

ChEeSE: L'Europa si prepara ad entrare nell'era exascale

“Oggi, il tuo cellulare ha una potenza di calcolo maggiore di quella della NASA nel 1969, quando mandò due astronauti sulla Luna. [...] La PlayStation di oggi, che costa 300 dollari, ha la potenza di calcolo di un supercomputer militare del 1997, il cui costo ammontava a milioni di dollari.” [Michio Kaku, fisico teorico]

Come dimostra Ray Kurzweil, informatico e saggista statunitense, il progresso tecnologico è un processo a crescita esponenziale. Se dovessimo descrivere la curva di crescita della capacità di calcolo dal primo computer elettronico programmabile (il *Colossus*, usato dagli Inglesi durante la seconda guerra mondiale per decifrare i messaggi dei nazisti) ai supercomputer odierni, non potremmo di certo usare una relazione lineare. Negli ultimi decenni abbiamo assistito ad una incredibile accelerazione tecnologica nell'ambito dell'**HPC (High Performance Computing, calcolo ad elevate prestazioni o supercalcolo)**, che ha permesso di raggiungere prestazioni di calcolo un tempo impensabili. Prendiamo la velocità di calcolo, che viene generalmente espressa in **FLOPS (FLoating Point Operations Per Second)**, ovvero il numero di operazioni (in virgola mobile) che una macchina può effettuare in un secondo: se il 2010 ha visto l'avvento del calcolo **petascale** (10^{15} FLOPS, cioè un milione di miliardi di FLOPs), il 2020 ha segnato l'inizio dell'era **exascale**, ovvero quella delle macchine in grado di raggiungere 10^{18} FLOPS (cioè un miliardo di miliardi di FLOPs, quindi mille volte di più). Gli sforzi tecnologici dietro questi incredibili numeri non risiedono solo nello sviluppo dell'hardware, ma anche in quello del software: un'infrastruttura di calcolo perfettamente in grado di sostenere prestazioni elevate è infatti inutile se gli algoritmi implementati sono inefficienti o, peggio ancora, incapaci di gestire le specifiche caratteristiche del calcolo HPC. Ma perché abbiamo bisogno del calcolo exascale? Per due ragioni fondamentali: 1) l'attuale capacità di calcolo non è sufficiente per risolvere determinati problemi scientifici; 2) i tempi di computazione attuali sono incompatibili con certe applicazioni specifiche. Nel seguito di questo articolo vedremo alcuni esempi delle sfide scientifiche per cui abbiamo bisogno di macchine exascale.

Il sistema HPC europeo

Mentre Stati Uniti e Cina si contendono lo scettro per il supercomputer più veloce al mondo, l'Italia ricopre, a marzo del 2021, una posizione di tutto rispetto con due supercomputers tra i primi 11 ([TOP500 List](#)): l'**HPC5** dell'ENI, all'8° posto in classifica (ma 6° se si guarda la [Green500 List](#), essendo uno dei più sostenibili al mondo), e **Marconi-100** del CINECA, all'11° posto. Dal canto suo, l'Unione Europea scende in campo con strategie comunitarie ambiziose e budget considerevoli, determinata ad avere un ruolo sempre più importante nello sviluppo del calcolo ad alte prestazioni.

Nel 2018 è stata istituita l'**EuroHPC, l'Impresa comune europea per il calcolo ad alte prestazioni**, che ha il duplice obiettivo di sviluppare un'infrastruttura di supercalcolo e di finanziare attività di ricerca e innovazione. Il budget destinato a questa impresa per il solo biennio 2019-2020 è pari a **1400 milioni di euro**, una somma ingente che testimonia l'interesse dell'Unione Europea in questa sfida tecnologica. L'infrastruttura di supercalcolo europea, che al momento include 7 sistemi HPC, verrà ampliata includendo 3 macchine petascale entro il 2021 (una di queste in Italia, al CINECA) e 2 macchine exascale entro il 2023 (la prima è prevista per il 2022). Infrastruttura e tecnologia non sono però i soli pilastri su cui si fonda questa scommessa europea: un terzo e fondamentale pilastro è quello

delle applicazioni, rappresentato da **14 Centri di Eccellenza Europei**, i veri protagonisti del progresso tecnologico e del consolidamento della leadership europea nel campo del supercalcolo.

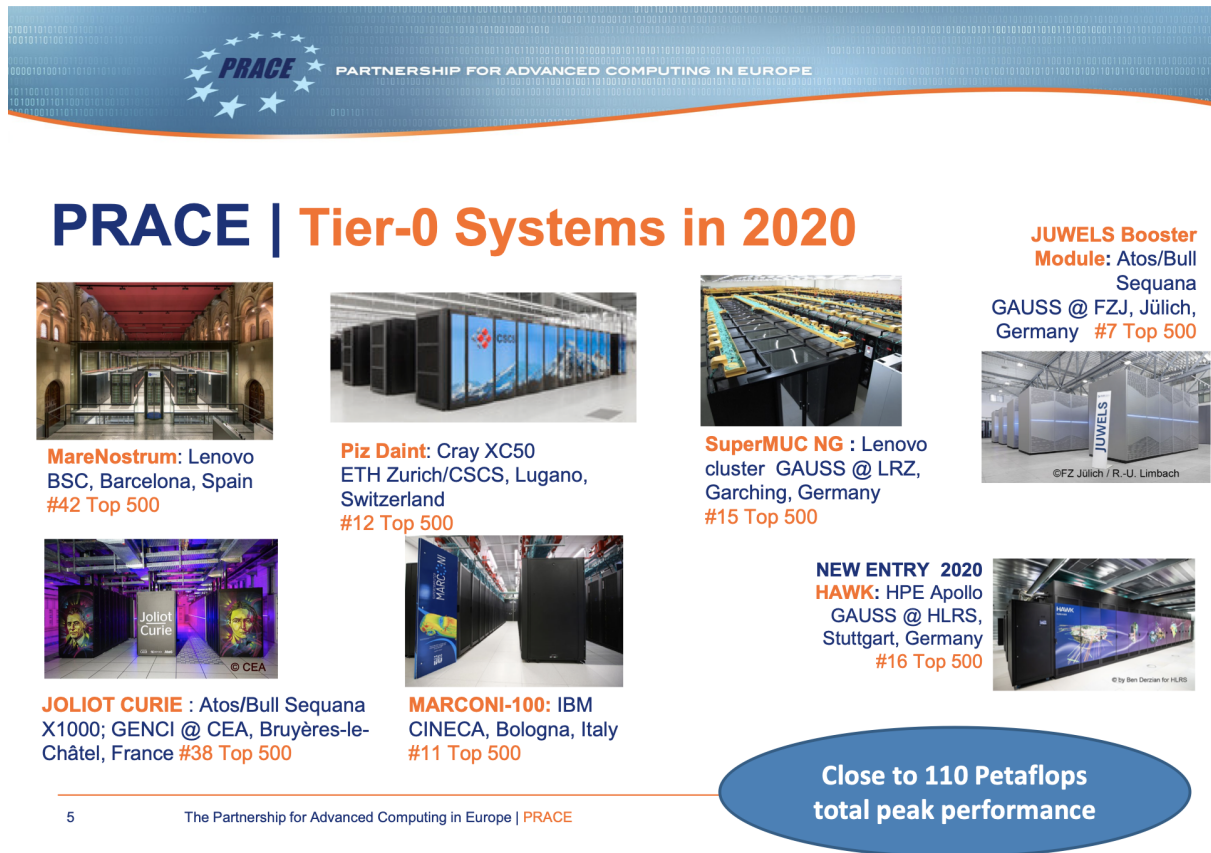


Figura 1: L'infrastruttura europea HPC allo stato attuale (prace-ri.eu)

ChEESE

ChEESE (<https://cheese-coe.eu/>) è un consorzio di 13 partners provenienti da 7 Paesi europei (l'Italia partecipa con INGV e CINECA). L'obiettivo di questo progetto europeo è quello di creare un **Centro di Eccellenza Europeo per il supercalcolo nell'ambito della Terra Solida**, in grado di recepire le nuove tecnologie petascale (2020) ed exascale (2022). Vietato arrivare impreparati, dunque.

Nello specifico, ChEESE si pone i seguenti obiettivi:

- Preparare 10 codici europei per l'architettura exascale emergente che permettano di riprodurre un determinato processo fisico attraverso delle simulazioni numeriche;
- Affrontare 15 sfide scientifiche della Terra Solida che necessitano di calcolo ad alte prestazioni;
- Fornire 12 pilot demonstrators (PD), ossia esempi pilota di applicazioni dei codici a problemi della Terra Solida, per dimostrare l'utilità del supercalcolo nella costruzione di servizi utili per la società (es. hazard sismico, o vulcanico, o da tsunami);
- Integrare le istituzioni, il mondo accademico, quello della ricerca e l'industria privata in un obiettivo comune.

Le 15 sfide scientifiche che ChEESE si propone di affrontare non sono solo molto ambiziose ma possono anche avere un grande impatto sulla società. Esse sono infatti relative allo studio di terremoti, tsunami, vulcani e campo geomagnetico e sono organizzate nei seguenti servizi: 1) **Urgent computing** (simulazioni in condizioni di emergenza, per cui è necessario ridurre drasticamente i tempi di computazione al fine di fornire servizi tempestivi); 2) **Stima probabilistica dell'hazard** (proiezione della pericolosità di un fenomeno in atto o possibile nel futuro a breve o a più lungo termine); 3) **Early warning** (allerta rapida); 4) **Tomografia sismica per la caratterizzazione del sottosuolo**.

Il ruolo dell'INGV

Nell'ambito del progetto ChEESE, l'INGV ricopre un ruolo di primo piano, con più di 20 ricercatori coinvolti nello sviluppo di 2 codici di simulazione numerica e 7 applicazioni a problemi della Terra Solida, nella creazione di un portale web che funga da repository per codici e datasets, nella definizione di strategie per l'implementazione dei servizi ChEESE e nell'inclusione di diversi prodotti, software e servizi in EPOS-ERIC (<https://www.epos-eu.org/>). Nel lungo termine, ci si aspetta che questi sviluppi possano migliorare significativamente i servizi di sorveglianza sismica e vulcanica, di monitoraggio e allerta tsunami (<http://www.ingv.it/cat/it/>), nonché le stime di pericolosità di breve e lungo termine che l'INGV fornisce alla Protezione Civile (si veda anche <https://cheese-coe.eu/media/news/research-societal-relevance-how-cheese-and-urgent-computing-may-enhance-ingvs-hazard>).

Vulcani

In ambito vulcanologico, l'INGV è leader di un PD per lo sviluppo del codice di simulazione **ASHEE** per la dinamica di eruzioni esplosive. L'ottimizzazione del codice di calcolo e dei *workflows* consentiranno di aumentare di un ordine di grandezza l'accuratezza spaziale delle simulazioni numeriche, permettendo in particolare una migliore descrizione fisica dei processi multifase ed una migliore analisi multiscala dei processi di dispersione turbolenta. D'altra parte, la diminuzione dei tempi di calcolo permetterà di utilizzare anche i modelli tridimensionali in modo probabilistico, utilizzando *ensembles* di simulazioni per la valutazione dell'incertezza nelle previsioni.

Per quanto riguarda la pericolosità vulcanica, l'INGV in ChEESE è leader di un PD che illustra l'utilità del calcolo exascale nei *workflows* per le stime di impatto delle ceneri vulcaniche (cadute al suolo o in sospensione in atmosfera) simulate con il codice **FALL3D**. L'esempio illustrativo è relativo ai vulcani Campi Flegrei a Napoli, e Jan Mayen, un'isola vulcanica che si trova nel Nord Atlantico. Le stime di pericolosità da caduta di ceneri vengono calcolate per il breve termine, ossia eseguendo qualche centinaio di simulazioni che considerano la previsione meteorologica del vento per i pochi giorni a seguire, e per il lungo termine (cioè per i prossimi decenni), eseguendo migliaia di simulazioni che tengono conto della statistica climatologica del vento. In entrambi i casi, il calcolo exascale permette di eseguire le simulazioni in un tempo ragionevole e, nel caso del breve termine, di eseguirle in un tempo utile affinché la stima di pericolosità possa essere utilizzata nella pratica per mitigare il rischio. Inoltre, il calcolo exascale permette di eseguire le simulazioni su un'area molto estesa, di migliaia di km, ad una risoluzione di circa 2 km (si veda ad esempio Figura 2). Questo si traduce nella capacità di simulare, e quindi di valutare, la pericolosità associata anche ad eventi molto rari ma di impatto altamente disastroso.

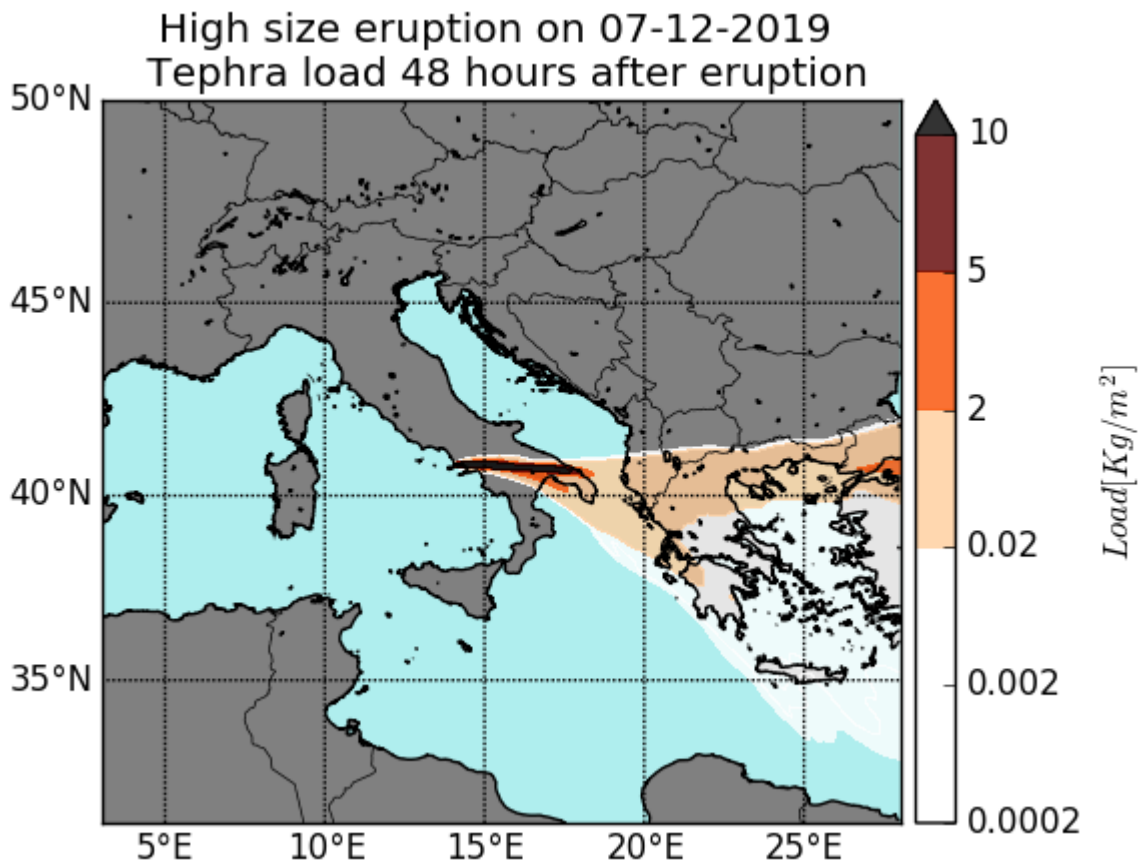


Figura 2: Esempio di stima di pericolosità di ceneri accumulate al suolo (con probabilità di superamento del 5%), dovuta ad un'eruzione altamente esplosiva ai Campi Flegrei, calcolata utilizzando le previsioni del vento del 7 dicembre 2019 per le 48 ore successive, su un dominio di 2000 km x 2000 km

Terremoti

I ricercatori dell'INGV sono impegnati nello sviluppo delle applicazioni ChEASE legate all'urgent computing in caso di forti terremoti ed alla tomografia sismica. Nel primo caso, l'obiettivo è quello di rendere fattibile, grazie alle risorse di calcolo exascale, la produzione di mappe di scuotimento ("shakemaps") molto accurate e basate su modelli fisici verosimili entro tempi molto brevi (ore) dal momento in cui si è verificato un evento sismico di media-grande magnitudo. Al momento, infatti, queste mappe vengono prodotte sulla base di relazioni empiriche che possono essere migliorate con le simulazioni numeriche per meglio rappresentare la complessa unicità del singolo evento sismico. Lavorare allo sviluppo delle applicazioni di urgent computing, assimilando anche dati in tempo reale, è di cruciale importanza, in quanto la stima tempestiva ed accurata della distribuzione attesa dello scuotimento rappresenta un supporto essenziale per le operazioni di soccorso. Uno strumento molto efficace legato alle *shakemaps* è rappresentato dagli *shakemovies*, che permettono di visualizzare in modo estremamente efficiente lo scuotimento atteso in aree via via più distanti dall'epicentro. La Figura 3 ne mostra un esempio, relativo al recente evento sismico in Croazia.

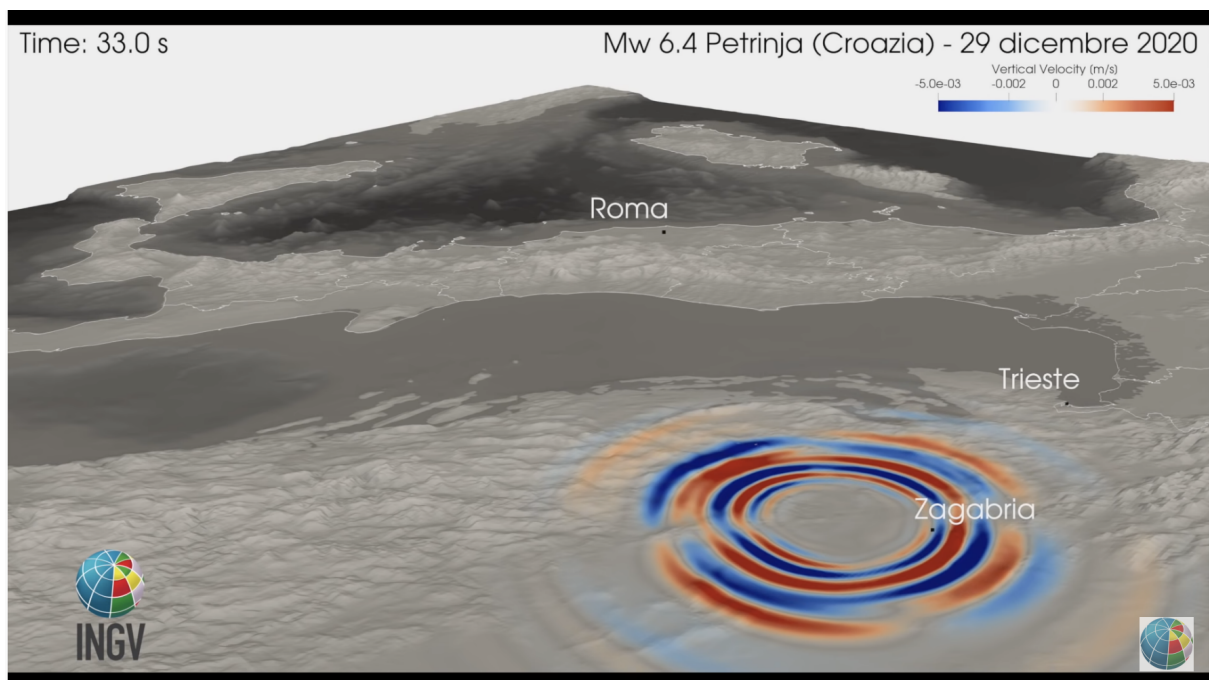


Figura 3: Estratto dal [video](#) di INGV terremoti che mostra la propagazione delle onde sismiche del terremoto in Croazia (Mw 6.4) del 29 dicembre 2020. Il colore blu indica movimenti del suolo verso il basso, mentre il colore rosso indica movimenti del suolo verso l'alto. La gradazione del colore rappresenta invece diversi valori di velocità verticale (m/s): tanto più il blu (rosso) è intenso, tanto più velocemente il suolo si muove verso il basso (l'alto).

Un aspetto molto importante che emerge da simulazioni come questa è la distribuzione disomogenea delle velocità del suolo: punti equidistanti dall'epicentro non subiscono lo stesso tipo di sollecitazione a causa di condizioni locali (topografia, tipo di suolo, ...) che possono influenzare in maniera significativa la propagazione delle onde sismiche. Questa è una delle ragioni principali per cui si rende necessario ottenere delle stime di scuotimento che siano basate su modelli fisici realistici.

Ed a tale scopo è legato il coinvolgimento dell'INGV nell'altra applicazione chiave legata allo studio dei terremoti, e cioè la tomografia sismica. Questa tecnica di **imaging** (ovvero di rappresentazione attraverso un'immagine, metodo molto simile alle radiografie o ecografie in ambito medico) permette di produrre modelli realistici e fisicamente plausibili del sottosuolo. Migliorare la qualità di tali modelli permetterebbe di ottenere delle migliori simulazioni di propagazione delle onde sismiche e, dunque, in un circolo virtuoso, delle stime ancora più accurate di scuotimento del suolo.

Tsunami

Per quanto riguarda lo studio degli tsunami nell'ambito del progetto ChESEE, l'INGV è fortemente coinvolto in due PD per la stima della pericolosità di breve termine - quindi immediatamente dopo l'occorrenza di un evento sismico - e di lungo termine.

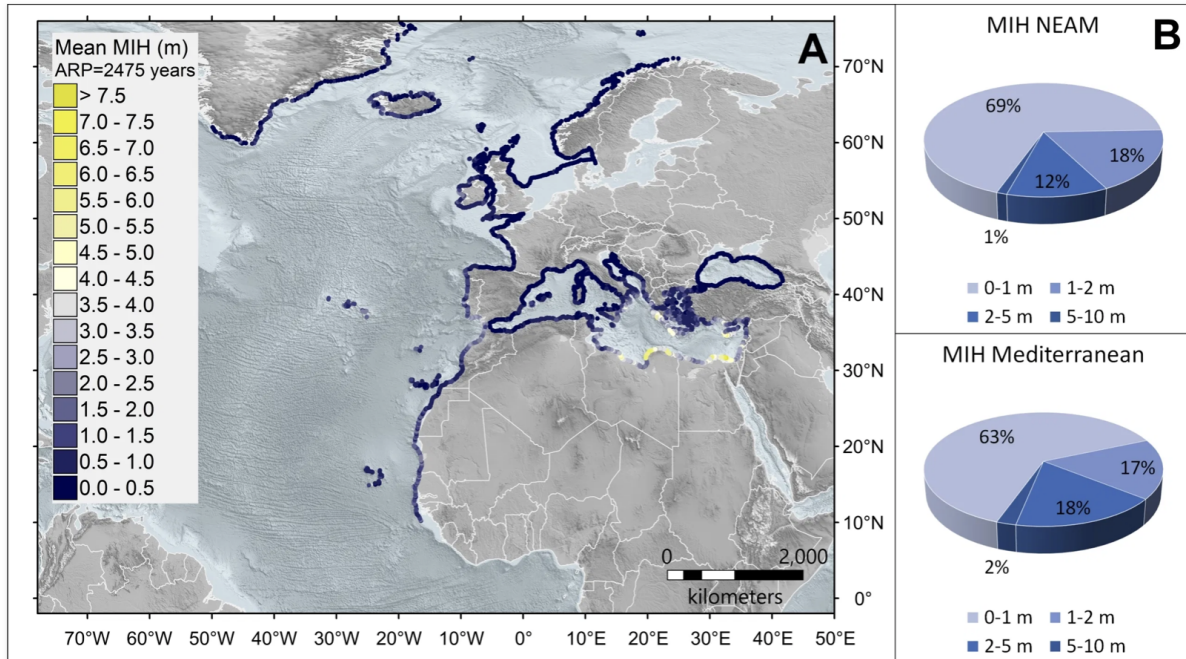


Figura 4: Uno degli outputs di NEAM Tsunami Hazard Model 2018 (NEAMTHM18), il modello di PTHA relativo all'area del Mediterraneo, Atlantico NE ed aree connesse. La mappa mostra la MIH ("Max Inundation Height") attesa per un tempo di ritorno medio di 2475 anni.

Il primo PD riguarda la **stima probabilistica della pericolosità da tsunami** (PTHA: Probabilistic Tsunami Hazard Assessment). Le metodologie PTHA permettono di valutare la probabilità che una data misura dell'intensità dello tsunami (ad esempio, l'altezza massima di run-up) venga superata in una particolare località in un determinato periodo di tempo. Dietro questa frase apparentemente semplice si nasconde un problema estremamente complesso: è necessario infatti combinare una quantità molto elevata di parametri e una varietà di dati di diversa natura per stimare come l'incertezza ad essi associata si propaga sui modelli di pericolosità. Difatti, la variabilità dei parametri di input, quali l'esistenza o meno di faglie sismogenetiche e la loro tipologia, la frequenza di terremoti di una certa magnitudo su una data faglia, le incertezze sulla batimetria e sulla topografia, determina una estrema variabilità negli scenari di tsunami plausibili e nella loro probabilità di accadimento. Al momento, le risorse di calcolo pongono dei limiti alla possibilità di esplorare l'incertezza totale e di modellare la propagazione degli tsunami su scala locale, per i quali la complessità del processo e la necessità di un'alta risoluzione richiedono di per sé tempi e risorse di calcolo estremamente elevati. Le macchine exascale permetterebbero dunque di superare tali limiti, migliorando così la nostra capacità di restituire scenari più verosimili e, al contempo, di quantificare l'incertezza totale.

Il secondo PD, coordinato da INGV, è invece relativo all'**early warning** ed al **post-disaster assessment** (stima rapida dell'impatto). Poiché la velocità di propagazione delle onde da tsunami è molto minore di quella delle onde sismiche, è infatti possibile utilizzare le registrazioni sismiche per stimare tempestivamente l'impatto dello tsunami atteso sulle coste. Per fare questo, tuttavia, è necessario tenere conto dell'incertezza riguardo la sorgente sismica che potrebbe aver innescato lo tsunami: infatti, nei minuti che seguono il verificarsi di un evento sismico, la conoscenza circa la sorgente sismica si basa solo su delle stime approssimate. Il calcolo exascale permetterebbe di esplorare in tempi molto ridotti un numero elevato di sorgenti compatibili con le prime osservazioni e di stimare meglio, in maggiore dettaglio e più rapidamente l'impatto atteso sulla costa, caratterizzando accuratamente l'incertezza.

Angela Stallone, Laura Sandri, Stefano Lorito, Manuela Volpe, Emanuele Casarotti, Tomaso Esposti
Ongaro, Lorenzo Cugliari.