causata da terremoti di subduzione in una sezione dell'Arco Ellenico. Il sito obiettivo \bar{a} per la propagazione dell'onda è a Siracusa, sulla costa orientale della Sicilia. Il caso studio considera la sismicità crostale generata nella regione di Cefalonia-Lefkada. Questa area sorgente comprende un totale di $Q_{tot} = 23272$ scenari sismici e $M = 1000$ modelli alternativi per il calcolo di $\Lambda(\psi_{\bar{a}} \geq 1m)$.

Si suppone che i terremoti siano generati in specifiche posizioni epicentrali con diverse magnitudo, profondità e meccanismi di faglia. Senza perdita di generalità, sono state fatte le seguenti ipotesi:

- i. Soglia di $\tilde{\psi} = 1m$ a 50m dalla costa.
- ii. Si considera un solo epicentro che fornisce un numero pari a $Q = 721$ di scenari sismici $\sigma_{\bar{x}}$.

Ogni $\sigma_{\bar{x}}$ è caratterizzato da un set di parametri $\bar{x} = (x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8)$ [11]. Per alleviare l'onere computazionale della SPTHA, è stata sviluppata la procedura schematizzata in Figure 1. Innanzitutto, viene risolto un problema di ottimizzazione che produce l'insieme ottimale di scenari sismici che contribuiscono maggiormente a $\Lambda(\psi_{\bar{a}} \geq 1m)$ dell'Eq. (6). Poi, i valori delle loro *feature* identificati. L'ottimizzazione è eseguita con un

approccio euristico *wrapper-based* basato su MODEA in cui il motore DE [10] ricerca iterativamente insiemi di scenari candidati tra il dataset originale di $Q = 721$ scenari, le cui prestazioni sono valutate rispetto a una funzione di costo, fino al raggiungimento di un criterio di arresto specificato. Una volta identificato l'insieme ottimale di scenari, le loro *feature* comuni sono recuperate con un'analisi statistica.

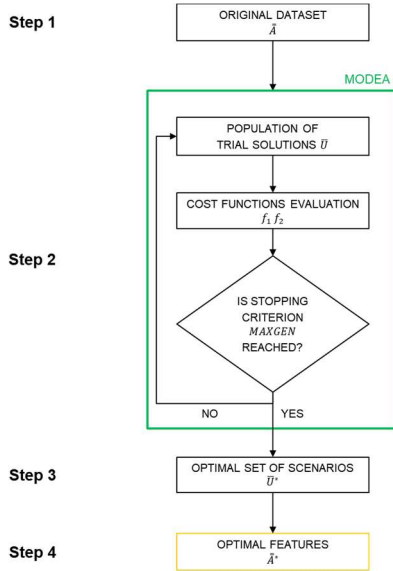


Figure 1: approccio euristico *wrapper-based* basato su MODEA

Nel dettaglio, la procedura consiste in:

Step 1: Considerare il dataset originale

Il dataset originale $\bar{A} = [Q \times 9]$ è:

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

dove $x_{1,q}$ è il valore del parametro x_1 nel q -esimo scenario, $x_{2,q}$ è il valore del parametro x_2 nel q -esimo scenario, etc., mentre $\Lambda(\psi_{\bar{a}} \geq 1m | \sigma_{\bar{x}_q}) = \frac{1}{M} \sum_{m=1}^M \lambda(\sigma_{\bar{x}_q})_m Pr(\psi_{\bar{a}} \geq 1m | \sigma_{\bar{x}_q})_m$ è il tasso annuale (del q -esimo scenario) di superamento della soglia.

Step 2: Applicazione del MODEA per individuare gli scenari più rilevanti

Il MODEA ricerca il minimo globale di un insieme di funzioni obiettivo (di costo) $F = \{f(\cdot)\}$, di uno (o più) vettori decisionali \bar{U} , che sono tipicamente stringhe di cifre binarie [12]. Nel caso di interesse

per questo lavoro, \bar{U} indica se il q -esimo scenario sismico è considerato nella soluzione di prova MO (q -esimo bit uguale a 1) o no (q -esimo bit uguale a 0). La ricerca MODEA viene eseguita campionando inizialmente in modo casuale i bit dei vettori NP che compongono le stringhe della popolazione iniziale.

Poi, iterativamente, la popolazione viene arricchita dalla soluzione \bar{U} che meglio si adatta alle funzioni obiettivo, attraverso un processo di selezione guidato da un insieme di parametri, cioè il fattore di scala F e la probabilità di crossover CR . Nel nostro caso, le funzioni obiettivo considerate sono:

1. Minimizzazione di Q (i.e., il numero di scenari $\sigma_{\bar{x}_q}$ considerati nella soluzione):

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

2. Minimizzazione dell'errore quadratico SE tra frequenza annuale di superamento $\Lambda(\psi_{\bar{a}} \geq 1m)$ e la frequenza annuale di superamento, calcolata considerando i $Q^* = \min(\sum_{q=1}^Q U_q)$ scenari selezionati $\Lambda^*(\psi_{\bar{a}} \geq 1m)$:

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

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

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

La procedura di ricerca termina quando viene raggiunto il criterio di arresto (es. il numero massimo di generazioni $MAXGEN$) viene raggiunto.

Step 3: Set ottimale di scenari

Il vettore soluzione ottimale \bar{U}^* (i.e., il set ottimale di scenari) è selezionato dal fronte di Pareto, in quanto soluzione con minor numero Q^* di *entries* uguali a 1 (i.e., gli scenari considerati nella soluzione di MO).

Step 4: Individuazione delle *feature* ottimali

Per individuare le *feature* più importanti, occorre innanzitutto ricavare la matrice delle *feature* $\bar{A}^* = [Q^* \times 9]$, come prodotto di Hadamard tra il dataset originale \bar{A} e \bar{U}^* (con $(Q - Q^*)$ vettori colonna nulli):

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

Quindi si confronta la matrice \bar{A}^* , colonna per colonna, con il dataset originale \bar{A} per valutare gli elementi comuni (i.e., il subset di *feature* ottimali).

In questo lavoro si predilige la soluzione caratterizzata dal minor numero di scenari $Q^* = 38$ (i.e., una riduzione del 95% rispetto a Q) con un relativamente basso $SE = 8.5^{-30} \text{years}^{-2}$ (i.e., un errore percentuale di 0.085%) nella stima di $\Lambda(\psi_{\bar{a}} \geq 1m | H_{14})$, il che significa che la maggior parte dei $Q = 721$ scenari sismici ha un contributo trascurabile nel calcolo di $\Lambda(\psi_{\bar{a}} \geq 1m)$ seppur aumentandone l'onere computazionale. Un analista, quindi, a seguito di questo risultato, può simulare gli scenari caratterizzati da:

- magnitudo $x_1 \in (6.5000, 6.8012, 7.0737, 7.3203, 7.5435, 7.7453)$;
- profondità $x_2 \in (1, 7.56, 9.43, 11.58, 14.12)$;
- *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 $x_6 \in (318.5, 638.11, 1194.98, 2108.29, 3524.55, 5608.92)$;
- lunghezza $x_7 \in (22.68, 34.39, 50.10, 70.44, 95.87, 126.69)$;
- *slip* $x_8 \in (0.67, 0.95, 1.30, 1.73, 2.24, 2.82)$.

Questi risultati sono attesi, sulla base della capacità tsunamigenica dei terremoti (cfr. [13] e riferimenti ivi contenuti). Essi dipendono sia dal particolare caso di studio analizzato che dalla specifica soglia di tsunami di $\psi_{\bar{a}} \geq 1m$ scelta. Intensità di tsunami più grandi, ad esempio $\psi_{\bar{a}} \geq 10m$ avrebbero comportato magnitudini diverse (probabilmente maggiori). D'altra parte, i risultati per gli angoli *strike*, *dip* e *rake* sono probabilmente più generali, e sono probabilmente ancora validi per intensità di tsunami maggiori.

3. Conclusions

In questa tesi abbiamo proposto due metodi di SA per affrontare i problemi computazionali della SPTHA relativi a:

1. L'individuazione dei parametri del modello che influenzano maggiormente PGA;
2. L'individuazione delle *feature* del modello sismico più rilevanti per decidere a priori gli scenari sismici da simulare.

Abbiamo proposto un nuovo metodo di *Bootstrapped Modularised Sensitivity Analysis* (BMGSA) basato su *bootstrapping*, MGSA e strategie di *ensemble* per identificare i parametri di input che influenzano maggiormente l'output di un modello di PSHA, ipotizzando che sia disponibile solo un dataset di input-output mentre

il modello non è disponibile. La novità e la forza del metodo proposto BMGSA sta nel fatto che per essere applicato necessita solo di dati e non del codice di simulazione. La capacità del metodo proposto è testata su un caso di studio di riferimento. I risultati sono stati confrontati con un metodo GSA standard di letteratura basato sulla varianza, mostrando che metodo proposto e GSA standard concordano sull'identificazione delle tre variabili di input più importanti. Inoltre, il BMGSA ha dimostrato di essere affidabile anche quando applicato a dataset molto piccoli. L'applicazione al PSHA della tecnica sviluppata dimostra la sua capacità di classificare correttamente l'importanza del fattore di incertezza epistemica esistente, avendo bisogno solo dei dati di input e di output. Questo permette di applicare la tecnica a qualsiasi modello di rischio in cui l'incertezza epistemica deve essere valutata. La sua applicazione sistematica agli studi di rischio per individuare i parametri più influenti permetterebbe ai professionisti del rischio sia di migliorare i controlli di integrità durante la valutazione sia di concentrare la ricerca futura verso la riduzione dell'incertezza epistemica attraverso un'ulteriore caratterizzazione dei fattori più importanti.

In seguito, è stato presentato un nuovo approccio per ridurre il numero di scenari sismici da analizzare per SPTHA. L'approccio è un approccio euristico *wrapper-based* basato su un MODEA, che seleziona le feature degli scenari sismici da simulare. L'approccio proposto è stato applicato ad un caso di studio con riferimento alla stima del tasso annuale di superamento di una soglia di altezza $\tilde{\psi} = 1m$ delle onde di tsunami causate da terremoti crostali che potrebbero essere generate nella regione di Cefalonia-Lefkada nella Grecia nord-occidentale e propagate ad un sito target \bar{a} sulla costa orientale della Sicilia (Italia). L'approccio proposto si dimostra in grado di ridurre significativamente il numero di caratteristiche che descrivono la variabilità della sorgente sismica e, quindi, il numero di scenari da considerare nell'analisi, senza influenzare la precisione della stima del tasso annuale di superamento. È stata fornita un'interpretazione geofisica dei risultati.

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] 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.
- [5] R. K. Mcguire, “Probabilistic seismic hazard analysis: Early history,” *Earthq. Eng. Struct. Dyn. Earthq. Engng Struct. Dyn.*, vol. 37, pp. 329–338, 2008.
- [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. Molkenhain, 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).