Future Generation Computer Systems 146 (2023) 47–61

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

The EU Center of Excellence for Exascale in Solid Earth (ChEESE): Implementation, results, and roadmap for the second phase

Arnau Folch [a,*], Claudia Abril [b,a], Michael Afanasiev [m], Giorgio Amati [d], Michael Bader [f], Rosa M. Badia [a], Hafize B. Bayraktar [c,1], Sara Barsotti [b], Roberto Basili [c], Fabrizio Bernardi [c], Christian Boehm [m], Beatriz Brizuela [c], Federico Brogi [c], Eduardo Cabrera [a], Emanuele Casarotti [c], Manuel J. Castro [h], Matteo Cerminara [c], Antonella Cirella [c], Alexey Cheptsov [e], Javier Conejero [a], Antonio Costa [c], Marc de la Asunción [h], Josep de la Puente [a], Marco Djuric [e], Ravil Dorozhinskii [f], Gabriela Espinosa [a], Tomaso Esposti-Ongaro [c], Joan Farnós [a], Nathalie Favretto-Cristini [k], Andreas Fichtner [m], Alexandre Fournier [j], Alice-Agnes Gabriel [g], Jean-Matthieu Gallard [f], Steven J. Gibbons [i], Sylfest Glimsdal [i], José Manuel González-Vida [h], Jose Gracia [e], Rose Gregorio [a], Natalia Gutierrez [a], Benedikt Halldorsson [b,2], Okba Hamitou [l], Guillaume Houzeaux [a], Stephan Jaure [l], Mouloud Kessar [j], Lukas Krenz [f], Lion Krischer [m], Soline Laforet [l], Piero Lanucara [d], Bo Li [g], Maria Concetta Lorenzino [c], Stefano Lorito [c], Finn Løvholt [i], Giovanni Macedonio [c], Jorge Macías [h], Guillermo Marín [a], Beatriz Martínez Montesinos [c,3], Leonardo Mingari [a], Geneviève Moguilny [j,k], Vadim Montellier [k], Marisol Monterrubio-Velasco [a], Georges Emmanuel Moulard [l], Masaru Nagaso [k], Massimo Nazaria [c], Christoph Niethammer [e], Federica Pardini [c], Marta Pienkowska [m], Luca Pizzimenti [c,4], Natalia Poiata [j], Leonhard Rannabauer [f], Otilio Rojas [a], Juan Esteban Rodriguez [a], Fabrizio Romano [c], Oleksandr Rudyy [e], Vittorio Ruggiero [d,5], Philipp Samfass [f], Carlos Sánchez-Linares [h], Sabrina Sanchez [j], Laura Sandri [c], Antonio Scala [c,1], Nathanael Schaeffer [k], Joseph Schuchart [e,6], Jacopo Selva [c,7], Amadine Sergeant [k], Angela Stallone [c], Matteo Taroni [c], Soelvi Thrastarson [m], Manuel Titos [b], Nadia Tonelllo [a], Roberto Tonini [c], Thomas Ulrich [g], Jean-Pierre Vilotte [j], Malte Vöge [i], Manuela Volpe [c], Sara Aniko Wirp [g], Uwe Wössner [e]

[a] Barcelona Supercomputing Center (BSC), Spain
[b] Icelandic Meteorological Office (IMO), Iceland
[c] Istituto Nazionale di Geofisica e Vulcanologia (INGV), Italy
[d] CINECA, Italy
[e] High Performance Computing Center Stuttgart (HLRS), Germany
[f] Technical University of Munich (TUM), Germany
[g] Ludwig Maximilian University (LMU), Germany
[h] University of Malaga (UMA), Spain
[i] Norwegian Geotechnical Institute (NGI), Norway
[j] Universite Paris Cite, Institut de Physique du Globe de Paris (IPGP), France
[k] National Centre for Scientific Research (CNRS), France
[l] ATOS, France
[m] ETH Zurich, Switzerland

* Correspondence to: Geociencias Barcelona (GEO3BCN-CSIC), Spain.
E-mail address: afolch@geo3bcn.csic.es (A. Folch).

1 Also at Università degli Studi di Napoli Federico II, Napoli, Italy.
2 Also at University of Iceland, Iceland.
3 Now at Earthquake Research Institute, University of Tokyo, Japan.
4 Now at SOFT STRATEGY SpA, Italy.
5 Now at LV srl, Italy.
6 Now at University of Tennessee Knoxville (UTK), USA.
7 Now at Università degli Studi di Napoli Federico II, Naples, Italy.

https://doi.org/10.1016/j.future.2023.04.006

A R T I C L E   I N F O

Article history:
Received 5 October 2022
Received in revised form 27 March 2023
Accepted 1 April 2023
Available online 5 April 2023

Keywords:
Center of Excellence (CoE)
EuroHPC
Exascale transition
Code scalability
Geophysics
Natural hazards
HPC service enabling
Urgent computing
Early warning forecast

A B S T R A C T

The EU Center of Excellence for Exascale in Solid Earth (ChEESE) develops exascale transition capabilities in the domain of Solid Earth, an area of geophysics rich in computational challenges embracing different approaches to exascale (capability, capacity, and urgent computing). The first implementation phase of the project (ChEESE-1P; 2018–2022) addressed scientific and technical computational challenges in seismology, tsunami science, volcanology, and magnetohydrodynamics, in order to understand the phenomena, anticipate the impact of natural disasters, and contribute to risk management. The project initiated the optimisation of 10 community flagship codes for the upcoming exascale systems and implemented 12 Pilot Demonstrators that combine the flagship codes with dedicated workflows in order to address the underlying capability and capacity computational challenges. Pilot Demonstrators reaching more mature Technology Readiness Levels (TRLs) were further enabled in operational service environments on critical aspects of geohazards such as long-term and short-term probabilistic hazard assessment, urgent computing, and early warning and probabilistic forecasting. Partnership and service co-design with members of the project Industry and User Board (IUB) leveraged the uptake of results across multiple research institutions, academia, industry, and public governance bodies (e.g. civil protection agencies). This article summarises the implementation strategy and the results from ChEESE-1P, outlining also the underpinning concepts and the roadmap for the on-going second project implementation phase (ChEESE-2P; 2023–2026).

1. Introduction

In May 2022, High Performance Computing (HPC) finally entered the exascale era after the deployment at the U.S. Oak Ridge National Laboratory (ORNL) of the Frontier supercomputer, a Cray Inc. system powered with 9,408 nodes and 37,632 AMD MI250X GPUs scoring 1.102 High Performance Linpack (HPL) exaflop/s. Following Frontier's footsteps, similar world-class computational infrastructures are underway in several countries according to their strategic HPC roadmaps. In the European Union (EU), the EuroHPC Joint Undertaking (JU) has already deployed LUMI and Leonardo (428 and 255 HPL peak petaflop/s; third and forth in the June 2022 top-500 rank respectively) and initiated the deployment of MareNostrum-5, with an expected HPL peak performance of 314 petaflop/s. These three pre-exascale systems are part of a larger European investment that also includes subsequent procurements for two exascale computers and the setup of a number of mid-range petascale systems in several EU member states to complement the extreme-scale European ecosystem. In parallel and over the last years, the EU has made a substantial investment to develop exascale transition capabilities both in terms of hardware (e.g. the Open-RISC-V Strategic Agenda) and of open software stack technology, including the establishment of the Centers of Excellence (CoEs), to advance in lighthouse applications, and of the National Competence Centres (NCCs), to leverage the uptake of HPC codes and services by Small and Medium-sized Enterprises (SMEs), industry, and academia. Several research and innovation actions, originally under the EU Horizon 2020 funding programme (calls EINFRA-2015, INFRAEDI-2018, INFRAEDI-2019) and afterwards under the umbrella of Horizon Europe and the JU (call EUROHPC-JU-2021-COE) have supported a number of CoEs to promote the transition towards exascale in target scientific domains. In particular, the first implementation phase of the Center of Excellence for Exascale in Solid Earth (ChEESE-1P) targeted the preparation of community flagship codes and services in the domain of Solid Earth.

Solid Earth (SE) science is rich in computational challenges that embrace the different approaches to exascale, requiring large-scale computational infrastructures both to address fundamental scientific questions and to manage the occurrence of geohazards and their impacts. On the one hand, many SE problems involve multiple spatial scales spanning orders of magnitude (from local to global), time ranges, multiphysics with complex couplings, and substantial data volumes, requiring large monolithic executions that have been unaffordable by petascale systems in terms of capability. Exascale capability challenges in SE include, for example, high-frequency seismic full-waveform inversion and related high-resolution Earth's sub-surface imaging, modelling of the Earth's convective dynamo, or multiphase and multicomponent full-scale simulation of volcanic plumes. On the other hand, SE problems entail uncertainties that are handled using large ensembles of model realisations, e.g. in ensemble-based data assimilation schemes, to solve inverse problems, or in probabilistic long-term and short-term hazard assessments that combine many scenarios spanning the uncertainty ranges. All these problems make use of complex workflows to orchestrate hundreds to millions of smaller-scale simulations that can potentially aggregate into an overall exascale capacity workload. Depending on the degree of coupling between single model realisations, capacity workloads can range from tightly coupled, as in the case of ensemble-based data assimilation cycles (e.g. methods based on Kalman filters and derivates), to loosely coupled as in realisations of probabilistic hazard scenarios. Finally, geohazardous phenomena such as earthquakes, tsunamis or volcanic eruptions require emergency simulations that are affordable with current systems (in terms of capability) but that need to be solved under strict time constraints. In these urgent computing applications, exascale resources can reduce substantially the time to solution and hence have the potential to substantially transform the way society deals with early warning, short-term forecast, and emergency management. In the transition to exascale, scientists are becoming able to combine capability computing (higher resolution, increased degree of model sophistication, multiscale, multiphysics) and capacity computing (multiple scenarios to deal with uncertainties, data assimilation, fusion of models and observations and hybrids thereof) into emerging products and services delivering to public authorities and decision-makers high-fidelity predictions under strict time constraints, thereby helping to save lives and livelihood.

This manuscript reviews the implementation strategy and the results from the first phase of the ChEESE Center of Excellence (ChEESE-1P), which ran from November 2018 to March 2022.

Table 1
Members of the ChEESE-1P consortium grouped by type of institution and country. ETH includes the Swiss National Supercomputing Centre (CSCS) as an autonomous unit.

Type of institution	Partner	Acronym	Country
Supercomputing Centres	Barcelona Supercomputing Center	BSC	Spain
	High Performance Computing Center Stuttgart	HLRS	Germany
	CINECA	CIN	Italy
Research and Monitoring	Istituto Nazionale di Geofisica e Vulcanologia	INGV	Italy
	Icelandic Meteorological Office	IMO	Iceland
	Institut de Physique du Globe de Paris	IPGP	France
	Centre National de la Recherche Scientifique	CNRS	France
Research and Academia	Swiss Federal Institute of Technology	ETH	Switzerland
	Technical University of Munich	TUM	Germany
	Ludwig-Maximilians University	LMU	Germany
	University of Malaga	UMA	Spain
	Norwegian Geotechnical Institute	NGI	Norway
Private Company	Bull-ATOS	BA	France

The review covers the exascale preparation strategies adopted for the different flagship codes and related workflows, the Pilot Demonstrators, and the final enabling and validation of HPC services in cooperation with end users and stakeholders. The paper outlines also the underpinning concepts and the roadmap for the second project implementation phase (ChEESE-2P) that kicked-off in January 2023 and will run for 4 years (2023–2026). Note that, for the sake of space, only a succinct overview of the ChEESE-1P results can be included here. For further details the reader is pointed to more than 60 publications that resulted from the project, most of which are cited throughout the text.

2. Overview of ChEESE-1P

The first implementation phase of the Center of Excellence for Exascale in Solid Earth (ChEESE-1P), a consortium of 13 partners from 7 different European countries (see Table 1), addressed scientific, technical, and socio-economic computational challenges in the domains of computational seismology, tsunami science, volcanology, and magnetohydrodynamics. ChEESE-1P was articulated around the fundamental concepts of flagship applications, Pilot Demonstrators (PDs), service enabling, and community building. The applications pillar conducted a number of transversal activities, including co-design, aimed at preparing 10 community flagship codes for the present and upcoming world-class supercomputing systems (Section 3). In parallel, the project implemented 12 PDs, intended as proofs of concept in which flagship codes and workflows were combined to: (i) address underpinning capability and capacity computational challenges and, (ii) test the optimised versions of the codes on petascale machines and exascale hardware prototypes available in the interim (Section 4). In addition, the 8 PDs that reached a Technology Readiness Level (TRL) of 5 or above during the execution of the project (i.e. underlying technology successfully demonstrated in relevant environments) were enabled in operational service environments on critical aspects of geohazards like probabilistic hazard assessment, urgent computing, and early warning and probabilistic forecast (Section 5). The definition of HPC service functional requirements, testing, and validation of project assets, was done in partnership with members of the ChEESE-1P Industry and User Board (IUB), which also collaborated on a market analysis for future service exploitation. The implementation and execution of the PDs and related service prototypes on the European petascale systems available during the first phase of the project (see Table 2) was possible thanks to a combination of consortium in-kind resources, national-level competitive projects, and the award of 4 Partnership for Advanced Computing in Europe (PRACE) project access, all together adding up a remarkable total of 217M of core hours for the various project activities. Finally, in terms of community building (Section 6), ChEESE-1P leveraged the use of HPC across multiple research institutions, academia, industry, and public governance bodies (e.g. civil protection agencies). This included numerous courses and training activities in order to engage interested parties in the usage and uptake of the ChEESE-1P results.

3. Flagship codes

ChEESE-1P prepared 10 community flagship codes that are of wide use across different areas of SE including 4 codes in computational seismology, 2 codes in magnetohydrodynamics (MHD), 2 codes in volcanology, and 2 codes in tsunami modelling (see Table 3 for code list, description, and references). At the beginning of the project, the 10 flagship codes were labelled in terms of (strong) scalability as:

Scalability level 1: petascale codes, good strong scalability (>90% parallel efficiency) proven up to about 10^4–10^5 cores or above (for hybrid node architectures, core numbers can be converted to the equivalent node number depending on the case). Four codes were already at petascale level at the beginning of the project (November 2018): Salvus, SPECFEM3D, SeisSol, and xSHELLS.

Scalability level 2: pre-petascale codes, with good strong scalability (>90% parallel efficiency) proven up to 10^3–10^4 cores or node equivalent. Two codes were initially at level 2, PARODY_PDAF and ExaHyPE.

Scalability level 3: codes with strong scalability bottlenecks occurring at around 10^3 cores or node equivalent. At the start of ChEESE-1P, 4 codes were at the lower scalability level of 3: ASHEE, T-HySEA, L-HySEA, and FALL3D.

As noted, both the baseline performances and the scalability levels of the flagship codes were very heterogeneous, reflecting the fact that some areas of SE have a stronger background in HPC than others (e.g. seismology versus volcanology). With this premise in mind, a number of transversal activities were defined to identify and solve the bottlenecks that partly or completely affected the performances of the codes and to act on co-design, understood as a set of software, tools and techniques for exploring key issues associated to future exascale hardware design. The long-term objective was (and still is) to bring codes and underlying workflows to a future strong scalability level 0, which actually means parallel efficiency on tier-0 exascale machines (i.e. good scalability up to about 10^6 cores or about 10^4 accelerated hybrid nodes).

3.1. Exascale code preparation strategy

The code preparation strategy entailed two iterations of code audits and related audit-driven optimisation activities. At the

Table 2
List of European petascale systems used in the various ChEESE-1P activities (code audits and optimisations, implementation of PDs, large-scale testbeds, and service production runs). Acronyms for centers are as follow: CIN (CINECA), LRZ (Leibniz Supercomputing Centre), CSCS (Swiss National Supercomputing Centre), JSC (Jülich Supercomputing Centre), CNRS (Centre National de la Recherche Scientifique), TGCC-CEA (Très Grand Centre de calcul du CEA), BSC (Barcelona Supercomputing Center).

System	Center	Peak performance (HPL PFlop/s)	Number of (thin) nodes	CPUs (per node)	GPUs (per hybrid node)
Marconi-100	CIN	32.0	980	32 IBM Power-9	4 NVIDIA V100
SuperMUC	LRZ	26.9	6,336	48 Intel Xeon Platinum	none
Piz Daint	CSCS	25.3	5,704	12 Intel Xeon E5–2690	1 NVIDIA P100
JURECA	JSC	18.5	768	128 AMD Rome (EPYC)	4 NVIDIA A100
Jean Zay	CNRS	15.9	2,140	40 Intel Cascade Lake	4 NVIDIA V100
Joliot-Curie (Irene-Rome)	TGCC-CEA	11.7	2,292	128 AMD Rome (EPYC)	none
MareNostrum-4	BSC	11.1	3,456	48 Intel Xeon Platinum	none
Joliot-Curie (Irene-SKL)	TGCC-CEA	9.0	1,656	48 Intel Skylake	none
CTE-Power	BSC	1.5	52	40 IBM Power-9	4 NVIDIA V100
Nord-3	BSC	0.25	756	16 Intel SandyBridge	none

Table 3
List of flagship codes in ChEESE-1P including code reference, code area, initial and final code scalability level (see text for details), and a short code description. Code areas are as follow: CS (Computational Seismology), MHD (magnetohydrodynamics), V (volcanology), T (tsunami modelling).

Code (reference)	Area	Level (start)	Level (end)	Short code description
Salvus [1]	CS	1	1	Solves dynamic (visco)acoustic and elastic wave propagation on fully unstructured hyper-cubic and simplicial meshes using a spectral-element approach. Designed for large-scale, time-domain, full-waveform modelling and inversion ranging from laboratory ultrasound studies to planetary-scale seismology.
SPECFEM3D [2]	CS	1	1	Solves acoustic (fluid), elastic (solid), coupled acoustic/elastic-poroelastic or seismic wave propagation in any type of conforming mesh. It can also be used for ocean acoustics (doi:10.5281/zenodo.6394675).
SeisSol [3]	CS	1	1	Solves seismic wave propagation (elastic, viscoelastic) and dynamic rupture problems on heterogeneous 3D models using high-order Discontinuous Galerkin (DG) discretisation and local time-stepping on unstructured adaptive tetrahedral meshes (doi:10.5281/zenodo.7598601).
xSHELLS [4]	MHD	1	1	Solves rotating incompressible flows and magnetic fields in spherical shells including the coupled induction equation for MHD and temperature. Based on a semi-spectral approach combining finite differences in radius and spherical harmonics, semi-implicit second-order time scheme.
PARODY_PDAF [5]	MHD	2	1	Solves the Navier–Stokes and Maxwell equations in the MHD approximation in spherical shell geometry, offering the possibility of ensemble simulations using the Parallel Data Assimilation Framework (PDAF) library (doi:10.5281/zenodo.6415313).
ExaHyPE [6]	CS	2	2	Solves systems of hyperbolic PDEs including seismic wave propagation problems using high-order DG discretisation with local time-stepping on tree-structured adaptive Cartesian meshes (doi:10.5281/zenodo.6387677).
ASHEE [7]	V	3	2	Solves multiphase fluid dynamic model for compressible volcanic plume mixtures composed of gaseous components and solid particle phases (doi:10.5281/zenodo.6560777).
T-HySEA [8] L-HySEA [9]	T	3	2	T-HySEA solves propagation of tsunamis generated by earthquakes including seafloor deformation, propagation and inundation of coastal areas under the shallow water approach in spherical coordinates and using high-order Finite Volume discretisation. Similarly, L-HySEA solves the Savage–Hutter equations for landslide triggered tsunamis in a two layer shallow water system where the landslide represents the bottom layer and the tsunami the top layer (doi:10.5281/zenodo.6400825).
FALL3D [10]	V	3	2	Solves set of advection-diffusion-sedimentation (ADS) equations for the atmospheric transport and ground deposition of particles and aerosols using a Finite Volume explicit scheme (doi:10.5281/zenodo.7267808).

beginning of the project, all codes were audited in collaboration with the EU Performance Optimisation and Productivity Centre of Excellence in HPC (POP CoE). Level 1 codes were evaluated with the Score-P profiling tool [11], which provides a deep insight in particular performance aspects such as large scale executions, whereas the evaluation of level 2 and level 3 codes relied on Extrae, a simpler profiling package that generates Paraver trace-files for analyses and that is flexible enough to highlight potential code performance bottlenecks [12]. Hints from the code audits were then mapped into corresponding single-node and multi-node optimisation activities as shown in Table 4. A second audit round served to better plan the optimisation tasks until the end of the project and to compare the resulting POP metrics and the test cases (code runtimes) with the first audit values, thereby providing a quantitative and objective interim monitoring of the code optimisation activities. Overall, the largest effort focussed on code acceleration using CUDA and OpenACC (including efficiency of CPU–GPU and GPU–GPU data transfer for most of the ported applications), followed by parallel and asynchronous I/O. In fact, at the end of ChEESE-1P, 8 out of 10 flagship codes had been fully or partially ported to GPU environments, and all codes initially labelled at scalability level 3 were already at level 2. Details on code optimisations and tunings can be summarised as follows:

1. **Salvus**. Salvus uses the Portable, Extensible Toolkit for Scientific Computation (PETSc, [13]) DMPLEX class for unstructured mesh management. The first POP audit revealed that the serial mesh loading and interpolation (originally done on a single node), imposed memory limits on the maximum mesh size that Salvus could handle. Accordingly, optimisation activities on Salvus included the parallelisation of DMPLEX and the enabling of parallel

Table 4
Audit-driven optimisation activities performed on the different ChEESE-1P codes: (1) Salvus, (2) SPECFEM3D, (3) SeisSol, (4) xSHELLS, (5) PARODY_PDAF, (6) ExaHyPE, (7) ASHEE, (8) T-HySEA, (9) L-HySEA, (10) FALL3D.

Activity/Code number	1	2	3	4	5	6	7	8	9	10	Total
Porting to accelerators and GPU optimisation support	–	–	✓	✓	✓	✓	✓	✓	✓	✓	8
Parallel and asynchronous I/O improvements	✓	✓	–	–	✓	–	✓	✓	✓	✓	7
Algorithmic optimisation	–	–	–	–	–	✓	✓	✓	✓	✓	5
Code vectorisation	–	✓	–	–	–	✓	–	–	–	✓	3
OpenMP parallelisation	–	–	–	✓	✓	–	–	–	–	✓	3
Load balance	–	–	✓	–	–	✓	–	✓	✓	–	4
Communications efficiency	–	✓	–	–	–	–	✓	–	–	–	2
Memory usage	–	–	–	✓	–	–	–	–	–	–	1

HDF5 I/O, resulting in a drastic reduction of the code start-up time and the extension to much larger meshes [14–16] (see Fig. 1a). Improvements added to PETSc DMPLEX class were tested up to 214 billion elements (7th order, 512 grid points per element) on 3,000 nodes of MareNostrum-4 supercomputer and were distributed to the community in the PETSc 3.13 version release.

2. **SPECFEM3D**. SPECFEM3D legacy performance metrics were already very good (Fig. 1b), but the code produced several files per MPI rank leading to bottlenecks when the number of processes increased. The first POP audit also revealed that vectorisation over elements instead of vectorisation inside elements limited the code vectorisation capabilities. Dedicated optimisation activities focussed on vectorisation on CPU (with a speedup gain up to 3x) and on implementing parallel HDF5 in order to reduce the number of output files, including asynchronous I/O to overlap computation with IOs, leading to a 2x speedup gain above 5,120 processors (Fig. 2a). Finally, CUDA kernels were ported to Heterogeneous Interface for Portability (HIP) for providing also AMD GPU support.

3. **SeisSol**. The SeisSol legacy version was parallelised using MPI plus OpenMP and the underlying Yet Another Tensor Toolbox (YATeTo, [17]) for Discontinuous Galerkin (DG) methods, which did not have GPU support. During ChEESE-1P, YATeTo was extended for use with GPUs, including a change in the parallelisation mode to launch batches of matrix–matrix multiplications for operations on GPUs [18]. On the other hand, the local time stepping was identified as a main challenge that was addressed by means of CUDA graphs and improved mesh partitioning for GPUs. The fused kernels lead to a GPU overall performance increase of 2x (Fig. 2b), whereas data transfer optimisation for computing friction mechanics improved it by a factor of 3x.

4. **xSHELLS**. The first audit showed that the xSHELLS scalability was limited by data-dependencies between processes in the sparse linear solver, leading to latency increasing with the number of processes. In addition, xSHELLS uses the Spherical Harmonic Transform library (SHTns, [19]) that was not fully efficient on GPUs. Tailored optimisation activities for xSHELLS focused on the use of MPI-3 shared memory within nodes to improve parallelism and reduce explicit communication overhead and the CUDA part of the SHTns library. These resulted in a much better scaling at large node counts, the code being more than 2x faster on the 128 nodes tests, now used in production [20,21]. The new GPU mode is 4x to 8x faster on a node-to-node comparison on Marconi-100 machine (hybrid nodes with 4 NVIDIA V100 GPUs).

5. **PARODY_PDAF**. Scaling in PARODY_PDAF [22,23], a code originally at scalability level 2, was largely constrained by the limited extension of the OpenMP regions and the sequential I/O. Audit-driven optimisation activities focussed on improving parallel efficiency by extending the OpenMP regions, enabling parallel I/O using MPI-IO, and a preliminary porting to GPUs using OpenACC and interfacing with the CUDA version of the SHTns library (that is shared with xSHELLS). All these activities resulted on a 1.75x speedup gain on Joliot-Curie machine tests using 18,432 cores, already bringing the code to a scalability level close to 1 (Fig. 1c).

6. **ExaHyPE**. ExaHyPE uses the Peano framework [24] for solvers operating on dynamically adaptive Cartesian meshes, which has a tree-oriented load balancing (coarse granular). The first audit suggested that the tree traversal at MPI-level and the lack of code vectorisation were critical. Accordingly, dedicated activities for ExaHyPE focussed on increasing the percentage of vectorised code to nearly 100% and on starting the rewrite of Peano (4th generation) with a much improved parallelisation approach including support for GPUs [25–27]. On the other hand, a reactive load balancing and code resilience strategy was implemented by means of task sharing, including reactive task offloading and replication-based resilience (i.e. task sharing between replicated ranks to mitigate performance pain and detection of silent data corruption via error indicators), which yielded to a 2x performance gain on several tests [28,29].

7. **ASHEE**. The ASHEE code [30], based on the OpenFOAM toolkit, presented I/O issues on a few thousands of cores and memory bandwidth and MPI communications problems yielding to a poor parallel efficiency (the initial code level was of 3). The dedicated optimisation activities focussed on implementing asynchronous I/O and on porting to GPUs the linear algebra parts using external libraries (*i.e.* no need to change the underlying OpenFOAM code). On the other hand, the memory usage and the intra-node scaling were improved with the use of mixed precision, yielding to a 2x faster matrix assembly (in single precision) and to a 3x gain in the linear algebra part (in double precision) using GPU offloading. Several scaling tests with mixed precision gave good parallel efficiency up to 12,000 cores, meaning the code is currently at level 2.

8/9. **T-HySEA and L-HySEA**. The first POP audit of the code(s), already fully ported to CUDA at the start of the project, revealed a poor load balancing, a lack of asynchronous transfers between CPU and GPU memories, and a non-optimal netCDF file output. Audit-driven activities for T/L-HySEA focussed on implementing asynchronous transfers between CPU and GPU memories and direct GPU-to-GPU memory transfers, as well as compressing the netCDF files in single precision. These yielded to an overall improvement of 30% in code speedup and a decrease of the total time to solution by about 20% [31] (Fig. 2c).

10. **FALL3D**. FALL3D was originally a level 3 code that presented I/O limitations and memory bottlenecks leading to an overall lack of scalability. During ChEESE-1P, the code was heavily refactored with improved communication/computation ratios, better data management, introduction of netCDF parallel I/O and, finally, fully ported to GPUs using OpenACC [32,33]. The overall speed-up increased by 5x (90% parallel efficiency on 64 Joliot-Curie nodes; code already at level 2) and the runtime was also reduced drastically, with the accelerated version more than 10x faster compared with the original plain MPI version on the same number of nodes (tests on Marconi-100 and CTE-Power systems, Fig. 2d).

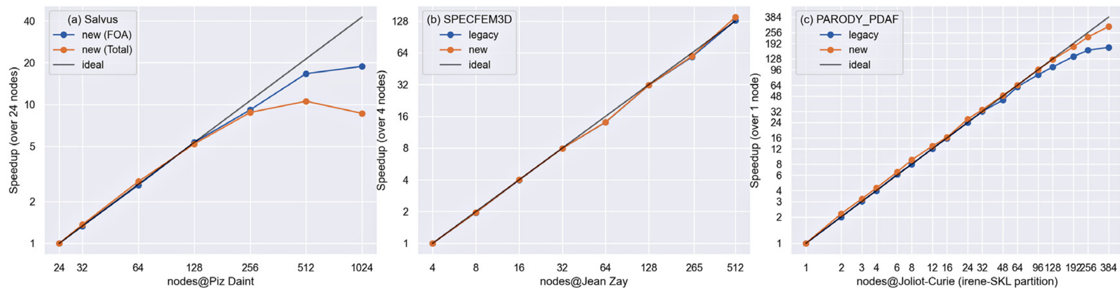

Fig. 1. Strong scaling examples for some level 1 codes on different architectures. (a) Salvus at Piz Daint up to 1024 nodes (*i.e.* 1024 NVIDIA P100 GPUs). Results are for a mesh containing 256M 4th-order spectral elements showing the second POP audit focus-of-analysis (FOA) and the total execution time of the optimised (new) code version. Note that legacy results are not shown for this test because, before the improvements in PETSs, the largest Salvus mesh size was limited by the memory of a single node (around 16M elements for Piz Daint). Almost perfect strong scaling is observed until 125k elements per GPU, below which the time spent in MPI communication takes over. (b) SPECFEM3D at Jean Zay up to 512 nodes (*i.e.* 20,480 cores). Both legacy and new versions show a very good scaling, the new version being faster by about 1% on average. (c) PARODY_PDAF at Joliot-Curie (Irene-SKL partition) up to 384 nodes (*i.e.* 18,432 cores). At large node count, the new code version exhibits performance gain of about 1.75x.

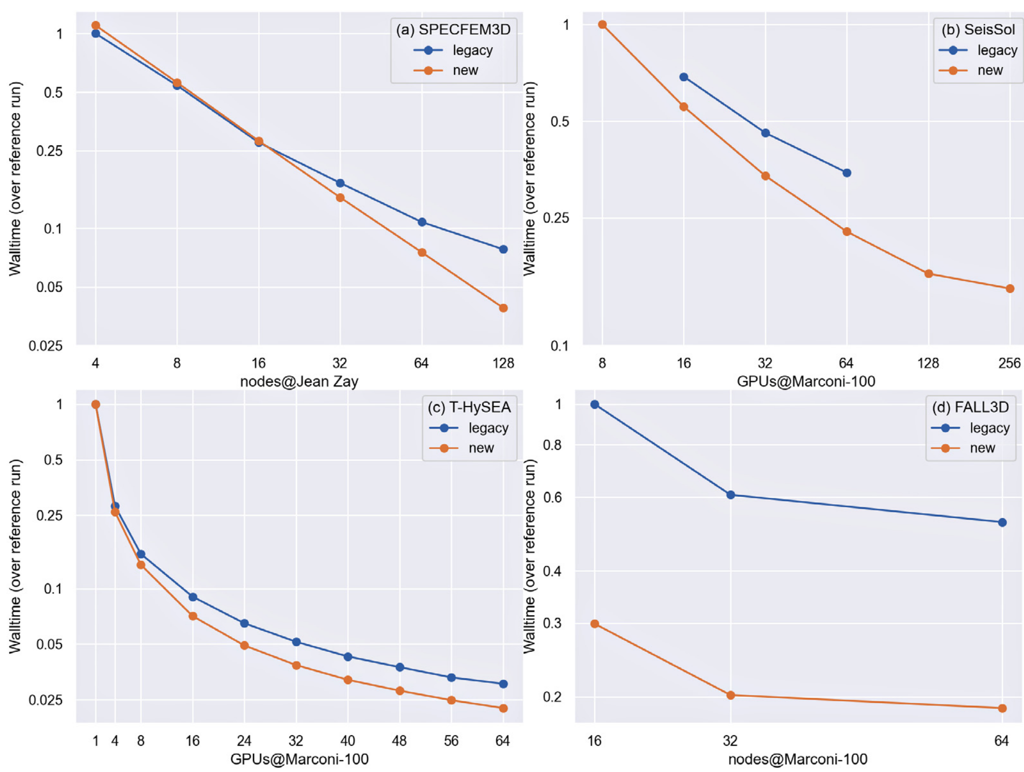

Fig. 2. Examples illustrating code time to solution improvements after audit-driven optimisations on smaller-scale tests. (a) SPECFEM3D at Jean Zay up to 128 nodes (*i.e.* 5,120 cores). The plot compares the legacy I/O with the new asynchronous HDF5 I/O strategy overlapping computation with writing for a 13M elements mesh and a simulation producing an output file size of 0.5Tb. (b) SeisSol at Marconi-100 up to 256 NVIDIA V100 GPUs (*i.e.* 64 nodes). This example, on a 10M elements mesh, illustrates the overall performance improvement of SeisSol-GPU after improved mesh partitioning for GPUs and local time stepping. (c) T-HySEA at Marconi-100 up to 64 NVIDIA V100 GPUs (*i.e.* 16 nodes). The example corresponds to a 84M cells domain and illustrates the gain of asynchronous transfers between CPU and GPU memories. (d) FALL3D at Marconi-100 up to 64 nodes for a 18M cells case showing the impact of GPUs using OpenACC and CUDA-Aware MPI to directly communicate all GPU devices.

3.2. Co-design activities

ChEESE-1P entailed several co-design activities aimed at exposing the specificities and the dependencies of the mini-apps (*i.e.* reduced versions of the applications that test more effectively numerical kernels) and full applications on different hardware architectures in terms of performance, scalability, and power consumption. Target hardware technologies were selected considering that exascale systems will be heterogeneous with processors based on Intel/AMD x86 and ARM processor architectures having high-levels of core parallelism, e.g. the Advanced Vector Extensions (AVX) AVX-512 for x86 and the Scalar Vector Extensions (SVE) for ARM, and with hybrid nodes equipped with accelerators, e.g. NVIDIA and AMD GPUs. With this premise, the co-design tasks followed a 3-level methodology as follows:

Co-design level A used mini-app proxies inherited from the previous Mantevo project [34], which embed numerical kernels contained in stand-alone applications. The level A step allowed co-design activities to start quicker, while the ChEESE mini-apps were still under development, and included mini-FE (as a proxy for unstructured implicit finite element codes), mini-AERO (proxy for explicit unstructured finite volume codes solving the compressible Navier–Stokes equations), and Bench-IO (proxy for benchmarking the performance of parallel file systems).

Co-design level B implemented mini-apps for 4 flagship codes; mini-ASHEE (based on the standard OpenFOAM solver pisoFOAM

Table 5
List of Pilot Demonstrators (PDs) in ChEESE-1P and associated Technology Readiness Level (TRL) at the start/end of the project. Service Type: (UC) Urgent Computing, (EWS) Early Warning System, (PHA) Probabilistic Hazard Assessment, (SI) Seismic Inversion.

PD	Name	Initial TRL	Final TRL	Service type
PD1	Urgent seismic simulations	3	5	UC
PD2	Faster than real-time tsunami simulations	3	9	EWS
PD3	High-resolution volcanic plume simulation	2	4	none
PD4	Physics-based tsunami-earthquake interaction	2	4	none
PD5	Physics-based probabilistic seismic hazard assessment	3	6	PHA
PD6	Probabilistic volcanic hazard assessment	3	7	PHA
PD7	Probabilistic tsunami hazard assessment	3	6	PHA
PD8	Probabilistic Tsunami Forecast for early warning and rapid post event assessment	3	8	EWS/UC
PD9	Seismic tomography	3	6	SI
PD10	Array-based statistical source detection and restoration and ML from monitoring	2	4	none
PD11	Geomagnetic forecasts	2	4	none
PD12	High-resolution ensemble-based volcanic ash dispersal forecast	3	9	UC

Table 6
Technology Readiness Level (TRL) scale in the EU funded projects arena and its adaptation to ChEESE-1P PDs and services. This tailored TRL definition was used as a Key Performance Indicator (KPI) to monitor the progress of the PDs.

TRL	EU definition	ChEESE-1P PD/service concept
1	Basic principles observed	Equations and Quantities of Interest defined
2	Technology concept formulated	Efficiency metrics defined and implementation strategy assessed
3	Experimental proof of concept	Proof of scalability and efficiency of individual components
4	Technology validated in lab	Proof of models/data interoperability in workflows
5	Technology validated in relevant environment	Individual components demonstrated in relevant environments
6	Technology demonstrated in relevant environment	Use-case tests demonstrated in relevant environments
7	System prototype demonstration	Individual components demonstrated in operational environments
8	System complete and qualified	Use-case tests demonstrated in operational environments
9	System proven in operational environment	Service deployed and operation-ready

for compressible single-phase flows), mini-FALL3D (including both the CPU and the GPU versions of the main FALL3D kernel), mini-SeisSol (executes kernels on artificial data), and mini-SPECFEM3D (contains the most computationally intensive kernel of the code). All the mini-apps were tested on architectural testbeds representative of industry trends including CPUs (e.g. Intel Xeon Scalable processors, AMD EPYC Zen3, ARM ThunderX2 and Graviton2) and GPUs (NVIDIA V100 and A100 and AMD MI50 and MI100). Other exascale hardware prototypes, e.g. OpenSequana and RISC-V prototypes from the EuPEX and EuPilot European initiatives, were not yet available during ChEESE-1P but will be considered during the next implementation phase (see Section 7). Level B co-design focussed on mini-app performance at the node level to study aspects like, e.g. memory bandwidth, influence of the instruction set generated by compilers, performance gain obtained using novel GPU generations, or energy efficiency with respect to the compilation flags and the hardware architecture. The ultimate goal was to guarantee that flagship codes are performant on emerging exascale architectures and that the code optimisations done during the project will remain valid.

Finally, co-design level C focussed on full applications at single and multi-node levels in order to expose full application characterisation of CPU and GPU versions, evaluate performance gain of optimisations and porting of full applications and, finally, evaluate the energy efficiency with respect to the application optimisation level and the hardware prototype.

4. Pilot demonstrators

ChEESE-1P has implemented 12 Pilot Demonstrators (PDs) in which flagship codes and workflows were combined to address a number of capability and capacity computational challenges in the Solid Earth domain. Table 5 lists the different PDs together with their respective initial and final Technology Readiness Level (TRL) as defined in the project and used as a Key Performance Indicator (KPI) for monitoring the project evolution and success (see Table 6). PDs constituted, together with the flagship codes, the key ingredient of ChEESE-1P. On the one hand, PDs served to address relevant scientific problems in computational seismology, tsunami modelling, volcanology, and magnetohydrodynamics. On the other hand, those PDs reaching a TRL equal or greater than 5 during the course of the project were further developed to enable and validate potential HPC-based services on different aspects of geohazards like urgent computing, early warning systems, probabilistic hazard assessment, or seismic tomography and inversion (see Section 5).

4.1. Pilot demonstrators implementation and results

PD1 [35] addressed the capability challenge of delivering physics-based regional ground shaking maps at frequencies relevant for earthquake engineering and civil protection purposes (up to 10 Hz) and within a few hours following a significant seismic event. A Software as a Service (SaaS) workflow, called Urgent Computing Integrated Services for Earthquakes (UCIS4EQ), was developed with a backend that comprises a set of processes executed in an orderly and automated manner and orchestrated by a workflow manager. A background automatic earthquake alert service continuously queries earthquake origin and magnitude from the Federation of Digital Seismograph Networks (FDSN), disambiguates duplicated events coming from different agencies, unifies real-time earthquake hypocentre locations and moment magnitudes, assesses the potential earthquake impact, and determines whether the event has to trigger an urgent simulation based on pre-defined impact thresholds and geographical constraints. Should this happen, a smart center control performs an AI-based near real-time estimation of the Centroid Moment Tensor (CMT) to characterise the source term and its uncertainties [36], configures the underlying seismic simulator [37], and launches executions on EuroHPC infrastructures using an urgent seismic computing protocol. The final step in UCIS4EQ is a post-processing service that tailors results according to end-user requirements and delivers ground shaking proxies through a dedicated web-based frontend. PD1 was tested with the automated

UCIS4EQ set-up for the 30 Oct 2020 Samos–Izmir earthquake using Salvus up to 20 Hz, and manually in the South Iceland Seismic Zone (SISZ) using SeisSol. The former case included a full-scale capability execution on MareNostrum-4 machine using 3,000 computing nodes (84,000 cores), reaching an extrapolated performance of 9.6 PFlop/s.

PD2 [38–40] focussed on a Faster-Than-Real-Time (FTRT) prototype for tsunami simulations using the T-HySEA code on massively parallel multi-GPU architectures in order to increase the model spatial resolution, enabling also simulations with nested grids to produce high-resolution (metric scale) coastal inundation maps. The underlying PD2 workflow retrieves earthquake information from the FDSN to launch a number of tsunamigenic scenarios for the given event. This constitutes also the building block for Tsunami Early Warning Systems (TEWS) with uncertainty quantification (see PD8) and for the generation of probabilistic hazard analysis based on tens to hundreds of thousands of scenarios (see PD7). The workflow can use thousands of GPUs simultaneously with very good weak scaling (above 95% of parallel efficiency). PD2 was tested using synthetic tsunami scenarios in the Mediterranean Sea (15 arc-sec HySEA model resolution) and the Pacific Ocean (60 arc-sec resolution) with a time to solution up to 200 times FTRT. Moreover, a high-resolution inundation demonstrator was run for the Eastern Mediterranean Sea (Sicily) for an hypothetical tsunami triggered by a Mw=8 earthquake with 4 grid nesting levels down to 10 m resolution (14 times FTRT using 32 V100 GPUs at Marconi-100 machine for each single tsunami scenario).

PD3 [41] addressed Large-Eddy Simulations (LES) of multiphase volcanic plumes in a stratified multicomponent atmosphere using the ASHEE code. In addition, the resulting time-averaged vertical profiles of mass flow rate were used to couple the local-scale ASHEE model with the regional-scale FALL3D ash transport and dispersion model [42], thereby increasing the coherence of the simulated data with respect to satellite-based observations. Parallel performances were measured on different machines, up to about 12,000 cores. PD3 was used to reconstruct scenarios for the A.D. 79 Vesuvius Plinian eruption (Italy) [43] and the 2015 Calbuco sub-Plinian eruption (Chile) on the Joliot Curie machine at CEA/TGCC (Irene-Rome and Irene-Skylake partitions). For these historical events, the Eruption Source Parameters (ESPs) and the atmospheric conditions (profiles) driving ASHEE were obtained from the bibliography and the NCEP/NCAR reanalysis [44] respectively.

PD4 [45,46] implemented a fully-coupled framework for the simulation of multi-physics earthquake-tsunami scenarios, covering dynamic earthquake rupture at complex faults, the generation and propagation of seismic waves and time-dependent displacement of the seafloor, and generation and propagation of acoustic and gravity waves in the Ocean, accounting for complex bathymetry/topography. To this purpose, the SeisSol code was extended towards fully-coupled Earth and Ocean models that combine seismic, acoustic and surface gravity wave propagation in elastic (Earth) and acoustic (Ocean) materials. PD4 was demonstrated considering different dynamic rupture and tsunami characteristics in several large megathrust tsunamigenic earthquakes, the A.D. 365 event in the Hellenic Arc, the 2004 Sumatra–Andaman earthquake and subsequent Indian Ocean tsunami [47], and the 2018 Palu Sulawesi earthquake-tsunami [48]. The later example included an extreme-scale fully-coupled simulation using 3,072 nodes (147,456 cores) of the SuperMUC-NG supercomputer at LRZ, achieving more than 1 TFlop/s per node.

PD5 enabled physics-based seismic hazard assessments with state-of-the-art multi-physics earthquake simulation software (SeisSol, SPECFEM3D and ExaHyPE). SeisSol was combined with the Seismic Hazard and Earthquake Rates In Fault Systems (SHERIFS, [49]) open source tool to estimate annual rupture probabilities which, in conjunction with ground shakings from simulated scenarios, can generate hazard curves and maps for selected site locations. ChEESE-1P collaborated with the Southern California Earthquake Center (SCEC) to install the CyberShake platform [50] on MareNostrum-4 supercomputer with the goal of undertaking PSHA studies on Iceland and, in a longer term, to adapt it to the geological and geophysical contexts of other target regions in Europe. PD5 was tested on one Italian and two Icelandic scenarios. In Northern Iceland, PD5 developed a new fault system model of the complex Tjörnes Fracture Zone (TFZ) consistent with all relevant geophysical information of the region. This model now lays the foundation for a complete revision of the PSHA in the zone, both physics-based and conventional PSHA. In particular, the new fault system model contains three degrees of fault complexity for the Húsavík-Flatey Fault (HFF), the largest and primarily submerged strike-slip fault in Iceland that poses the greatest hazard to the coastal communities. Dynamic earthquake rupture scenarios were proposed on the HFF to explore the realistic range of ground motion amplitudes and the tsunami potential of fault rupture [51,52]. In Southern Iceland, CyberShake and the Anelastic Wave Propagator AWP-ODC simulator (an SCEC code not under the umbrella of ChEESE) were run assuming a subset of 17 faults for hazard calculation at selected cities and towns. Finally, SPECFEM3D was used to simulate the Mw 6.3 2019 L'Aquila and the Mw 6.5 2016 Norcia earthquakes with frequency up to 1 Hz on Marconi-100 machine at CINECA, including a sensitivity analysis (75,000 scenarios) of the influence of velocity model and kinematic fault parameters on ground shaking. Simulations required about 24 CPU hours on 50,000 cores.

PD6 enabled high-resolution (km scale) short-term (days) and long-term (years) Probabilistic Volcanic Hazard Assessment (PVHA) for atmospheric tephra dispersal and fallout over country-scale domains and accounting for the natural variability linked to the eruption scale and type. PD6 input data considered bibliography from geological record for the ESPs, the ECMWF ERA-5 reanalysis over 30 years for climatology and, for the short-term, seismic and deformation data from the Osservatorio Vesuviano operational monitoring network. The Workflow Management System WMS-light (see next section) was used to orchestrate the concurrent execution and combination of thousands of FALL3D model scenarios, thereby providing probability and hazard maps with associated uncertainty quantification. PD6 was used for long-term PVHA from Beerenberg volcano at Jan Mayen island (Norway), including implications for air traffic in the North Atlantic [53], and for Campi Flegrei caldera in the Neapolitan area [54], currently under a long-lasting unrest period (ongoing since 2004).

PD7 [55–57] implemented a workflow for site-specific high-resolution tsunami hazard analysis. This included tsunami source disaggregation and discretisation, rapid simulation of inundation scenarios using the T-HySEA code with telescopic grids and, finally, a city-scale hazard aggregation. PD7 applications focussed on the Mediterranean Sea, using tsunami source input data from a previous regional assessment produced by the TSUMAPS-NEAM (North-East Atlantic and Mediterranean) project [58]. In particular, high-resolution (inundation modelled at 10 m model resolution) tsunami hazard maps were produced for Catania and Siracusa in Sicily by running large ensembles of scenarios, in chunks using each 32 nodes of Marconi-100 machine (i.e. 128 V100 GPUs) [59]. Overall, the PD7 results comprised more than 2 million individual inundation simulations for 8 critical locations in the Mediterranean Sea (Heraklion, Messina, Thessaloniki and others), which will become a very valuable open dataset to benchmark other future PTHA applications within the scope of the EU Geo-INQUIRE project (see Section 6).

PD8 [60,61] dealt with rapid probabilistic forecasts of tsunami inundations for integration in near-field Tsunami Early Warning Systems (TEWS), using ensembles to reflect the large uncertainties on tsunamigenic earthquake parameters (location, magnitude, mechanism, slip distribution) that typically exist shortly after an event. The main difference with respect to PD7 is that the definition of the scenarios in a TEWS context results from merging available real-time seismic event parameters from the CAT-INGV monitoring room or using FDSN webservices, long-term probabilities for unavailable parameters (as in PD7), and stochastic simulations for predicting earthquake slip distributions. In addition, the PD8 workflow also converts results to Alert Levels (ALs), which refer to specific actions such as the evacuation of a coastal zone, thus introducing a transparent approach to downstream uncertainty communication and management in a TEWS. PD8 was implemented in the Mediterranean Sea, with the underlying workflow using either pre-calculated tsunami scenarios (e.g. from the TSUMAPS-NEAM database) or running on-the-fly a large ensemble of T-HySEA simulations with the FTRT time to solution premise. The former option is ready for operational use in a TEWS; the latter option, provided the availability of computational resources, is a potential game-changer for rapid post-event assessment, and it was successfully tested in a large-scale capacity workflow execution using 805 nodes (3,220 V100 GPUs) of Marconi-100 machine, reaching an extrapolated performance of 25.7 PFlop/s.

PD9 considered the imaging of internal structures of the Earth at different scales using Salvus and SPECFEM3D in conjunction with the Large-scale Seismic Inversion Framework (LASIF, [62]). LASIF is a collaborative development between two ChEESE-1P partners (ETH Zurich and CNRS) for a data-driven workflow tool performing Full Waveform Inversions (FWIs) of continuous seismological data (waveforms) from the FDSN. For SPECFEM3D, a new code branch for efficient gradient computation and iterative model update (SPCEFEM3D_FWI) was developed [63,64] in order to optimise the most costly component of the workflow (i.e. forward modelling and adjoint inversion in complex 3D media). For Salvus, relevant tools for data selection and visualisation as well as new functionalities for generating inputs files were added. PD9 was validated inverting up to 9 Hz a real dataset from exploration geophysics in the North Sea consisting on 138 Ocean Bottom Seismometers (OBS, about 70,000 shots per seismometer) and using 256 nodes of Marconi-100 (1,024 V100 GPUs) over 24 h. A second PD9 use case focussed on the regional-scale Iberia–Pyrenees using stations in France, Spain and Morocco during the period 2011–2020 (inversion using 800 stations).

PD10 [65] implemented a scalable data-streaming workflow for automatic detection, location, and characterisation of seismic sources from continuous waveforms recorded by large seismic networks in order to improve seismic monitoring systems and enrich the seismic catalogues with previously undetected short transient signals (e.g. local earthquakes, phase arrivals from distant events). PD10 ported the BackTrackBB code [66] to PyCOMPSs [67], thereby improving efficiency, scalability, and parallel workflow task management in centralised HPC and cloud environments. PD10 was tested in two real-case datasets, the Vrancea seismic source in Romania, using continuous seismological data (waveforms) from the Romanian Seismic Network (RSN) at 46 stations spanning over a 1 month period (revealing 20 previously unregistered events), and a 4-year long continuous seismic dataset from stations in the Shikoku region, Japan [68].

PD11 [69] focussed on modelling the Earth's geomagnetic field, including both the simulation of past geomagnetic reversals with an unprecedented level of accuracy (using xSHELLS) and ensemble-based geomagnetic forecasts for the next few decades (using PARODY_PDAF), an aspect of particular interest when planning future space missions. In both cases, workflow tasks were orchestrated with WMS-light using archeomagnetic, volcanic and historical data from the HISTMAG database [70]. PD11 was used to simulate with a global 3D spherical resolution of the order of 10 km during the equivalent geological time of 4 My, producing 5 geomagnetic polarity reversals and 9 so-called excursions [71,72]. On the other hand, ensemble forecasts considered multiple dynamo models with 512 members on 5,760 cores of Irene-Skylake machine.

Finally, **PD12** [73–75] implemented an ensemble-based data assimilation system (workflow) combining the FALL3D atmospheric dispersal model with high-resolution geostationary Infrared satellite retrievals (Meteosat −11, GOES-16, and Himawari-8 satellites). To this purpose, a parallel data assimilation system was implemented in FALL3D in order to run volcanic ash forecasts at an unprecedented resolution and over wide (continent-size) areas and driven by outputs from different meteorological datasets, including global model forecasts (GFS), global analyses and re-analyses (GDAS, ERA5), and mesoscale models (WRF/ARW). PD12 was validated against multiple past events including, e.g. the 2010 Eyjafjallajökull eruption (Iceland) or the February 2013 Etna event (Sicily), reaching a TRL of 9 (used operationally in a real case, see Section 5 for details).

4.2. Data sources

Table 7 compiles the most relevant input data required to run the different PDs (see references in Section 4.1 for details). In a broad sense, input data types in PDs can be classified as: (i) static; available from multiple external sources and typically stored and pre-processed locally before feeding the workflow, (ii) dynamic (data stream cached locally); continuously gathered from external repositories (e.g. to screen the occurrence and characteristics of a given event) and, (iii) co-located with the HPC systems (e.g. large pre-defined model inputs such as grids or configuration files). ChEESE-1P followed a Data Management Plan (DMP) to ensure that the datasets generated and post-processed during the project were managed according to the Findable, Accessible, Interoperable, and Reusable (FAIR) guiding principles [76]. The associated metadata information is similar to the Common European Research Information Format (CERIF) metadata model, the same used by the European Plate Observatory System Research Infrastructure (EPOS-ERIC), but limited to the goals and data typology of the different PDs.

4.3. Modelling workflows and tools

In parallel with the development and implementation of the different PDs, ChEESE-1P considered different heterogeneous aspects of modelling that go beyond the strict code preparation and optimisation. These included:

1. A geometric mesh partitioning library (Amik-1.0) based on a Hilbert Space-Filling curve [77] that can be used either as standalone (e.g. for domain decomposition during code preprocess) or in runtime if coupled to a code-specific data redistribution library (e.g. for mesh adaptivity). With respect to other topological mesh partitioners (e.g. Zoltan, [78]), advantages of the geometric approach are that the domain decomposition load balance can be adjusted perfectly and that the mesh partition becomes independent of the number of CPUs, resulting in a much faster and highly-scalable partitioner.

2. Support to post-process and visualisation of large 3D and 4D datasets using the distributed data-parallel scientific visualisation in virtual reality (Vistle) software. This included coding special ChEESE extensions to import data from SeisSol and T/L-HySEA codes.

Table 7
Relevant input data sources for PDs. Data source: (FDSN) Federation of Digital Seismograph Networks, (ORFEUS) Observatories and Research Facilities for European Seismology, (IRIS) Incorporated Research Institutions for Seismology, (EMSC) European Mediterranean Seismological Center, (CMT) Global Centroid-Moment-Tensor, (GHEA) Global Historical Earthquake Archive, (EFSM20) European Fault-Source Model 2020, (NEAMTHM18) NEAM Tsunami Hazard Model 2018, (GFS) Global Forecast System. Data formats: (SEED) Standard for Exchange of Earthquake Data, (SAC) Seismic Analysis Code, (NPY) binary file format for arbitrary NumPy array on disk . Data type: (1) co-located with the HPC, (2) local; accessed from external sources, (3) cached locally (data stream).

PD input data	Data source	Data format	Data type
Seismic time series (waveforms) and parametric information	FDSN, ORFEUS, IRIS, EMSC	SEED, SAC, ASDF	(3)
Earthquake source parameters	seismic agencies	ASCII, QuakeML	(2)
Seismic catalogues	CMT project, GHEA	QuakeML, ASCII	(2)
Seismogenic sources	EFSM20	GeoJSON	(2)
Probabilities for earthquake mechanisms; regional hazard	NEAMTHM18	NPY, MAT-files	(2)
Multi-parametric monitoring networks	*e.g.* from INGV	TSDSystem	(3)
Geological and 3D velocity models	case-dependent	various	(1)
Bathymetry and Topography	open or proprietary	netCDF	(1)
Eruption Source Parameters (ESP)	bibliography (historical events)	text	(2)
Meteorological forecasts and re-analyses	*e.g.* GFS, ERA5	grib, netCDF	(1)–(2)
Archeomagnetic and historical data	HISTMAG	ASCII	(2)
Configuration file(s)	case/model specific	various	(1)

Table 8
Characteristics of the service validation live exercises. Service type: (UC) Urgent Computing, (EWS) Early Warning System, (PHA) Probabilistic Hazard Assessment. IUB members: (ISAVIA) Iceland air navigation service provider, (VAAC) Volcanic Ash Advisory Center, (PLINIVS) Italian civil protection competence centre, (ARISTOTLE) ARISTOTLE-eENHSP project, (NEAM-TSP) Tsunami Service Providers from the NEAM region, (IGN) Spanish National Geographic Institute, (SPC) Spanish Civil Protection, (GP) Global Parametrics. See Table 2 for hardware characteristics.

Exercise	PDs	Exercise date	Driven IUB	Type	Nodes	Machine	Time-to-solution
1a	PD12	12 Mar 2021	IMO and ISAVIA	UC	84	Nord-3	20 min
1b		4 Nov 2021	Buenos Aires VAAC		48	MareNostrum-4	25 min
2	PD6	4 Nov 2021	PLINIVS and ARISTOTLE	PHA	24	MareNostrum-4	3 h
3	PD8	5 Nov 2021	NEAM-TSP	EWS	805	Marconi-100	20 min
4	PD2	22 Nov 2021	IGN and SCP	EWS	72	Marconi-100	2 min
5	PD1	18 Jan 2022	GP and ARISTOTLE	UC	10	MareNostrum-4	<1 h
6	PD4	18 Jan 2022	various	PHA	85	Super-MUC	<6 h

3. At the beginning of the project, some of the PDs inherited consolidated workflow frameworks (e.g. based on scripting, on LASIF [62], or on other diverse tailored solutions) whereas, in contrast, other PDs (e.g. PD6 or PD11) completely lacked any kind of workflow support. This deficit motivated the implementation of an HPC Workflow Management System (WMS-light, [79]) as a general tool to execute component-based workflows, extendible beyond the geosciences domain. WMS-light is built on a light-weight approach that reduces tailoring of the application and/or infrastructure to the absolute minimum while still meeting the major workflow execution requirements. In fact, executable workflow units in WMS-light (i.e. the components) can actually be any user-defined software, such as binaries, shell scripts, or even other local workflow systems. WMS-light offers a non-intrusive programming model able to meet the requirements of most batch-based HPC workflows while keeping the specification simple and intuitive for users and developers by means of a simple JSON-format file. On the other hand, WMS-light components can be distributed across multiple heterogeneous computational infrastructures and queuing systems without any implication for their design.

5. Service enabling

Up to 8 PDs reaching a TRL of 5 or above were further enabled as potential services covering different aspects of geohazards (see Table 5) and made available to the scientific community and the stakeholders represented by the Industry and Users Board (IUB). The IUB engaged around 30 representatives from the public sector (e.g. civil protection agencies, met offices, governance bodies), industry and SMEs, European and global academic networks, as well as other European research projects. The role of the IUB was to provide independent feedback on project interim results, service functional requirements, operational service validation and, finally, identification of potential exploitation paths. In terms of service validation, PDs reaching TRLs of 5–6 were self-assessed by scientists in the project through the execution of use cases, whereas PDs reaching higher TRLs of 7-8-9 were validated during user-driven live exercises, with service functional and operational requirements imposed by end-users from the IUB. From the very beginning of the project, a Project Innovation Manager (PIM) performed continuous liaison activities and periodic meetings with IUBs to enlarge the number of memberships, foster their engagement and participation, and ensure a true co-design methodology for service enabling. Overall, ChEESE-1P entailed 14 use cases and 6 user-driven exercises engaging a total of 53 different institutions (represented by 99 individuals) from 26 different countries. Table 8 shows the time-to-solution requirements and the HPC resources allocated in showcasing the services, which can be summarised as follows:

Exercise 1 consisted of a demonstration of PD12 in operational settings, and it was actually conducted twice (1a and 1b). Firstly, the Icelandic Meteorological Office (IMO) participates in VOLCEX, a regular volcanic ash exercise in the EUR/NAT Regions promoted by the International Civil Aviation Organisation (ICAO) and involving EUROCONTROL, airlines, airports, meteorological offices, and Aeronautic Service Providers (ANSPs). The VOLCEX-21 exercise was held in Iceland and tested the developments of PD12 in an operational setting. Secondly, IMO also organised the VOLCICE exercise, a country-scale reduced version of VOLCEX, in conjunction with the U.K. Met. Office, the British Geological Survey (BGS), the Icelandic Civil Protection, and ISAVIA (national air service provider), the last two being members of the IUB. In the 2021 edition, the VOLCICE exercise simulated the response to an eventual explosive eruption at the Beerenberg volcano (Jan Mayen island, Norway) and PD12 delivered live forecasts using 21 ensemble members at 5 km grid resolution. Finally, PD12 participated also in a live demonstration with the Buenos Aires

VAAC (Argentina) for the South American airspace running 48 FALL3D scenarios at 2 km resolution.

Exercise 2 demonstrated PD6 for short-term Probabilistic Volcanic Hazard Assessment (PVHA) at Campi Flegrei (Italy) and consisted of running 300 FALL3D scenarios at 2 km grid resolution orchestrated by WMS-Light and in response to a fictitious volcanic unrest with synthetic real-time data furnished by the Osservatorio Vesuviano surveillance system. The resulting hazard and probability maps were pipped to PLINIVS, a competence centre of the Italian Civil Protection Department, for further quantification of tephra fallout impacts on Southern Italy infrastructures (e.g. roads, grid systems, power plants) and population.

Exercise 3 was a PD8 live demo during the World Tsunami Awareness Day (5th November) that is promoted by the UN Disaster Risk Reduction (UNDRR) facilitates in collaboration with the rest of the United Nations system. The event consisted of the simulation of a rapid probabilistic post-event tsunami assessment following the 30 October 2020 Mw 7.0 Samos–Izmir earthquake, and entailed a substantially large capacity workflow using 805 nodes of Marconi-100@CIN (i.e. 3,220 V100 GPUs) that delivered around 38,000 T-HySEA scenarios exploring all the uncertainty ranges. The exercise engaged representatives of the NEAM region Tsunami Service Providers (KOERI, NOA, and INGV), in charge for tsunami warning in the region, as well as national and European Civil Protection representatives.

Exercise 4 showcased the PD2 Faster Than Real Time (FTRT) tsunami simulations considering several tsunamigenic seismic source scenarios in the Gulf of Cadiz that were proposed on-the-fly by the Spanish Tsunami Early Warning System Center. It consisted of deterministic T-HySEA simulations of a single event but including variability in the tsunami source (135 scenarios), and used 68 nodes of M100@CIN (272 NVIDIA V100 GPUs) with a strict workflow time-to-solution constraint of 2 min, imposed by the FTRT requirement.

Exercise 5 delivered shake maps of ground motion proxies (peak ground acceleration, velocity, and others) in urgent computing mode for post-event evaluation of earthquake impacts. The PD1 Urgent Computing Integrated Services for EarthQuakes (UCIS4EQ) integrated with Salvus emulated the 2020 Samos–Izmir event at 1.5 Hz (5 Hz offline) in about 30 min, using a web-based dedicated user portal to modify parameters on-the-fly, launch and monitor the queuing status, perform automated data transfers using the EUDAT B2Drop utility and, finally, visualise the resulting maps. This showed the potential of wave propagation simulations to overcome the current data-driven sparse approaches based on Ground Motion Prediction Equations (GMPE) in case of quick post-event impact quantification.

Finally, **exercise 6** showcased the SeisSol code in 3D fully-coupled high-resolution (1.5 to 5 Hz) earthquake and tsunami dynamics and the ExaHyPE code for uncertainty quantification, both simulating the 2018 Palu-Sulawesi event. The live demonstration showed how interference of seismic and acoustic waves may be dominant in data recorded by offshore instruments, which motivates fully coupled acoustic–elastic coupling with gravity solving an entirely new class of earthquake-tsunami problems.

As success stories, it is worth mentioning that two services reached a TRL of 9 and became fully operational. Firstly, the workflow developed by PD2 is already exploited operationally by IGN (the Spanish Instituto Geográfico Nacional) and by the ARISTOTLE-eENHSP, delivering to the European Emergency Response Coordination Centre (ERCC). The ARISTOTLE tsunami service is also integrated in the Scientific Products Archiving and Document Assembly (SPADA) IT platform. ChEESE-1P exercises performed on-the-fly multi-scenario simulations in just a few minutes (i.e. FTRT), clearly showing the enormous potential of large-scale computing in allowing near-future TEWS to overcome the current emergency-management mechanisms based on decision matrices and pre-computed expected scenarios. Secondly, PD12 ran operationally on MareNostrum-4 at BSC during the eruption of Cumbre Vieja volcano at La Palma, Canary Islands (19 September–13 December 2021). The resulting ensemble-based deterministic and probabilistic ash and SO_2 dispersal forecasts were delivered daily (at 8:00 am LT) to the emergency management scientific committee and to the regional government and civil protection authorities for real operational decision-making.

The service enabling strategy in operational environments proved very useful to identifying target users, HPC access modes (urgent computing in particular), the digital landscape (data, models and workflows), and the information flow involved in the decision-making process. ChEESE-1P pioneered the definition of urgent computing services for natural disasters (with emphasis on seismic/tsunami simulations) to be used in the context of the EuroHPC JU as a model for future urgent computing (emergency) access mode.

6. Community building

A final general high-level objective of ChEESE-1P was that of community building by integrating around HPC research institutions, academia, industry, and public governance bodies concerned with geohazards. To this end and, in addition to the aforementioned engagement with the IUB members, the consortium pursued various actions and activities that included:

1. Training activities. ChEESE-1P organised 15 training courses including 5 PRACE Advanced Training Courses (PATC) on the use of flagship codes and 2 EU-ASEAN schools. In total, training activities involved over 400 attendees from 45 different countries in Europe, North and Latin-America, and ASEAN Member States.

2. A Zenodo collection of ChEESE-1P codes and tools (see Table 3 for references) for code distribution and data sharing with efforts already in place to integrate the ChEESE catalog as a future European Plate Observing System (EPOS-ERIC) service including definition of a metadata scheme for software and ensure interoperability with the EPOS catalog (starting from the EPOS Volcano Thematic Core Service software metadata model).

3. Dissemination of scientific results and exascale road map definition in geosciences. In addition to project-related publications and presentations in almost 100 international conferences, workshops and other events, ChEESE-1P promoted a women-only editorial board in the Frontiers in Earth Sciences journal research topic entitled "High-Performance Computing in Solid Earth Geohazards: Progresses, Achievements and Challenges for a Safer World". The CoE was also granted with the Galileo conference "Solid Earth and Geohazards in the Exascale Era" (Barcelona, 23–26 May 2023) that will involve 9 European Geosciences Union (EGU) Divisions to discuss a consensual exascale implementation roadmap in Solid Earth sciences. Dissemination activities also embraced wider audiences, with 15 general-public official videos, 34 press clippings in National Geographic, BBC and other media sites, and multiple activities during the International Day of Women and Girls in Science (11 February 2020, 2021 and 2022). These were actively promoted by the ShEESE (a combination of She and ChEESE) committee, which supported the visibility of female researchers in many different ways throughout the project.

4. Exploitation strategies and market analysis. Continuous actions existed to maximise the impact of the project and the transfer of knowledge and results into the research community, industry, public institutions, policymakers and society. To this end, exploitable results were identified and considered in a targeted market analysis screening for early adopters and longer-term users. As a result, the number of IUB members was continuously enlarged and the synergies with other Centers of Excellence and related initiatives encouraged.

Table 9
List of Pilot Demonstrators (PDs) and related Simulation Cases (SC) in ChEESE-2P. Service Type: (UC) Urgent Computing, (TCS-TSU) EPOS Thematic Core Service in tsunamis, (TCS-VO) EPOS Thematic Core Service in volcanoes.

PD	SC	PD/SC name	Target TRL	Service or outcome
PD1		Extreme-scale modelling of seismic hazards	7	
	SC1.1	Anthropogenic noise and seismicity at higher frequencies in Rade d'Hyères		datasets
	SC1.2	Earthquake simulators for the Hellenic arc and Iceland		datasets
	SC1.3	Physics-based PSHA and urgent computing with UQ for Iceland		UC
PD2		High-resolution seismic inversion and tomography	6	
	SC2.1	Seismic tomography of the Adriatic-Ionian region		datasets
PD3		Global probabilistic tsunami hazard and uncertainty quantification	8–9	
	SC3.1	The Global Tsunami Hazard Map		TCS-TSU
PD4		Complex multi-source tsunami modelling	5	
	SC4.1	Multi-source high-resolution scenario (the 2022 Tonga case)		UC
PD5		Ensemble-based volcanic dispersal across multiple scales	7–8	
	SC5.1	Towards the European tephra hazard		TCS-VO
	SC5.2	Volcanic dispersal at local scale (the 2021 La Palma case)		UC
PD6		Multiphase 3D volcanic explosion modelling	7	
	SC6.1	Urgent simulation of phreatic eruptions at Vulcano island		UC
	SC6.2	Long-term probabilistic hazard maps for phreatic eruptions at Vulcano island		datasets
PD7		The Earth's dynamo model	5	
	SC7.1	MHD turbulence closure models in planetary cores using AI		datasets
PD8		Geodynamics to geohazards	5	
	SC8.1	Magmatic systems and induced landslides: the Etna case		datasets
	SC8.2	Tectonic deformation as driver of seismic hazard in the Hellenic arc		datasets
PD9		Glacial outburst floods	4	
	SC9.1	Eastern Skaftá cauldron GLOF simulation, Iceland		datasets
	SC9.2	Calving glacier hazard in Greenland		datasets

Overall, ChEESE-1P was pivotal in leveraging a flourishing ecosystem of European projects and initiatives tackling computational geohazards that will benefit from current and upcoming exascale EuroHPC infrastructures and uptake substantial parts of the project developments. In particular, it is worth mentioning:

eFlows4HPC (2021–2024, GA No 955558). A EuroHPC-funded project for workflow software stack integrating HPC simulations with big data analytics and machine learning, with a pillar use case on natural catastrophes [80].

Geo-INQUIRE (2022–2024, GA No 101058518). A Horizon Europe project for virtual and trans-national service access to data and state-of-the-art numerical models and workflows for monitoring and simulation of the dynamic processes in the geosphere at unprecedented levels of detail and precision. Access to several services developed in ChEESE-1P will be enabled through Geo-INQUIRE to facilitate more HPC-driven geohazard applications.

DT-GEO (2022–2025, GA No 101058129). A Horizon Europe project to deploy a prototype Digital Twin (DT) on geophysical extremes with long-term ambition towards the Destination Earth initiative. A number of the workflows developed in ChEESE-1P will enter as components into the different engines of the various twin components to be developed in DT-GEO.

ChEESE-2P (2023–2026, GA No 101093038). The second project implementation phase, described in next section.

Finally, the **European Plate Observing System** (EPOS), constituted as a European Research Infrastructure Consortium (ERIC) in 2018 and that federates European national research infrastructures providing access to data and scientific products. EPOS is currently under a transition period to deploy Integrated Core Services Distributed (ICS-D), including HPC-based components. EPOS has been closely linked to ChEESE-1P since the very beginning through its involvement in the Scientific and Industry Users boards.

7. ChEESE-2P: towards the next project phase

ChEESE-2P (2023–2026) is continuing with the preparation and optimisation of flagship codes on computational seismology, tsunami science, volcanology, and magnetohydrodynamics in order to bring them to the next level and consolidate the role of the CoE in harnessing the scientific community, the EuroHPC infrastructures, the National Competence Centers (NCCs), and public and private industry parties. However, ChEESE-2P includes also community codes from geodynamics and glacier hazards, two disciplines that were untapped during the first project phase. Enlarging the application domains pursues the engagement of other geophysics communities and the attraction of potential end-users from other countries (e.g. eastern and northern Europe). Similarly to the first project phase, the flagship code preparation roadmap articulates around the adequacy of applications to the new EuroHPC heterogeneous hardware systems (including porting, optimisation, audits, and performance monitoring) and on co-design with major EU initiatives on OpenSequana (EuPEX) and RISC-V (EuPilot) exascale hardware prototypes, ensuring an effective integration of co-design hints back to the main applications. The second phase of the project also considers critical exascale technical challenges not fully addressed yet, including code resilience, fault-tolerance, (run-time) load balancing, energy efficiency, code performance portability and containerisation for deployment on pre-exascale tier-0 (LUMI, Leonardo, MareNostrum-5) and petascale tier-1 EuroHPC infrastructures. This includes the preparation of services and policy recommendations for an eventual EuroHPC emergency (urgent computing) access mode in case of geohazards, for which containerisation of components or whole workflows allowing fast and reliable deployment and setup of simulations is particularly relevant.

ChEESE-2P farms a new generation of 9 Pilot Demonstrators (Table 9) to replace those predecessors that reached a higher TRL and that, to some extent, were transferred elsewhere in the ecosystem for further development or downstream service implementation. The new PDs are underpinned by concepts like multi-physics, multi-scale, multi-source, multi-hazard, and integration of phenomena that will require developments in flagship codes in order to incorporate new physics (functionalities), couplings, and forcing terms. During the project, the PDs will materialise in various capability and capacity Simulation Cases

(SCs) defined on the basis of scientific interest, social relevance, window of opportunity (e.g. hazardous events that have occurred recently or are likely to occur in the near future) and homogeneous geographic coverage across Europe and beyond. The resulting datasets (e.g. hazard maps) will be EPOS-compliant and follow an EOSC-enabled quality approach, including automatic FAIR validation. In addition, and in order to foster the urgent computing access mode in EuroHPC tier-0/tier-1 systems for emergency management during high-impact geohazardous events (earthquakes, tsunamis, and volcanoes), 4 PDs will be exploited for urgent computing service enabling, including related technical challenges and access policy recommendations.

Finally, a last high-level objective is to synergise with the Geo-INQUIRE and DT-GEO projects. The former will provide Virtual and TransNational access to the ChEESE codes and workflows for a broader user community; the latter will foster the integration of codes, workflows and services to the Destination Earth initiative coordinated by the EC Directorate-General for Communications Networks, Content and Technology (DG-CNECT) in support of the European Green Deal, and will implement high-resolution operational digital twins with tailored access to high-quality knowledge for user-specific scenario development for decision support.

CRediT authorship contribution statement

Arnau Folch: Conceptualisation, Software, Validation, Writing – original draft, Writing – Review, Supervision, Resources, Project administration, Funding acquisition. **Claudia Abril:** Investigation, Validation. **Michael Afanasiev:** Software. **Giorgio Amati:** Software, Resources. **Michael Bader:** Methodology, Software, Investigation, Resources, Review, Project administration. **Rosa M. Badia:** Software, Resources. **Hafize B. Bayraktar:** Investigation. **Sara Barsotti:** Investigation, Validation, Resources. **Roberto Basili:** Validation. **Fabrizio Bernardi:** Validation, Investigation. **Christian Boehm:** Software. **Beatriz Brizuela:** Methodology. **Federico Brogi:** Software, Investigation. **Eduardo Cabrera:** Software. **Emanuele Casarotti:** Investigation. **Manuel J. Castro:** Investigation. **Matteo Cerminara:** Methodology, Software, Investigation, Resources. **Antonella Cirella:** Investigation. **Alexey Cheptsov:** Methodology, Software. **Javier Conejero:** Software. **Antonio Costa:** Methodology, Software, Investigation, Review, Project administration. **Marc de la Asunción:** Software. **Josep de la Puente:** Conceptualisation, Investigation, Resources, Project administration. **Marco Djuric:** Visualisation. **Ravil Dorozhinskii:** Software. **Gabriela Espinosa:** Project administration. **Tomaso Esposti-Ongaro:** Methodology, Investigation, Data curation, Review, Project administration. **Joan Farnós:** Supervision, Review, Project administration. **Nathalie Favretto-Cristini:** Investigation. **Andreas Fichtner:** Investigation, Review. **Alexandre Fournier:** Methodology, Software, Investigation, Review, Project administration. **Alice-Agnes Gabriel:** Methodology, Software, Investigation, Resources, Project administration. **Jean-Matthieu Gallard:** Investigation. **Steven J. Gibbons:** Methodology, Investigation, Resources, Review. **Sylfest Glimsdal:** Investigation. **José Manuel González-Vida:** Investigation. **Jose Gracia:** Software, Project administration. **Rose Gregorio:** Review, Project administration. **Natalia Gutierrez:** Software. **Benedikt Halldorsson:** Methodology, Investigation, Review, Project administration. **Okba Hamitou:** Software, Resources. **Guillaume Houzeaux:** Software. **Stephan Jaure:** Investigation. **Mouloud Kessar:** Software. **Lukas Krenz:** Software. **Lion Krischer:** Software. **Soline Laforet:** Software, Resources, Project administration. **Piero Lanucara:** Methodology, Project administration, Review, Resources. **Bo Li:** Investigation. **Maria Concetta Lorenzino:** Investigation. **Stefano Lorito:** Methodology, Validation, Investigation, Resources, Review. **Finn Løvholt:** Methodology, Investigation, Resources, Review, Project administration. **Giovanni Macedonio:** Software. **Jorge Macías:** Software, Investigation, Resources, Project administration. **Guillermo Marín:** Visualisation. **Beatriz Martínez Montesinos:** Investigation, Validation. **Leonardo Mingari:** Software, Investigation. **Geneviève Moguilny:** Investigation. **Vadim Montellier:** Methodology, Software, Investigation, Resources, Project administration. **Marisol Monterrubio-Velasco:** Methodology, Investigation. **Georges Emmanuel Moulard:** Software. **Masaru Nagaso:** Investigation. **Massimo Nazaria:** Software. **Christoph Niethammer:** Software, Project administration. **Federica Pardini:** Investigation. **Marta Pienkowska:** Methodology, Investigation, Resources, Review, Project administration. **Luca Pizzimenti:** Methodology. **Natalia Poiata:** Software, Investigation. **Leonhard Rannabauer:** Software. **Otilio Rojas:** Investigation. **Juan Esteban Rodriguez:** Software. **Fabrizio Romano:** Investigation. **Oleksandr Rudyy:** Software. **Vittorio Ruggiero:** Software, Resources. **Philipp Samfass:** Software. **Carlos Sánchez-Linares:** Methodology, Software. **Sabrina Sanchez:** Investigation. **Laura Sandri:** Validation, Investigation, Resources, Project administration. **Antonio Scala:** Investigation. **Nathanael Schaeffer:** Software, Investigation. **Joseph Schuchart:** Software. **Jacopo Selva:** Methodology, Validation, Investigation, Resources. **Amadine Sergeant:** Investigation. **Angela Stallone:** Investigation. **Matteo Taroni:** Validation. **Soelvi Thrastarson:** Investigation. **Manuel Titos:** Investigation, Validation. **Nadia Tonelllo:** Data curation. **Roberto Tonini:** Investigation. **Thomas Ulrich:** Investigation. **Jean-Pierre Vilotte:** Methodology. **Malte Vöge:** Investigation. **Manuela Volpe:** Validation, Investigation, Resources, Review. **Sara Aniko Wirp:** Investigation. **Uwe Wössner:** Visualisation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Codes and data available in a Zenodo repository.

Acknowledgements

This work has been funded by the European Union Horizon 2020 research and innovation program under the ChEESE project, Grant Agreement No. 823844, by the European High Performance Computing Joint Undertaking (JU), Grant Agreement No 101093038, and by 4 different Partnership for Advanced Computing in Europe (PRACE) projects from calls 20 and 21 for granting ChEESE activities with a total of 170M core hours on different machines: VOHA (ID 2019215114), TSUCAST (ID 2019215169), SEISVIEW (ID 2019215212) and TsuHazAP (ID 2020225386). We thank two anonymous referees for their constructive comments. This article is in memoriam of our colleague and friend Dimitri Komatitsch (1970–2019) [81] and to his invaluable legacy in computational seismology.

References

[1] MONDAIC, Salvus product, 2022, URL https://mondaic.com. (Last access 23 September 2022).
[2] V. Monteiller, M. Nagaso, A. Sergeant, SPECFEM_FWI-V1.0, 2022, http://dx.doi.org/10.5281/zenodo.6394675.
[3] K. Uphoff, et al., SeisSol: v1.0.0-rc, 2022, http://dx.doi.org/10.5281/zenodo.7038298.
[4] N. Schaeffer, xSHELLS code, 2022, URL https://nschaeff.bitbucket.io/xshells. (Last access 23 September 2022).

[5] A. Fournier, S. Sanchez, parody_pdaf, 2022, http://dx.doi.org/10.5281/zenodo.6415313.
[6] M. Bader, T. Weinzierl, ExaHyPE engine, 2022, URL https://www.exahype.org. (Last access 23 September 2022).
[7] M. Cerminara, ASHEE numerical code, 2022, http://dx.doi.org/10.5281/zenodo.6560777.
[8] J. Macías, C. Sánchez Linares, M. de la Asuncion, M. Castro, J. González Vida, T. Morales de Luna, S. Ortega, C. Escalante, edanya-uma/TsunamiHySEA: Tsunami-HySEA open-source, 2022, http://dx.doi.org/10.5281/zenodo.6400815.
[9] J. Macías, C. Sánchez Linares, M. de la Asuncion, M. Castro, J. González Vida, S. Ortega, T. Morales, C. Escalante, edanya-uma/LandSlideHySEA: Landslide-HySEA open-source, 2022, http://dx.doi.org/10.5281/zenodo.6400825.
[10] A. Folch, A. Costa, G. Macedonio, L. Mingari, FALL3D, 2022, http://dx.doi.org/10.5281/zenodo.6343786.
[11] A. Knupfer, C. Rossel, D. an Mey, S. Biersdorff, K. Diethelm, D. Eschweiler, M. Geimer, M. Gerndt, D. Lorenz, A. Malony, W.E. Nagel, Score-p: a joint performance measurement run-time infrastructure for periscope, scalasca, tau, and vampir, Technical Report, Oak Ridge National Lab. (ORNL), 2012, http://dx.doi.org/10.1007/978-3-642-31476-6.
[12] A. Munera, S. Royuela, G. Llort, E. Mercadal, F. Wartel, E. Quiñones, Experiences on the characterization of parallel applications in embedded systems with Extrae/Paraver, in: 49th International Conference on Parallel Processing, ICPP 20, 2020, pp. 1–11, http://dx.doi.org/10.1145/3404397.3404440.
[13] S. Balay, S. Abhyankar, M.F. Adams, S. Benson, J. Brown, P. Brune, K. Buschelman, E. Constantinescu, L. Dalcin, A. Dener, V. Eijkhout, W.D. Gropp, V. Hapla, T. Isaac, P. Jolivet, D. Karpeev, D. Kaushik, M.G. Knepley, F. Kong, S. Kruger, D.A. May, L.C. McInnes, R.T. Mills, L. Mitchell, T. Munson, J.E. Roman, K. Rupp, P. Sanan, J. Sarich, B.F. Smith, S. Zampini, H. Zhang, H. Zhang, J. Zhang, PETSc/TAO Users Manual, Technical Report. ANL-21/39 - Revision 3.17, Argonne National Laboratory, 2022.
[14] D.P. van Herwaarden, M. Afanasiev, S. Thrastarson, A. Fichtner, Evolutionary full-waveform inversion, Geophys. J. Int. 224 (1) (2020) 306–311, http://dx.doi.org/10.1093/gji/ggaa459.
[15] S. Thrastarson, M. van Driel, L. Krischer, C. Boehm, M. Afanasiev, D.-P. van Herwaarden, A. Fichtner, Accelerating numerical wave propagation by wavefield adapted meshes. Part II: full-waveform inversion, Geophys. J. Int. 221 (3) (2020) 1591–1604, http://dx.doi.org/10.1093/gji/ggaa065.
[16] V. Hapla, M.G. Knepley, M. Afanasiev, C. Boehm, M. van Driel, L. Krischer, A. Fichtner, Fully parallel mesh I/O using PETSc DMPlex with an application to waveform modeling, SIAM J. Sci. Comput. 43 (2) (2021) C127–C153, http://dx.doi.org/10.1137/20M1332748.
[17] C. Uphoff, M. Bader, Yet Another Tensor Toolbox for discontinuous Galerkin methods and other applications, 2019, CoRR. arXiv:1903.11521.
[18] R. Dorozhinskii, M. Bader, SeisSol on distributed multi-GPU systems: CUDA code generation for the modal discontinuous Galerkin method, in: HPC Asia 2021: The International Conference on High Performance Computing in Asia-Pacific Region, ACM, Virtual Event, Republic of Korea, 2021, pp. 69–82, http://dx.doi.org/10.1145/3432261.3436753.
[19] N. Schaeffer, Efficient spherical harmonic transforms aimed at pseudospectral numerical simulations, Geochem. Geophys. Geosyst. 14 (3) (2013) 751–758, http://dx.doi.org/10.1002/ggge.20071.
[20] D. Cébron, J. Vidal, N. Schaeffer, A. Borderies, A. Sauret, Mean zonal flows induced by weak mechanical forcings in rotating spheroids, J. Fluid Mech. 916 (2021) http://dx.doi.org/10.1017/jfm.2021.220.
[21] T.T. Clarté, N. Schaeffer, S. Labrosse, J. Vidal, The effects of a Robin boundary condition on thermal convection in a rotating spherical shell, J. Fluid Mech. 918 (2021) http://dx.doi.org/10.1017/jfm.2021.356.
[22] L. Nerger, W. Hiller, Software for ensemble-based data assimilation systems—Implementation strategies and scalability, Compute. Geosci. 55 (2013) 110–118, http://dx.doi.org/10.1016/j.cageo.2012.03.026.
[23] A. Fournier, L. Nerger, J. Aubert, An ensemble Kalman filter for the time-dependent analysis of the geomagnetic field, Geochem. Geophys. Geosyst. 14 (10) (2013) 4035–4043, http://dx.doi.org/10.1002/ggge.20252.
[24] T. Weinzierl, The Peano software—parallel, automaton-based, dynamically adaptive grid traversals, ACM Trans. Math. Software 45 (2) (2019) 14.
[25] A. Reinarz, D.E. Charrier, M. Bader, L. Bovard, M. Dumbser, K. Duru, F. Fambri, A.-A. Gabriel, J.-M. Gallard, S. Köppel, L. Krenz, L. Rannabauer, L. Rezzolla, P. Samfass, M. Tavelli, T. Weinzierl, ExaHyPE: An engine for parallel dynamically adaptive simulations of wave problems, Comput. Phys. Comm. 254 (2020) 107251, http://dx.doi.org/10.1016/j.cpc.2020.107251.
[26] K. Duru, L. Rannabauer, A. Gabriel, H. Igel, A new discontinuous Galerkin method for elastic waves with physically motivated numerical fluxes, J. Sci. Comput. 88 (2021) http://dx.doi.org/10.1007/s10915-021-01565-1.
[27] K. Duru, L. Rannabauer, A.-A. Gabriel, O.K.A. Ling, H. Igel, M. Bader, A stable discontinuous Galerkin method for linear elastodynamics in 3D geometrically complex elastic solids using physics based numerical fluxes, Comput. Methods Appl. Mech. Engrg. 389 (2022) 114386, http://dx.doi.org/10.1016/j.cma.2021.114386.
[28] J. Schuchart, P. Samfass, C. Niethammer, J. Gracia, G. Bosilca, Callback-based completion notification using MPI continuations, Parallel Comput. 106 (2021) 102793, http://dx.doi.org/10.1016/j.parco.2021.102793.
[29] P. Samfass, T. Weinzierl, A. Reinarz, M. Bader, Doubt and redundancy kill soft errors—Towards detection and correction of silent data corruption in task-based numerical software, in: 2021 IEEE/ACM 11th Workshop on Fault Tolerance for HPC At EXtreme Scale, FTXS, 2021, http://dx.doi.org/10.1109/FTXS54580.2021.00005.
[30] F. Brogi, S. Bnà, G. Boga, G. Amati, T.E. Ongaro, M. Cerminara, On floating point precision in computational fluid dynamics using OpenFOAM, 2022, arXiv. http://dx.doi.org/10.48550/ARXIV.2209.06105.
[31] J. Macías, M.J. Castro, S. Ortega, J.M. González-Vida, Performance assessment of Tsunami-HySEA model for NTHMP tsunami currents benchmarking. Field cases, Ocean Model. 152 (2020) 101645, http://dx.doi.org/10.1016/j.ocemod.2020.101645.
[32] A. Folch, L. Mingari, N. Gutierrez, M. Hanzich, G. Macedonio, A. Costa, FALL3D-8.0: a computational model for atmospheric transport and deposition of particles, aerosols and radionuclides – Part 1: Model physics and numerics, Geosci. Model Dev. 13 (3) (2020) 1431–1458, http://dx.doi.org/10.5194/gmd-13-1431-2020.
[33] A.T. Prata, L. Mingari, A. Folch, G. Macedonio, A. Costa, FALL3D-8.0: a computational model for atmospheric transport and deposition of particles, aerosols and radionuclides – Part 2: Model validation, Geosci. Model Dev. 14 (1) (2021) 409–436, http://dx.doi.org/10.5194/gmd-14-409-2021.
[34] M.A. Heroux, D.W. Doerfler, P.S. Crozier, J.M. Willenbring, H.C. Edwards, A. Williams, M. Rajan, E.R. Keiter, H.K. Thornquist, R.W. Numrich, Improving Performance Via Mini-Applications, Technical Report. SAND2009-5574, Sandia National Laboratories, 2009.
[35] J. de la Puente, J. Rodríguez, M. Monterrubio-Velasco, O. Rojas, A. Folch, Urgent supercomputing of earthquakes: Use case for civil protection, in: PASC 20: Proceedings of the Platform for Advanced Scientific Computing Conference, Vol. 9, 2020, pp. 1–8, http://dx.doi.org/10.1145/3394277.3401853.
[36] M. Monterrubio-Velasco, J.C. Carrasco-Jimenez, O. Rojas, J.E. Rodriguez, A. Fichtner, J. de la Puente, A statistical approach towards fast estimates of moderate-to-large earthquake focal mechanisms, Front. Earth Sci. 10 (2022) 743860, http://dx.doi.org/10.3389/feart.2022.743860.
[37] M. Pienkowska, J. Rodríguez, J. de la Puente, A. Fichtner, Deterministic modelling of seismic waves in the Urgent Computing context: progress towards a short-term assessment of seismic hazard, in: EGU General Assembly 2021, EGU21-15516, 2021, http://dx.doi.org/10.5194/egusphere-egu21-15516.
[38] J. Macías, M. de la Asunción, Faster and faster tsunami simulations with ChEESE, in: American Geophysical Union, Fall Meeting, NH33A-06, 2019, 2019AGUFMNH33A..06M.
[39] F. Løvholt, S. Lorito, J. Macias, M. Volpe, J. Selva, S. Gibbons, Urgent Tsunami computing, in: 2019 IEEE/ACM HPC for Urgent Decision Making, UrgentHPC, 2019, pp. 45–50, http://dx.doi.org/10.1109/UrgentHPC49580.2019.00011.
[40] B. Gaite, J. Macías, J.V. Cantavella, C. Sánchez-Linares, C. González, L.C. Puertas, Analysis of faster-than-real-time (FTRT) tsunami simulations for the Spanish Tsunami warning system for the atlantic, GeoHazards 3 (3) (2022) 371–394, http://dx.doi.org/10.3390/geohazards3030019.
[41] F. Brogi, G. Amati, G. Boga, M. Castellano, J. Gracia, T. Esposti-Ongaro, M. Cerminara, Optimization strategies for efficient high-resolution volcanic plume simulations with OpenFOAM, in: EGU General Assembly, EGU21-15310, 2021, http://dx.doi.org/10.5194/egusphere-egu21-15310.
[42] M. Cerminara, F. Pardini, A. Folch, L. Mingari, Volcanic ash: Coupling three-dimensional multiphase plume simulations with ash dispersion models, in: PASC '21: Proceedings of the Platform for Advanced Scientific Computing Conference, Association for Computing Machinery, New York, NY, USA, 2021.
[43] D.M. Doronzo, M.A. Di Vito, I. Arienzo, M. Bini, B. Calusi, M. Cerminara, S. Corradini, S. de Vita, B. Giaccio, L. Gurioli, G. Mannella, G.P. Ricciardi, I. Rucco, D. Sparice, M. Todesco, E. Trasatti, G. Zanchetta, The 79 CE eruption of vesuvius: A lesson from the past and the need of a multidisciplinary approach for developments in volcanology, Earth-Science Reviews 231 (2022) 104072, http://dx.doi.org/10.1016/j.earscirev.2022.104072.
[44] E. Kalnay, M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K.C. Mo, C. Ropelewski, J. Wang, A. Leetmaa, R. Reynolds, R. Jenne, D. Joseph, The NCEP-NCAR 40-year reanalysis project, Bull. Am. Meteorol. Soc. 77 (3) (1996) 437–472, http://dx.doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.
[45] E.H. Madden, M. Bader, J. Behrens, Y. van Dinther, A.-A. Gabriel, L. Rannabauer, T. Ulrich, C. Uphoff, S. Vater, I. van Zelst, Linked 3-D modelling of megathrust earthquake-tsunami events: from subduction to tsunami run up, Geophys. J. Int. 224 (1) (2020) 487–516, http://dx.doi.org/10.1093/gji/ggaa484.

[46] S. Aniko Wirp, A.-A. Gabriel, M. Schmeller, E. Madden, I. van Zelst, L. Krenz, Y. van Dinther, L. Rannabauer, 3D linked subduction, dynamic rupture, tsunami, and inundation modeling: Dynamic effects of supershear and tsunami earthquakes, hypocenter location, and shallow fault slip, Front. Earth Sci. 9 (2021) 626844, http://dx.doi.org/10.3389/feart.2021.626844.
[47] T. Ulrich, A. Gabriel, E. Madden, Stress, rigidity and sediment strength control megathrust earthquake and tsunami dynamics, Nat. Geosci. 15 (2022) 67–73, http://dx.doi.org/10.1038/s41561-021-00863-5.
[48] L. Krenz, C. Uphoff, T. Ulrich, A.-A. Gabriel, 3D acoustic-elastic coupling with gravity: the dynamics of the 2018 Palu, Sulawesi earthquake and tsunami, in: SC '21: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, No. 63, 2021, pp. 1–14, http://dx.doi.org/10.1145/3458817.3476173.
[49] T. Chartier, O. Scotti, H. Lyon-Caen, SHERIFFS: Open-source code for computing earthquake rates in fault systems and constructing hazard models, Seismol. Res. Lett. (2019) http://dx.doi.org/10.1785/0220180332.
[50] T. Jordan, S. Callaghan, R. Graves, F. Wang, K. Milner, C. Goulet, P. Maechling, K. Olsen, Y. Cui, G. Juve, K. Vahi, J. Yu, E. Deelman, D. Gill, CyberShake models of seismic hazards in Southern and Central California, in: Proceedings of the 11th National Conference in Earthquake Engineering, Earthquake Engineering Research Institute, Los Angeles, CA., 2018.
[51] F. Kutschera, S. Wirp, B. Li, A.-A. Gabriel, B. Halldórsson, C. Abril, Physics-based earthquake-tsunami modelling of the Húsavík-Flatey transform fault zone in North Iceland, in: EGU General Assembly, EGU22-13387, 2022, http://dx.doi.org/10.5194/egusphere-egu22-13387.
[52] L. Rannabauer, ExaSeis: Dynamic Rupture Benchmarks and an earthquake at the Husavik-Flatey fault, 2022, http://dx.doi.org/10.5281/zenodo.6386282.
[53] M. Titos, B. Martínez Montesinos, S. Barsotti, L. Sandri, A. Folch, L. Mingari, G. Macedonio, A. Costa, Long-term hazard assessment of explosive eruptions at Jan Mayen (Norway) and implications for air traffic in the North Atlantic, Nat. Hazards Earth Syst. Sci. 22 (1) (2022) 139–163, http://dx.doi.org/10.5194/nhess-22-139-2022.
[54] B. Martínez, M. Titos, L. Sandri, O. Rudyy, A. Cheptsov, G. Macedonio, A. Folch, S. Barsotti, J. Selva, A. Costa, On the feasibility and usefulness of High Performance Computing in Probabilistic Volcanic Hazard Assessment: an application to tephra hazard from Campi Flegrei, Front. Earth Sci. 10 (2022) http://dx.doi.org/10.3389/feart.2022.941789.
[55] S. Gibbons, S. Lorito, J. Macías, F. Løvholt, J. Selva, M. Volpe, C. Sánchez-Linares, A. Babeyko, B. Brizuela, A. Cirella, M. Castro, M. de la Asunción, P. Lanucara, S. Glimsdal, M. Lorenzino, M. Nazaria, L. Pizzimenti, F. Romano, A. Scala, R. Tonini, J. González Vida, M. Vöge, Probabilistic tsunami hazard analysis: High performance computing for massive scale inundation simulations, Front. Earth Sci. 8 (2020) http://dx.doi.org/10.3389/feart.2020.591549.
[56] F. Løvholt, S. Glimsdal, C. Harbitz, On the landslide tsunami uncertainty and hazard, Landslides 17 (2020) 2301–2315, http://dx.doi.org/10.1007/s10346-020-01429-z.
[57] S.J. Gibbons, S. Lorito, M. de la Asunción, M. Volpe, J. Selva, J. Macías, C. Sánchez-Linares, B. Brizuela, M. Vöge, R. Tonini, P. Lanucara, S. Glimsdal, F. Romano, J.C. Meyer, F. Løvholt, The sensitivity of tsunami impact to earthquake source parameters and manning friction in high-resolution inundation simulations, Front. Earth Sci. 9 (2022) http://dx.doi.org/10.3389/feart.2021.757618.
[58] R. Basili, et al., The making of the NEAM tsunami hazard model 2018 (NEAMTHM18), Front. Earth Sci. 8 (2020) 616594, http://dx.doi.org/10.3389/feart.2020.616594.
[59] R. Tonini, P.D. Manna, S. Lorito, J. Selva, M. Volpe, F. Romano, R. Basili, B. Brizuela, M.J. Castro, M. de la Asunción, D.D. Bucci, M. Dolce, A. Garcia, S.J. Gibbons, S. Glimsdal, J.M. González-Vida, F. Løvholt, J. Macías, A. Piatanesi, L. Pizzimenti, C. Sánchez-Linares, E. Vittori, Testing tsunami inundation maps for evacuation planning in Italy, Front. Earth Sci. (9) (2021) 628061, http://dx.doi.org/10.3389/feart.2021.628061.
[60] J. Selva, S. Lorito, M. Volpe, F. Romano, R. Tonini, P. Perfetti, F. Bernardi, M. Taroni, A. Scala, A. Babeyko, F. Løvholt, S.J. Gibbons, J. Macias, M.J. Castro, J.M.G. Vida, C. Sánchez-Linares, H.B. Bayraktar, R. Basili, F.E. Maesano, M.M. Tiberti, F.M. Mele, A. Piatanesi, A. Amato, Probabilistic tsunami forecasting for early warning, Nat. Commun. 12 (2021) http://dx.doi.org/10.1038/s41467-021-25815-w.
[61] M. Taroni, J. Selva, A testable worldwide earthquake faulting mechanism model, Seismol. Res. Lett. 92 (6) (2021) 3577–3585, http://dx.doi.org/10.1785/0220200445.
[62] L. Krischer, A. Fichtner, S. Zukauskaite, H. Igel, Large-scale seismic inversion framework, Seismol. Res. Lett. 86 (4) (2015) 1198–1207, http://dx.doi.org/10.1785/0220140248.
[63] V. Monteiller, S. Beller, B. Plazolles, S. Chevrot, On the validity of the planar wave approximation to compute synthetic seismograms of teleseismic body waves in a 3-D regional model, Geophys. J. Int. 224 (3) (2020) 2060–2076, http://dx.doi.org/10.1093/gji/ggaa570.
[64] A. Hoffmann, V. Monteiller, C. Bellis, A penalty-free approach to PDE constrained optimization: application to an inverse wave problem, Inverse Problems 37 (5) (2021) 055002, http://dx.doi.org/10.1088/1361-6420/abe4a9.
[65] N. Poiata, J. Conejero, R. Badia, J.-P. Vilotte, Data-streaming workflow for seismic source location with PyCOMPSs parallel computational framework, in: EGU General Assembly, 2021, http://dx.doi.org/10.5194/egusphere-egu21-15141.
[66] N. Poiata, C. Satriano, J.-P. Vilotte, P. Bernard, K. Obara, Multiband array detection and location of seismic sources recorded by dense seismic networks, Geophys. J. Int. 205 (3) (2016) 1548–1573, http://dx.doi.org/10.1093/gji/ggw071.
[67] E. Tejedor, Y. Becerra, G. Alomar, A. Queralt, R.M. Badia, J. Torres, T. Cortes, J. Labarta, PyCOMPSs: Parallel computational workflows in Python, Int. J. High Perform. Comput. Appl. 31 (1) (2017) 66–82, http://dx.doi.org/10.1177/1094342015594678.
[68] N. Poiata, J.-P. Vilotte, P. Bernard, C. Satriano, K. Obara, Imaging different components of a tectonic tremor sequence in southwestern Japan using an automatic statistical detection and location method, Geophys. J. Int. 213 (3) (2018) 2193–2213, http://dx.doi.org/10.1093/gji/ggy070.
[69] A. Fournier, J. Aubert, V. Lesur, E. Thébault, Physics-based secular variation candidate models for the IGRF, Earth, Planets Space 73 (2021) http://dx.doi.org/10.1186/s40623-021-01507-z.
[70] P. Arneitz, R. Leonhardt, E. Schnepp, B. Heilig, F. Mayrhofer, P. Kovacs, P. Hejda, F. Valach, G. Vadasz, C. Hammerl, R. Egli, K. Fabian, N. Kompein, The HISTMAG database: combining historical, archaeomagnetic and volcanic data, Geophys. J. Int. 210 (3) (2017) 1347–1359, http://dx.doi.org/10.1093/gji/ggx245.
[71] N. Schaeffer, A. Fournier, G. Marin, J. Calderón, Geomagnetic reversal from a numerical simulation, 2022, http://dx.doi.org/10.6084/m9.figshare.16825555.v1.
[72] N. Schaeffer, A. Fournier, Magnetic reversal in a geodynamo simulation, 2022, http://dx.doi.org/10.6084/m9.figshare.14847429.v1.
[73] S. Osores, J. Ruiz, A. Folch, E. Collini, Volcanic ash forecast using ensemble-based data assimilation: an ensemble transform Kalman filter coupled with the FALL3D-7.2 model (ETKF–FALL3D version 1.0), Geosci. Model Dev. 13 (2020) 1–22, http://dx.doi.org/10.5194/gmd-13-1-2020.
[74] L. Mingari, A. Folch, A.T. Prata, F. Pardini, G. Macedonio, A. Costa, Data assimilation of volcanic aerosol observations using FALL3D+PDAF, Atmos. Chem. Phys. 22 (3) (2022) 1773–1792, http://dx.doi.org/10.5194/acp-22-1773-2022.
[75] A. Folch, L. Mingari, A.T. Prata, Ensemble-based forecast of volcanic clouds using FALL3D-8.1, Front. Earth Sci. 9 (2022) http://dx.doi.org/10.3389/feart.2021.741841.
[76] M. Wilkinson, M. Dumontier, I. Aalbersberg, et al., The FAIR Guiding Principles for scientific data management and stewardship, Sci. Data 3 (2016) 160018, http://dx.doi.org/10.1038/sdata.2016.18.
[77] R. Borrell, J. Cajas, D. Mira, A. Taha, S. Koric, M. Vázquez, G. Houzeaux, Parallel mesh partitioning based on space filling curves, Comput. & Fluids 173 (2018) 264–272, http://dx.doi.org/10.1016/j.compfluid.2018.01.040.
[78] E. Boman, U. Çatalyürek, C. Chevalier, K. Devine, The Zoltan and Isorropia parallel toolkits for combinatorial scientific computing: Partitioning, ordering and coloring, Sci. Program. 20 (2012) 713587, http://dx.doi.org/10.3233/SPR-2012-0342.
[79] A. Cheptsov, C. Niethammer, WMS-light, 2022, http://dx.doi.org/10.5281/zenodo.6340199.
[80] J. Ejarque, R.M. Badia, L. Albertin, G. Aloisio, E. Baglione, Y. Becerra, S. Boschert, J.R. Berlin, A. D'Anca, D. Elia, F. Exertier, S. Fiore, J. Flich, A. Folch, S.J. Gibbons, N. Koldunov, F. Lordan, S. Lorito, F. Løvholt, J. Macías, F. Marozzo, A. Michelini, M. Monterrubio-Velasco, M. Pienkowska, J. de la Puente, A. Queralt, E.S. Quintana-Ortí, J.E. Rodríguez, F. Romano, R. Rossi, J. Rybicki, M. Kupczyk, J. Selva, D. Talia, R. Tonini, P. Trunfio, M. Volpe, Enabling dynamic and intelligent workflows for HPC, data analytics, and AI convergence, Future Gener. Comput. Syst. 134 (2022) 414–429, http://dx.doi.org/10.1016/j.future.2022.04.014.
[81] J. Tromp, S. Chevrot, Dimitri Komatitsch (1970–2019), 2019, http://dx.doi.org/10.1029/2019EO118653.

Arnau Folch holds a BS in Physics at the University of Barcelona (UB) and a Ph.D. on applied Mathematics at the Technical University of Catalonia (UPC).He is a research Professor at Geociencias Barcelona (GEO3BCN-CSIC) and visitor fellow at the Barcelona Supercomputing Center (BSC), Spain Author of over 100 refereed publicationsin journal papers, he has participated 45 national and European competitive research projects and in multiple contracts with private companies. Coordinator of the EuroHPC Center of Excellence for Exascale in Solid Earth (ChEESE) and of the DT-GEO project. Member of the Strategic Advisory Board of the Destination Earth initiative Having access to some of the most powerful supercomputing facilities in Europe, he has developed a wide expertise in optimised computing strategiesand in modelling of natural hazards.