D5.1 Pilot analysis and development strategy

Version 1.3

Document Information

Document date	24/04/2020
Contract Number	823844
Project Website	https://cheese-coe.eu/
Contractual Deadline	30/04/2020 (M18)
Dissemination Level	CONFIDENTIAL
Nature	Report
Author	Tomaso Esposti Ongaro
Contributors	Marta Pienkowska and Josep De la Puente (PD1), Jorge Macias (PD2), Alice Gabriel (PD5), Laura Sandri and Sara Barsotti (PD6), Finn Løvholt (PD7), Stefano Lorito, Manuela Volpe and (PD8), Vadim Monteiller (PD9), Arnau Folch (PD12), Joan Farnos (WP6), Josep De La Puente (WP4), and PD teams.

Reviewers	Piero Lanucara (CINECA), Giovanni Macedonio (INGV)

The ChEESE project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No 823844

Change Log

Version	Description of Change
v1.0	Initial draft for internal working
v1.1	Reviewed by the Project Coordinator
v1.2	Reviewed by Pilot leaders
v1.3	Reviewed by WP5 coordinator and sent out to internal reviewers
v1.4	Final review by WP5 coordinator, passed to project coordinator and project manager

Index

1. Introduction ... 6
2. From Pilot Demonstrators to Services ... 6
2.1 Definition of Technology Readiness Levels in ChEESE ... 6
2.2 Involvement and role of the end-users ... 8
3. Service typologies ... 8
3.1 Urgent Computing ... 9
3.2 Probabilistic Hazard Assessment ... 9
3.3 Early Warning ... 10
3.4 Imaging of the subsurface ... 11
3.5 Services to the scientific community ... 11
4. Data, Metadata and Model access policies ... 11
Data Taxonomy ... 12
Metadata model ... 12
Visions and Solutions ... 13
5. Ethical and liability issues ... 14
6. PD sheets ... 15
PD1. Urgent Seismic Computing ... 15
Service definition ... 15
Potential target users ... 17
System design and architecture ... 18
Environment ... 20
Computational resources ... 20
PD2. Faster-Than-Real-Time tsunami simulation ... 21
Service definition ... 21
Potential target users ... 22
Service design and architecture ... 24
Environments ... 25
Computational Resources ... 25
PD5. Physics-based probabilistic seismic hazard assessment ... 27
Service definition ... 27
Potential target users ... 28
System design and architecture ... 28
Environment ... 30

Computational Resources 30
PD6. Probabilistic volcanic hazard assessment 31
Service definition 31
Potential target user 33
System design and architecture 34
Environment 37
Computational Resources 38
PD7. Probabilistic tsunami hazard assessment 39
Service definition 39
Potential target users 40
System design and architecture 41
Environment 42
Computational Resources 42
PD8: Probabilistic Tsunami Forecast (PTF) for early warning and rapid post event assessment 44
Service definition 44
Potential target users 45
System design and architecture 46
Environments 49
Computational resources 50
PD9. Seismic tomography 52
Service definition 52
Potential target users 52
Service design and architecture 53
Environment 54
Computational Resources 54
PD12: High-resolution volcanic ash dispersal forecast 55
Service definition 55
Potential target users 55
Service design and architecture 56
Environment 58
Computational Resources 58
A1. Table 1. ChEESE Technology Readiness Levels 59
A2. Table 2. IUB interest in ChEESE Services 60

1. Introduction

One of the main objectives of WP5 is to coordinate the effort to make Pilot Demonstrators (PDs), developed in WP4, available as Services exploitable by a broader user community, represented by the ChEESE Industrial and User Board (IUB). As a measure of this effort, we aim to advance above 4 the Technology Readiness Level (TRL) of 8 of the 12 Pilot Demonstrators in ChEESE.

To achieve this goal, we involve (in collaboration with WP6) the geophysical community (in coordination and synergy with other pan-European initiatives) and non-academic stakeholders (industrial partners, observatories, civil protection authorities) belonging to the IUB, in the proper definition of the Services and in the Validation process, addressing realistic geophysical Use-Cases, especially focusing on hazard and risk assessment goals.

This document reports about the strategy to develop ChEESE Services:

- In Section 2, a description of the target Services, and the roadmap to co-design them through a synergy between code developers, HPC centres, and the IUB is described. Moreover, the TRL scale, appropriately adapted to the ChEESE Services and products is presented.
- Section 3 presents the three main Service prototypes addressed in ChEESE.
- To make Services available to a broad community, a preliminary assessment of Data, Workflows and Model access issues, coherent with policies already established in the field of Solid Earth research infrastructures and science products (in particular, in the framework of the EPOS Implementation Phase: https://www.epos-ip.org/about/epos-implementation-phase/epos-implementation-phase-project), is addressed in Section 4.
- Section 5 summarizes about the strategy to address the main ethical and liability issues raised by the development of ChEESE HPC Services.
- Section 6 reports the details of each PD development strategy, with the preliminary definition of Use-Cases, Functional Requirements and Validation criteria and methodology. A preliminary estimate of the required HPC resources needed and policy issues foreseen to deploy operational Services in the future is finally presented.

2. From Pilot Demonstrators to Services

Pilot Demonstrators designed in WP4 aim at demonstrating that HPC codes developed and optimized in WP2, together with High Performance Workflows for data and model integration, developed and optimized in WP3, can be exploited to address challenging scientific problems in the field of Solid Earth sciences. WP5 aims at demonstrating that such applications have an added value for the scientific community and for the whole society, in terms of their potential contribution to hazard assessment and risk mitigation, and a potentially disruptive economic impact. Eight of the twelve Pilot Demonstrators (those characterized by a target Technology Readiness Level above 4) are brought into WP5.

2.1 Definition of Technology Readiness Levels in ChEESE

Technology Readiness Levels (TRL) is a type of measurement system used to assess the maturity levels of a particular technology. Their definition in Horizon 2020 (https://ec.europa.eu/research/participants/data/ref/h2020/wp/2014_2015/annexes/h2020-wp1415-annex-g-trl_en.pdf) is necessarily rather general, as it has to be applied to different fields of science and technology. In ChEESE, we have adapted the TRL scale to the specific framework of the development of Exascale HPC services based on computational geophysical models and related workflows (Table 1).

Environments

The ChEESE TRL scale is organized so that different levels are associated to 4 different Environments:

Environment	TRL
Non-HPC laboratory environment	1-2
(Pre-) Exascale laboratory HPC environment	3-4
HPC (relevant) production environment	5-6
Operational environment	7-8-9

ChEESE definition keeps the original distinction between laboratory (where technology is developed), relevant (where technology is tested), and operational (where technology is applied) environments. However, for ChEESE Pilots, the main difference between the three different environments will concern the target users, the service interface, and the (data/software/workflow/products) access policy.

Typical Operational Environments in ChEESEs include:

- Volcano Observatories
- Earthquake and Tsunami Warning Centres
- Civil Protection Operative Centres
- Governmental Agencies
- Private companies (insurance, geo-resources, risk mitigation)

We also stress the distinction between Operational Environment and Operational Service. To implement an Operational Service (i.e., a Service used to take actions), an agreement should be formalized between the developers, or the Centre of Excellence, and users on the Ethical and Liability issues. This characterizes TRL 9 products. None of the developing HPC services is aiming at TRL 9, i.e., no HPC service will be operational in this sense in ChEESE.

Components and Workflows

A component is each individual step of a workflow (e.g., the simulation software, data conversion algorithms, pre-processing and post-processing procedures, data transfer, etc.), including the HPC Exascale-enabled codes developed under ChEESE. In each of the 4 environments, the difference between lower and higher TRL relates mostly to the differing accessibility to individual components (lower) and workflows (higher).

- ***TRL 1 – 2.*** This is a preliminary stage of model formulation and development of parallel HPC codes. All Flagship codes in ChEESE are already parallelized and have features and capability of running on parallel machines. Therefore, no ChEESE product has TRL 1. At TRL 2, the HPC performance of all ChEESE numerical codes have been assessed by means of appropriate Code Audit (WP2).
- ***TRL 3 – 4.*** Optimizing the performances of the numerical flagship codes and workflows on pre-Exascale platforms are the main objectives of ChEESE Work Packages WP2 (flagship codes and procedures) and WP3 (workflows). Successful achievement of this TRL will be demonstrated, in the laboratory, against target benchmarks and ChEESE Pilot Demonstrators (WP4). Target users are limited to the specialists who have developed the codes and workflows.
- ***TRL 5 – 6.*** At these levels, codes and workflows will be engineered to solve scientific problems of general scientific interest (Pilot Demonstrators of WP4) but the target user would potentially include application specialists (i.e., belonging to the broader scientific community represented in the IUB), not necessarily involved in model and workflow development.
- ***TRL 7 – 8.*** At these levels, codes and workflows will be engineered as HPC Services to solve problems of interest for the end-users. Target users are in general non-specialists in the field of computer science or geophysics. For validation of the service, pre-defined functional requirements specified by the stakeholders must be met.
- ***TRL 9.*** As stated above, at TRL 9 Service are implemented to be Operational (i.e., they are used to take actions). An agreement should be formalized between the developers, or the Centre of Excellence, and users on the Ethical and Liability issues.

Table 1 summarizes the main features of the ChEESE TRL scale.

2.2 Involvement and role of the end-users

In the transition from TRL<=4 (laboratory) to TRL>=5 (production), end-users are involved in the co-design of Services by defining, in synergy with the ChEESE teams and the PD leaders, the appropriate use-cases and requirements for the validation of the Pilot Demonstrators.

Deliverable D6.6 (Exploitation Plan) reports in detail the methodology adopted to involve the IUB members in the definition and potential exploitation of the Services. Table 2 in the Appendix summarizes the interests expressed in participation to the different Pilot Demonstrators. In Section 6, the specific involvement of IUB into Service development is reported for each Pilot Demonstrator.

3. Service typologies

Development of HPC Services poses several new challenges, both from a scientific-technological point of view and for what concerns implementation policies. ChEESE teams are addressing scientific-technological issues within WP2, WP3 and WP4, and are working with IUB in the definition and co-design of services in WP5. Preliminary considerations about policy issues are reported in Section 5.

We here summarize the main features of service design. The first three typologies can be implemented individually (as in the case of post-event Urgent Seismic computing for seismic risk management) or seen as components of compound procedures for supporting decision making for natural hazards, as in the case of Early Warning systems for volcanic eruptions or tsunami. The fourth typology (subsurface imaging) is not directly aimed at hazard assessment but it is a fundamental component of many Solid Earth science services. Service to the scientific community are finally outlined.

3.1 Urgent Computing

Urgent Computing (UC) (or Urgent Super-Computing, USC) is the procedure to obtain on demand very fast solutions in HPC applications where time is critical to accelerate the rates at which information is gleaned from the full cycle of data use, distilled and collated into forms that will be used for decision making. This involves, in particular, the automation of the different stages of these workflows, including data processing and ingestion, computing and result analysis mechanisms. In the case of very large scale simulations, we use the term urgent supercomputing, which further requires the establishment of specific services, resource management and system policies to enable efficient and time-responsive solutions in Tier-0 or Tier-1 supercomputing infrastructures for scenarios and workflow patterns of high socio-economic impact.

The development of operational UC services requires the identification of pathways for co-designing and co-implementing a list of data, compute, network services, together with resource management and access policy. ChEESE would consider "urgent" simulations as possible use cases within a testing phase to identify and assess the required services, resource management and policy access in relation with different workflow patterns and data logistics while checking the feasibility of effectively contributing to future emergencies.

From a scientific and technological point of view, UC poses several new challenges:

- Efficient data management and logistics
- HPC Efficiency on heterogeneous Exascale architectures
    - of numerical solvers: improving domain size, resolution, accuracy
    - of workflows: complex pre- and post-processing, model-data interoperability (including data-assimilation)
    - uncertainty management
- Workflow Automation, improved resource management and scheduling
- Redundancy
- Fault tolerance

3.2 Probabilistic Hazard Assessment

In the Solid Earth, processes leading to hazardous events are characterized by broad variability in magnitude, intensity, temporal/spatial occurrence and local geological conditions, and by extreme unpredictability. Among these processes, those leading to catastrophic events are particularly challenging for risk managers due to their low probability of occurrence but extremely severe impact (https://www.undrr.org/publication/sendai-framework-disaster-risk-reduction-2015-2030). To face the wide range of variability, scenario-based hazard and risk assessment strategies can be adopted, in which a relatively small number of representative event sizes and hazard development conditions are analysed, by means of numerical models, and used to eventually mitigate the risk.

A different strategy, promoted by several Pilot Demonstrators in ChEESE, is to consider the broad variability of hazardous events, together with their probability of occurrence, and perform large ensemble simulations. In Solid Earth sciences, we usually distinguish between short- and long-term Probabilistic Hazard Assessment (PHA), the ultimate difference being the possibility to integrate real-time monitoring data in the probability estimates for short-term analysis. The final product of PHA differs from that of scenario-based simulations, being provided in probabilistic terms, e.g. as a local probability of exceedance of a given, or a set of, hazard threshold.

Although there may be exceptions (e.g., in the case of short-term PHA for airborne volcanic ash concentration), PHA for Solid Earth phenomena usually does not have strict time constraints (i.e., it is a tool for planning, not for urgent action); however, it requires a careful accounting of all numerous sources of uncertainties, including model-related ones (e.g., parameter variability) and those related to epistemic and aleatoric uncertainty on initial and boundary conditions. Quantification of uncertainty is a fundamental step to PHA, and for decision-making in risk management processes.

From the point of view of ChEESE technological development and for Service deployment, Probabilistic Hazard Assessment poses several new challenges:

- HPC Efficiency
    - of numerical solvers: improving domain size, resolution, accuracy
    - of workflows: complex pre- and post-processing, model-data interoperability
- Managing of a large number of scenarios (large ensemble size).
- Large amount of data to store, post-process and analyse.

3.3 Early Warning

Early Warning (EW) Services rely on routine forecasting procedures run in Monitoring/Warning Centres and Observatories, having a formal responsibility of issuing a warning in case of triggering events. Depending on the cases, these can be recognition of unrest signals (e.g., for volcanic eruption) or an earthquake occurrence (e.g., for tsunami). Warning centres are in charge of reporting to decision-makers about potential impact scenarios, in either a deterministic or probabilistic framework (or a mixed one). EW also more strongly rely on quality and robustness of real-time data. Because they are associated to contingencies, they usually have strict time constraints (“early”). EW somehow merges the technological requirements of Urgent Computing and Probabilistic Hazard Assessment and poses several new HPC challenges:

- Efficient data management and logistics
- HPC Efficiency on heterogeneous Exascale architectures
    - of numerical solvers: improving domain size, resolution, accuracy
    - of workflows: complex pre- and post-processing, model-data interoperability (including data-assimilation)

    - uncertainty management
- Managing of a large number of scenarios (large ensemble size).
- Workflow Automation, improved resource management and scheduling
- Redundancy
- Fault tolerance

3.4 Imaging of the subsurface

Images of the inner structure of the lithosphere are the fundamental input data for many Solid Earth models, prerequisites for modelling volcanic systems or earthquake propagation. In addition, they are of paramount importance for exploitation of geo-resources. Images of the subsurface structure are obtained by solving an inverse problem, in which the difference between measured data and simulated signal at the surface is optimized by some minimization algorithm. In ChEESE, focus is on Seismic Tomography, i.e., seismic travel time data are compared to an initial Earth model and the model is modified until the best possible fit between the model predictions and observed data is found. Solving inverse problems poses severe computational challenges because the resolution of the image depends on the frequency of data and spatial-temporal resolution of forward numerical simulations used in the minimization algorithm. HPC challenges thus include:

- Efficient data management and logistics
- HPC Efficiency on heterogeneous Exascale architectures
    - of numerical solvers: improving domain size, resolution, accuracy
    - of workflows: complex pre- and post-processing, model-data interoperability
- Managing of a large number of simulations (to ensure convergence of the inverse problem).
- Large datasets to store, post-process and analyse.

3.5 Services to the scientific community

Services to the scientific community include accessibility to model and workflows with accessible pre- and post-processing, accessible databases of numerical results, development of community models and support to users. Such services will be better defined in the second half of the project, and will also depend on the access policies that will be adopted by the CoE and by the individual PD teams.

4. Data, Metadata and Model access policies

The Data Management Plan (DMP) has been set up in WP1 (Deliverable 1.2). The DMP describes the life cycle for all data sets that will be collected, processed or generated by the research project. It is a document outlining how research data will be handled during a research project, and even after the project is completed, describing what data will be collected, processed or generated and following what methodology and standards, whether and how this data will be shared and/or made open, and how it will be curated and preserved.

We here report about the strategy to implement an access policy compliant with the requirements of the EPOS-ERIC (Earth Plate Observing System - European Infrastructure

Consortium). EPOS has been working in its Implementation Plan to structure the scientific and technological community, to harmonize data policies and promoting interoperability, and to collect information about European Research Infrastructures in the field of Solid Earth Sciences, developing a federated governance model into the ERIC. Moreover, it has developed a novel e-infrastructure: The Central Hub of the Integrated Core Services. The current level of provision of software by EPOS is minor. Moreover, it has to be specified that EPOS will not provide resources for storing data or software, which remains in the duties of the community (the Thematic Core Services, or TCS). In this sense, EPOS acts through a federated governance model.

The strategy to integrate ChEESE products within the EPOS ecosystem has been discussed with EPOS coordinator and development Team. It will be based on three main points:

1. Fit EPOS data taxonomy
2. Fit EPOS Metadata model
3. Share visions and solutions:
    - to provide Workflows as Services
    - about sustainability
    - about the societal impact of Science

Data Taxonomy

In order to facilitate the integration among the communities, EPOS data had been categorized in the following levels:

- **Level 0**: raw data, or basic data
- **Level 1**: data products coming from nearly automated procedures
- **Level 2**: data products resulting by scientists' investigations
- **Level 3**: integrated data products coming from complex analyses or community shared products

As a first step of our strategy, data used and produced under ChEESE will be organized according to the EPOS categories, and keeping the focus on interoperability among data and between models and data. EPOS TCS representatives will be contacted by ChEESE to meet community standards on formats and to guarantee FAIRness of data (https://www.go-fair.org/).

Metadata model

EPOS maintains a Metadata catalog (https://epos-ip.org/what-metadata-epos), which reflects accurately the assets (such as datasets, or software or workflows/services or access to equipment) offered by the Thematic Core Services (TCS) as part of the overall EPOS system. The catalog will be used: (a) to provide the attributes for the role-based security system, commonly referred to as AAAI (Authentication, Authorisation, Accounting Infrastructure); (b) to answer queries on multiple attributes to return the metadata records describing assets in the TCSs. It also provides an inventory of assets so that EPOS management can demonstrate to stakeholders that EPOS has a valuable collection of assets and the means to provide access to them.

ChEESE data and products will be accompanied by appropriate Metadata, according to technical specifics set by EPOS. By this way, ChEESE repositories (for software and data products) will be findable by appropriate API to EPOS-ICS (Integrated Core Services), including the Web repository for codes and toolkits (which is not described here and will be detailed in future ChEESE Deliverable 4.13). EPOS adopts a minimum set of common metadata elements required to operate the ICS level, taking into consideration the heterogeneity of the many TCSs involved in EPOS.

ChEESE Pilot Demonstrator Teams will work in strict connection with the EPOS Team to implement appropriate Metadata model. This will be tested during the PD Validation stage (Deliverable D5.4).

Visions and Solutions

ChEESE and EPOS will share visions and solutions about the following key aspects of their development strategy. The publication of a joint position paper has been planned.

Workflows as Services

There is no current engagement of EPOS to provide Workflow as a Service (WaS). At the same time, ChEESE is aiming at developing prototype services using HPC Workflows, with massive use of computational resources and intense data access. Different models of Virtual Research Environments (VRE) can be conceived, either aiming at setting up collaborative framework or managing WAS, with virtual interface to models, input and output data. The latter requires a high level of commitment. To go beyond a prototype model, the *sustainability* model and *ethical issues* has to be carefully discussed in the scientific community, and it has to be designed with the stakeholders.

Sustainability

In providing access to data and products, EPOS is not taking responsibility for providing computational resources. The products are produced by the community, with their own processing chain. It should be discussed whether future CoE engagement will be aimed at providing access to computational resources to EPOS TCS, with other actors potentially involved in this process (e.g., the European Open Science Cloud – EOSC), or other forms of procurement of resources can be envisaged. ChEESE will share with EPOS its data preservation strategy, i.e., whether models or products will be preserved. Providing persistent access to large databases or to complex model workflows implies indeed defining a sustainability plan, and completely different business models, that are being discussed in ChEESE exploitation plan.

Ethical issues and societal impact of Science

EPOS is discussing about involving the Hazard and Risk Management community as EPOS TCS. This would help discussing societal implication of research products, from an ethical point of view. ChEESE is aiming at computing hazard and emergency management products. While sticking to principle of Open-Science, providing open access to these services might be in some cases an issue (see Section 5).

5. Ethical and liability issues

Strictly speaking (according to H2020 manual on Ethics), there are no ethical issues to be addressed in developing ChEESE services. However, some critical issues involving some ethical concepts concern the impact that scientific research might have in decision-making process and in public perception of natural risk. Such issues will be preliminary addressed by ChEESE, although services will be developed as prototypes or demonstrators.

For Urgent Computing, as for Early Warning services, urgency or earliness is determined by a specific commitment to deliver a shared scientific product to an authoritative stakeholder, distilled and collated into forms that can be used for decision making, through a given protocol and within recognized procedures (targeted to emergency management). National/regional authorities and authoritative scientific organizations or international bodies are usually in charge of managing the emergency following shared protocols. The *triggering* of UC and EW procedures is a complex task needing clear identification of the actors driving such process and the responsibilities in the entire workflow deployment. Concerning the release of final products, some scientific results and representation (e.g., probabilistic hazard maps or curves for PHA) need an expert audience to be correctly interpreted and can generate either undue alarms or a false sense of security. In addition, results developed in a Bayesian framework are subject to continuous update and need to be conveyed with the correct language. Liability in the decision-making process, when informed by scientific products, has to be rigorously established before implementation of operational services. Uncertainty associated to hazard products has to be robustly quantified and properly communicated.

All PD leaders are or will discuss with IUB members (in particular, with decision-makers) potential critical policy issues, which will need to be addressed in the perspective of a future deployment of operational services. Some of these issues will be reported in more detail in D5.3 and D5.4 at the end of the project.

6. PD sheets

This section describes the strategy to develop HPC Services for 8 Pilot Demonstrators, targeting TRL 5-8. In particular, it reports the target Use Cases and preliminary Functional Requirements to be met for Validation, which have been in most cases agreed with the end-users in the IUB. This description is complemented by information collected in the ChEESE Exploitation plan (Deliverable 6.6, 30 April 2020).

PD1. Urgent Seismic Computing

PD1	Urgent Seismic Computing
Leader	ETH
Participants	BSC, INGV, LMU, TUM, IMO
Codes	ExaHyPE, Salvus, SPECFEM3D
TRL initial	3
TRL target	5-6

Service definition

Service type. The proposed service targets the impact assessment in the immediate aftermath of an earthquake (Urgent Computing).

Objectives. Deadly earthquakes are unpredictable and relatively rare, but have a high socioeconomic impact. Moreover, the impact itself is difficult to estimate, as every event has specific characteristics (location, magnitude, directivity) which, together with local amplification and de-amplification effects due to geological structures, make their outcome very singular. Mitigating and managing impacts of such deadly earthquakes is of paramount importance. A priori studies attempt to understand the local conditions (geology, fault system) and derive information from past events to make long term risk assessments. A posteriori studies, on the other hand, take place immediately after an earthquake in order to make early assessments of its impact.

Scientific and technological advancements. At present, empirical relations derived from historical events are the principal methodology used to make such early assessments. The approaches, however, face a serious limitation - with large earthquakes being so rare we often lack sufficient data to constrain the relations, rendering them imprecise. Physics-based numerical simulations are powerful tools that can provide highly accurate shaking information. However, seismic wave propagation is currently computationally prohibitive at high frequencies relevant for earthquake engineering risk assessments. Moreover, the simulations require fine tuning of the parameters, as uncertainties in the geological model and in earthquake source information translate into uncertainties in final results.

HPC added value. Developments of HPC infrastructures will render routine executions of such simulations possible, enabling new approaches to assess seismic hazard. Urgent seismic simulations within hours after an event could potentially deliver more accurate short-time reports of the consequences of moderate to large earthquakes.

The proposed service is based around a novel HPC urgent seismic simulation workflow under development within PD1 (Figure 1). The workflow starts with an alert system that monitors earthquakes and triggers an alert should a high-risk even occur. It then estimates source parameters and defines other simulation inputs, and finally it launches a deterministic simulation of the propagation of seismic waves in the vicinity of the earthquake (Figure 2).

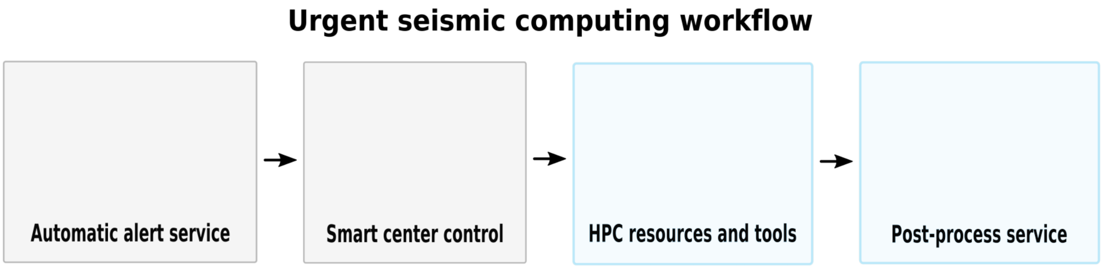

Figure 1 Conceptual diagram of the Urgent Computing Integrated Services for EarthQuakes (UCIS4EQ) workflow developed in PD1. UCIS4EQ is composed of four main components: Automatic alert service (scanning earthquake information in real time), Smart Centre control (collating source information and generating simulation inputs), HPC facilities (running the wave propagation simulation), and Post-process.

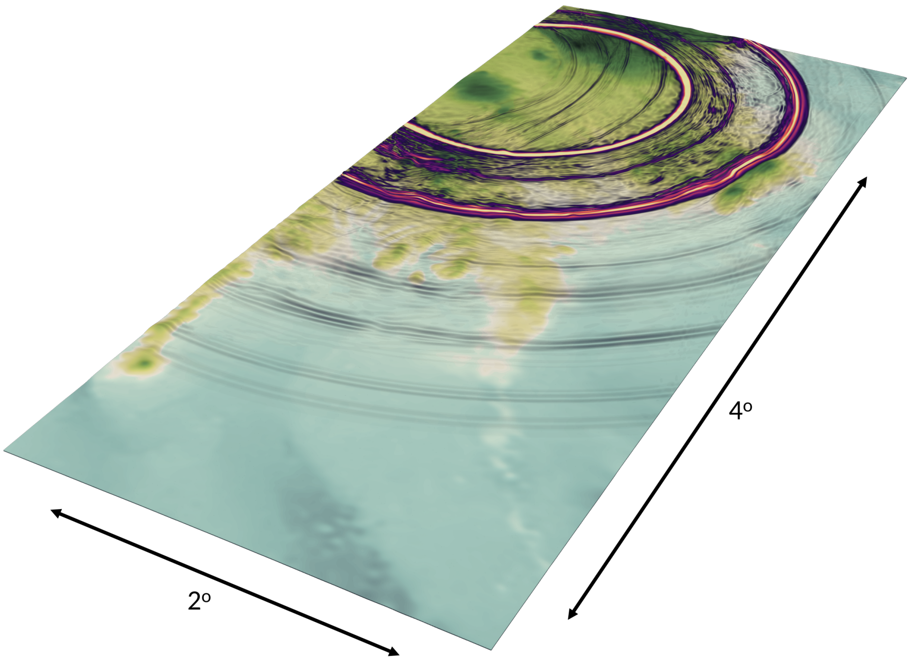

Figure 2. A 3D Wavefield snapshot of a HPC simulation in South-West Iceland. The model uses a 1-D velocity model of the subsurface. This particular test simulation was run on 200 GPU nodes (2400 cores) on PizDaint (CSCS) and took a total of 42 minutes. The simulation resolved frequencies up to 2.66 Hz.

The main objective is to produce physics-based ground shaking maps and other information potentially useful for rapid impact assessments.

Impact. An Urgent Seismic Computing protocol (USCp) - a technical guide to executing seismic simulations in a HPC environment considering strong time constraints - is being developed parallel to the workflow itself in the context of PD1. An urgent computing application must manage complex interactions between users, computational resources, elevated executed priority policies, and working sessions. HPC infrastructures currently provide virtually no support for urgent computing, and relevant protocols are crucial for the functioning of a future urgent seismic simulations service.

Potential target users

Within the ChEESE IUB, members that have expressed the most interest in future service deployment are:

1. The Icelandic Civil Protection and Icelandic Meteorological Office (IMO).
2. ARISTOTLE European Project
3. Global Parametrics Holding Company Limited (GP)

With the intended final TRL of the pilot at 5-6, the service developed within ChEESE will remain targeted at field specialists. The components of the service are under intensive development and real test cases are developed with IUBs for feasibility testing purposes. The demonstrator aims to test the future service in a controlled environment for specific use cases. It should be noted that all three IUB members above are willing to, up to a certain extent, be involved in the evaluation and/or development of PD1 associated services. In particular:

The **Icelandic institutions** are actively involved in the development of the pilot, and have been providing crucial information for a use case development in Iceland for civil protection purposes. The workflow is being tested on the South Iceland Seismic Zone (SISZ), and most of the work is focused on calibrating the models and understanding the benefits of deterministic simulations at high frequencies. Discussions are ongoing on future exploitation of potential services, with both urgent post-event simulations, as well as urgent simulations of large earthquake scenarios triggered by recorded swarms of smaller events on active faults (with such swarms treated as potential 'precursors' of larger events). The IMO has expressed interest in testing and running the workflow on their local infrastructures.

ARISTOTLE (All Risk Integrated System TOwards Trans-boundary hoListic Early-warning) is interested in exploring the possibility of integrating an urgent seismic computing service within the hazard-related services they provide to the EU Emergency Response Coordination Centre (ERCC). Aristotle aims to support the ERCC with regard to situation assessments in crisis situations, and urgent seismic simulations could contribute to the initial timely estimates of earthquake impact within hours after the event. Discussions with Aristotle have moved beyond preliminary stages towards a definition of a test case scenario in Turkey for one of the past events that Aristotle has tackled. We are waiting for further information from Aristotle, including a full report that has been generated for the chosen event, in order to define how the service could be exploited in the Aristotle framework. Aristotle expects a higher level of integration between simulation and data analysis for an applicable service.

Global Parametrics (GP) is developing parametric financing products which trigger a decision, financial in this case, based on crossing a predefined threshold of an index that captures frequency of catastrophic events, irrespective of the actual impact. Scientific and technological innovation has recently allowed for the experimentation with cross-boundary index-based products, such as forecast-based financing for weather-related events, that allows vulnerable communities to have access to financing even before the impact occurs. A similar solution is sought in the case of earthquakes, by introducing the concept of urgent financing, which would provide risk financing to affected communities within hours after the event occurs. It is envisioned that the output of urgent seismic simulations would be used to build indexes that would trigger payouts for products designed for earthquake-risk financing in vulnerable communities. Preliminary discussions revolve around a real-life test case in Mexico.

Moreover, we see the service as beneficial to both **scientific and HPC communities**. Developments towards the service include extensive testing of deterministic wave propagation in the context of seismic hazard assessment in new areas. Velocity models will be tested against recorded waveforms, and some new 3D velocity models of target areas might be proposed and shared with the scientific community. Furthermore, the developments in terms of earthquake simulation workflow might enable simpler interfaces for launching 3D simulations using HPC codes, by means of a unified translation of main earthquake characteristics to HPC code input files. The HPC community, on the other hand, will benefit from the developments on urgent computing protocols that are currently also discussed in the PRACE context (see e.g., https://prace-ri.eu/wp-content/uploads/WP243.pdf).

System design and architecture

Design and architecture of the service are discussed with potential users. The users, however, expressed distinct interests and therefore they have a range of requirements. We are currently working on use cases with each of the users, exploring possibilities in the exploitation of the proposed service beyond pure system benchmarking.

Use cases

Use case	Rationale	Summary
South Iceland Seismic Zone User: Icelandic Civil Protection	Strong motion data in Iceland is scarce and the empirical ground motion prediction equations (GMPEs) developed from data in other regions have been proven inadequate to represent Icelandic ground motions. Iceland is therefore interested to test a physics-based approach.	A domain has been defined around the SISZ, with receivers that correspond to locations of SIL stations (where seismic data since 1991 is available) and University EERC stations of the Icelandic Strong-motion Network (ISMN; where acceleration data is available). 1D and 3D velocity models have been defined, and example earthquakes have been chosen to calibrate the simulations and compare with recorded data. This scenario will be the first full-scale test of PD1.
Turkey User: ARISTOTLE	Aristotle has generated situation-assessment reports for some past earthquakes in Turkey, and given the high seismicity of the region suggested one of such past	The event has been selected by Aristotle, but the data has not been shared with ChEESE yet. Aristotle is waiting for permission from the Italian

	events as a test case of the service.	Civil Protection Department to share the information.
Mexico User: Global Parametrics	Introducing the concept of urgent financing for vulnerable communities in the context of seismic hazard. The Mexico use case has been suggested by GP given the interests of the Mexican government in the initial testing stages of such a financial product.	Detailed conversations with GP have started only recently, and a specific test region has not been selected yet. They are, however, well acquainted with PD1 developments due to an EuroLab4HPC project where they participated as external consultants.

Functional Requirements

Functional requirement	Rationale	Summary
Time to solution.	The time to solution is crucial in the context of urgent computing. Depending on use cases, together with users we will have to define the maximal time to solution of the wave propagation and of final product delivery. This will vary from a few hours (for cases like Aristotle), to a day (for urgent financing products). Critical factors are the highest frequency required, the amount of uncertainty analysis, access time to HPC resources and computational resources available.	The time to solution will most likely range from a few hours to a day or two, depending on the use case.
Resolution.	Resolution of simulations is related to the time to solution and the computational resources available, as longer time to solutions and more resources allow for higher frequency simulations. Resolution is a functional requirement, as frequency content of simulations has a strong influence on ground motion proxies (PGA, PGV, etc), and different buildings are sensitive to different frequencies of propagating waves. New, higher resolving, physical models of the subsurface are strong limitants to resolution, as is the lack of physical information on the shallowest layers of the Earth.	Overall the aim is to provide "high-frequency" deterministic results. On a regional scale, this means targeting above 1 Hz, with a maximum at approximately 10 Hz that is of interest for structural engineering (note that 10 Hz is currently not attainable with reasonable computational resources).

Number of simulations.	The number of required simulations within a given time might vary between cases, from single high-frequency deterministic runs, to tens of runs that better capture the uncertainty. In case we want to improve resilience of the system, we might duplicate runs and even split the runs in different HPC systems.	The number of simulations will range from a single simulation to an ensemble that is suitable for a statistical analysis.
Data formats.	Users may have different requirements on the outputs of the service, ranging from running the service themselves, obtaining only raw results, or expecting final processed ground motion intensity maps.	Data formats adopted by the end-users are currently unknown, but the service will provide processed ground motions.
Uncertainty quantification.	Specific treatment and representation of uncertainties might be required by different users.	Uncertainty quantification is work in progress, and it is hard to provide specific uncertainty treatments at this stage.

Environment

The service will be tested on realistic use cases, in relevant HPC environments: on Mare Nostrum at the BSC for the full workflow, as well as at PizDaint at CSCS for calibrating simulations. The validation on the selected use cases aims to prove the feasibility of the workflow, but for the time being the service is to be operated by experts. During the course of the validation the results will be shared with the users in order to better define final outputs and products.

Further development of the validation procedure (beyond ChEESE) includes generalizing the selected test cases to a range of regions and assuring the transportability of the workflow to other HPC clusters. IUB has expressed interest in hosting the service at their own compute servers.

Computational resources

In the particular case of PD1, more than in many other HPC applications, the more resources the better. Simulation HPC codes for earthquake simulations are amongst the most scalable software packages, hence it is easy to improve time-to-solution (critical for post-event damage assessment) by means of adding parallel compute power. In lack of precise numbers, which will be ready once our first demonstrations are run towards the end of the project, we can only assess the computational needs of this service, very generally, as “very high”.

PD2. Faster-Than-Real-Time tsunami simulation

PD2	Faster-Than-Real-Time tsunami simulation
Leader	UMA
Participants	BSC, INGV, TUM, NGI
Codes	Tsunami-HySEA, Landslide-HySEA
TRL initial	2
TRL target	6-7

Service definition

Service Type. The expected prototype service is a Faster Than Real Time (FTRT) environment for high-resolution simulations of earthquake generated tsunamis.

Objectives. The objective is developing a *FTRT Tsunami Simulation Service* that can produce many FTRT high-resolution tsunami simulations providing information about the propagation of the tsunami waves and their near-to-coast amplitude as well as high-resolution inundation maps in coastal areas over local high-resolution topography, available in the system or provided by the user. This service will also be used by other PDs as the computational core tool for producing a large number of simulations as required in Probabilistic Tsunami Forecast (PTF) and Hazard Assessment (PTHA).

Scientific and technological advancements. The service will facilitate the task of accurately simulating and doing faster the propagation over large domains of tsunami waves and also producing high-resolution inundation maps for tsunami hazard assessment both in extremely short computing times. Several scenarios should be computed at the same time and an aggregated scenario provided. If a new updated source definition is provided, new simulations should be launched. This service also opens the opportunity to develop PTF and PTHA tools with an increased spatial resolution. The application makes use of the ChEESE flagship code Tsunami-HySEA and uses telescopic grids to combine regional tsunami simulations at oceanic scale with local inundation simulations resolved to a few meters at city scale.

HPC added value. The availability of large computational resources is fundamental to achieving the objectives of the present service: (a) allowing increasingly faster simulations, the faster the simulation the greater the number of simulations that can be performed to assess the hazard in one single event, (b) providing the capability of consider high-resolution grids for the computation of the inundated areas and still computing FTRT, (c) making it possible to simulate enough scenarios to represent the tsunami hazard critical in PTF and PTHA applications.

Potential societal/economic impact. The ultimate goal of FTRT tsunami simulations is it use in Tsunami Early Warning Systems (TEWS) with the aim of saving lives. The potential impact in the event of a tsunami catastrophe for Society could be enormous, mainly by saving lives, but also mitigating the economic impact of the natural disaster. Users require accurate FTRT high-resolution tsunami simulations; this is the case of

TEWS and National Tsunami Warning Centers (NTWC). With this service, TEWS could provide not only one simulation of a given scenario but could refine and build confidence in predictions by considering many simulations simultaneously and/or considering new parameters defining the tsunami scenario. Moreover, this service also offers the possibility of high-resolution tsunami simulations, on a scale not done previously, in very short wall-clock times.

Impact on the scientific community. The tsunami scientific community in general will benefit from this service in many ways. The research tool provided for this service will boost tsunami simulation and tsunami research in many aspects, (a) facilitating a large number of tsunami simulations in very reduced computing times, (b) making extremely fast high-resolution inundation maps possible, (c) taking advantage of this for PTF and (d) making it possible to compute hundreds of thousands of tsunami simulations required for PTHA.

Potential target users

Targeted end-users include Tsunami Service Providers (TSPs), National Tsunami Warning Centres (NTWCs), Tsunami Researchers and Academic end-users, Insurance companies, Engineering companies, etc. We have currently collaborations with some stakeholders that can be classified as follows:

EDANYA-UMA previous collaborations

An important collaboration with the *NOAA Centre for Tsunami Research* (NCTR) of the Pacific Marine Environmental Laboratory (PMEL) started in 2011 and has been continued until the present year. First of all, our research collaboration was based on adapted Landslide-HySEA to the NCTR requirements in order to be used as initial conditions generator for landslide generated tsunamis. After this contract, currently our collaboration is based on adapting Tsunami-HySEA to be used in their SIFT (Short-term Inundation Forecasting for Tsunamis) system in order to be integrated in the United States Tsunami Warning Centres.

A permanent collaboration with the *CAT-INGV* (Centro di Allerta di Tsunami of the INGV) was established in 2012. The Tsunami-HySEA code has been tested and used in their operational system, and UMA maintains and provides improved and customized versions of the code to them.

A Collaboration Agreement was signed in 2015 between the *Joint Research Centre* (JRC) of the European Commission and the University of Málaga under which Tsunami-HySEA model was put at their disposal to be firstly integrated in the Tsunami Analysis Tool (TAT) of the European Crisis Management Laboratory (ECML) in JRC and secondly to be integrated as official numerical model for tsunami simulations in the *Global Disaster Alerting and Coordination System* (GDACS) of the European Union and UNESCO.

Since 2015 EDANYA-UMA has a collaboration agreement with the IGN (the Spanish NTWC) to jointly implement and develop the Spanish TEWS. The Tsunami-HySEA code has been installed and it is running at the IGN. Technical support has been provided and training courses done to IGN personnel. IGN is also a member of ChEESE IUB.

There exists a currently on-going project between IGME (Spanish Geological Survey) and CCS (Insurance Compensation Insurance, Spain) with the participation of EDANYA-UMA to estimate the economic losses associated with tsunamis in Spain.

Tsunami-HySEA has been installed at the Seismic Network of Puerto Rico and a customized version was implemented under a contract.

Collaborations with stakeholders from the IUB under ChEESE.

In October 2019, UMA officially started our partnership with the European *ARISTOTLE project*. To make this partnership happen, EDANYA-UMA had to work hard since July, 2019 because the ARISTOTLE tsunami service is an operational service provided to the European Civil Protection service in case of catastrophes and crisis.

Future collaborations.

There exists a joint project with the IGN to assist them in building the database of tsunamis scenarios for the Spanish TEWS at IGN. This project is scheduled to start from May 2020.

Among different international collaborations, future agreements with the Chilean SHOA (Servicio Hidrográfico y Oceanográfico de la Armada de Chile) and with the Costa Rican SINAMOT (Servicio Nacional de Monitoreo de Tsunamis) are highlighted. In the second case, during the last months, EDANYA-UMA has delivered both training courses and the Tsunami-HySEA installation in their servers. Respect SHOA, some specific versions of Tsunami-HySEA have been installed in their servers, giving support to them for the installation and management. UMA has renewed and signed an Agreement with the UTFSM (Universidad Técnica Federico Santa María) which is also linked with the collaboration with SHOA and in some sense collaborate in the Chilean NTWS.

Integration of Tsunami-HySEA and Landslide-HySEA in other systems.

Tsunami-HySEA has already been integrated in systems or platforms that are used for tsunami simulations. The first case was Tridec-Cloud, a cloud-based platform supported by the GFZ. They integrated Tsunami-HySEA in the Tridec system in 2016. https://trideccloud.gfz-potsdam.de/

At the same time, joint with our aforementioned collaboration with the JRC, EDANYA-UMA started to give support to their ITs in order to install Tsunami-HySEA in the GDACS system. Finally, they replaced their previous system (based on a database of precalculated scenarios) by a system based on real-time computation using Tsunami-HySEA.

More recently, in 2019, contact has been established by SeisCOMP3 (GFZ) and EDANYA-UMA provided technical support and a specific version of Tsunami-HySEA was developed to be integrated in their tsunami simulation software.

In recent dates two new integrations have been started. By one side EDANYA-UMA has been in contact with SHOA (Chile) in order to evaluate the integration of Tsunami-HySEA in their TEWS (SIPAT system). On other side, a research contract with NOAA (US) has been signed in 2020, where the main aim in the integration of Tsunami-HySEA in the SIFT system to be a complementary model and let to the TWC to use the MOST or the Tsunami-HySEA model for their simulations.

Service design and architecture

Use Cases

Use case	Rationale	Summary
Case 1. Pacific Ocean (Tohoku 2011)	It simulates the propagation of the 2011 Tohoku tsunami in the entire Pacific Ocean with a resolution of 1 arc-min, thus using a very big spatial domain, using a mesh of 84.2 million volumes	24 hours of simulation take 7343 seconds on a NVIDIA TESLA V100 GPU, and 280 seconds on 32 NVIDIA TESLA V100 GPUs
Case 2. Mediterranean Sea	It simulates a tsunami in the Mediterranean Sea (a more reduced scenario than Case 1) with a resolution of 30 arc-sec and a mesh of 10 million volumes	8 hours of simulation are performed in 275 seconds on a NVIDIA Tesla V100 GPU, and only in 23 seconds using 32 NVIDIA Tesla V100 GPUs
Case 3. Caribbean (with nested meshes)	It simulates a tsunami in Puerto Rico and its impact in a very localized area. The problem uses three resolution levels where the finest level has a resolution of 1 arc-sec. The total number of volumes is 23.6 million.	4 hours of simulation take 8525 seconds on a NVIDIA Tesla V100 GPU, and 550 seconds using 32 NVIDIA Tesla V100 GPUs

Functional Requirements

Functional requirement	Rationale	Summary
Time to solution	Use cases should be solved in few minutes (e.g. less than 5 min) FTRT simulations are a must in order to take measures as early as possible to protect the people in the coastal areas before the arrival of the tsunami	Cases 1 and 2 take less than 5 min using 32 and one NVIDIA Tesla V100 GPU, respectively. Case 3 takes 6.5 minutes using 64 NVIDIA Tesla V100 GPUs
Spatial resolution and domain size	Grids may be very big (e.g. up to 100 million volumes), and up to 24 hours of simulation time may be required. A tsunami covering an entire ocean may require a very big mesh for the coarsest resolution level. Several resolution levels may be needed to compute the inundation in a specific area, which usually is processed	Case 1 uses a mesh of 84.2 million volumes to cover the entire Pacific Ocean. The finest resolution level of Case 3 has a spatial resolution of 1 arc-sec, but we have worked with grids having spatial resolution as small as ⅓ arc-sec

	using a very high spatial resolution	
Data format	Support of NetCDF file format for I/O. NetCDF is a data format for scientific data widely used worldwide. There exist many libraries for many programming languages and also visualization tools. Furthermore, it also exists a parallel NetCDF library (PnetCDF) that allows multiple processes to read and write NetCDF files in parallel	Our solvers write all the output files in NetCDF format. Input files in gridded (GRD) format are supported

Environments

The service has been tested on realistic use cases, in relevant HPC environments (Power9 at the BSC, Spain; Davide and Marconi at CINECA, Italy; and at Piz Daint, Switzerland). Tests cases included the three cases above-described, producing information about the propagation and/or the inundated area, but also some other tests, in some cases massively computed in order to test other PDs within the ChEESE project, as for PD7 and PD8.

The service prototype will be run on a Linux environment where OpenMPI, parallel NetCDF and CUDA libraries are necessary. From the hardware side the only requirement is a NVidia GPU. Nevertheless, this model can be run in multiGPU environments.

The validation on the selected use cases aims to prove the feasibility of the workflow. For the time being, the service is to be operated by experts. During the course of the validation the results will be shared with the users in order to better define final outputs and products. The service will be ready for operation by other end-users by the end of the project.

For the ARISTOTLE case, the computing workflow has been integrated into the SPADA system, run at INGV. The workflow includes sending to UMA the input of parameters from INGV, using rabbitmq protocol, implemented in the SPADA system. The output is sent back to INGV and it is composed of some figures and data on maximal amplitude and arrival times. This information is currently being customized. Currently, this service is directly supported by computational resources allocated at the University of Malaga that are automatically launched in case of an event. In the near-future, when INGV has its own computational resources, the UMA computational resources will act as backup for this service.

Computational Resources

The typical amount of resources needed for a single scenario, was roughly estimated to 5-10 minutes of elapsed time using 32 V100 NVIDIA Graphic cards. HPC infrastructure is only needed to perform several test cases at the same time, which improves the efficiency of the on-fly simulations, that is, not only perform one single realization, but several of them and/or refine the prediction once more data about the true event is available. Moreover, if grid resolution is increased more resources will be needed.

The estimated required computational budget for the target Use Cases is reported in the following table (please note that the Total disk space requirements need to be decided, based on the adopted data preservation policy).

Use case	**# runs**	**# CPU cores/run**	**#GPU /run**	**Core hours / run**	**Total CPU M hours**	**Total memory (TiB)**	**Total disk space (TiB)**
Mediterranean	30000	2	2	5	0.3	Fits in nodes	
Tohoku	30000	2	2	10	0.6	Fits in nodes	
LANTEX	30000	2	2	10	0.6	Fits in nodes	

PD5. Physics-based probabilistic seismic hazard assessment

PD5	Physics-based probabilistic seismic hazard assessment
Leader	LMU
Participants	TUM, IMO, INGV, BSC
Codes	SeisSol, ExaHyPE, AWP-ODC
TRL initial	4
TRL target	6-7

Service definition

Service type. PD5 is a Probabilistic Seismic Hazard Assessment (PSHA) type of service prototype with a TRL of 6-7 (field specialists/end users).

Objectives. PD5 will contain general concepts for enabling physics-based seismic hazard assessment with state-of-the-art multi-physics earthquake simulation software (SeisSol, ExaHyPE, AWP-ODC) and conduct 3D physics-based seismic simulations to improve PSHA for validation scenarios provided by IMO (Iceland) and beyond. *We will make the pilot readily applicable to supplement established methods by stakeholders*. We expect the developed pilot to be readily applicable to different target regions and degrees of complexity of interest.

PSHA is the process of quantifying the rate (or probability) of exceeding various ground-motion levels at a site (or a map of the region) given all possible earthquakes. PSHA is a highly active field of research in the Scientific Community as well as of emerging interest for operational Risk and Hazard Assessment.

Scientific and technological advancements. Physics-based and HPC empowered PSHA allows the analysis of multiple forward models encapsulating the non-linear coupling of source, path and site effects affecting conjunctively ground shaking. Realistic model setups should acknowledge topography, 3D geological structures, rheology, and fault geometries with appropriate stress and frictional parameters, all of which contribute to complex ground motion patterns. For example, such fully dynamic multi-physics earthquake simulations can constrain dynamic and static stress transfers and physically limit "rupture jumping" and fault system interaction. Such methods also can shed light on the interplay and trade-offs of directivity, attenuation, anisotropy, off-fault deformation, fluid effects, severe frictional weakening, non-linear soils and other complex mechanisms.

The output we generate in PD5 will complement existing, mostly empirical SHA methods by providing

- Physics-based ground motion prediction equations (GMPEs), especially in the extreme near field
- Physics-based max. magnitude and fault-2-fault scenarios, e.g. for the definition of worst-case scenarios

- Physics-based 3D source, path (and site) effects on ground shaking
- Physics-based hazard maps

Potential societal/economic impact. We will enable the scientific and target users to perform and harvest results of 3D physics-based seismic simulations of different complexity levels for large amounts of hypothetical events. The results can be integrated into hazard maps, specifically considering their epistemic and observable uncertainties.

Potential target users

The PD5 consortium currently actively collaborates with the IUB members Fault2SHA working group and the Modeling Collaboratory for Subduction / Research Collaboration Network. PD5 results will be presented at the (virtual) EGU 2020 meeting as Li et al., EGU2020, "Physics-based constraints for probabilistic seismic hazard assessment in the Húsavík–Flatey fault zone, Northern Iceland".

In the scope of the ChEESE midterm GA PD5 connected with the following IUB members either in video interviews or via Email follow up:

ICP, Fault2SHA, Global Parametrics (GLOB PAR), ARM, Mitiga, Aristotele, Schlumberger, IASPEI, Intel, Protezione Civile, GEO-GSN, and the Modeling Collaboratory for Subduction / Research Collaboration Network.

We specifically mention GLOB PAR who expressed interest in porting PD5 to a new target region in Mexico.

We envision widening the field of potential target users, for example for the e-science environment of EPOS, the USGS Science Applications for Risk Reduction (SAFRR), Nuclear Power Plants, insurance companies, etc.

We also note a recently approved project related to PD5 and funded by the Southern California Earthquake Centre (SCEC) entitled "Validation of Broadband Ground Motion from Dynamic Rupture Simulations: towards better characterizing seismic hazard for engineering applications"

System design and architecture

Use Cases

Two Use Cases are envisaged, focusing on Northern and Southern Iceland. In North Iceland, extensional plate motion is accommodated by the Northern Volcanic Zone, a set of en-echelon volcanic systems, and the Tjörnes Fracture Zone (TFZ), a transform offset in the mid-Atlantic Ridge consisting of two parallel transform lineaments. The southern lineament, the Húsavík–Flatey fault (HFF) is a 100 km-long, segmented right-lateral strike slip fault, the largest transform fault in Iceland. The northern lineament, the Grímsey Oblique Rift (GOR) is composed of several active volcanic systems with N-S trending fissure swarms. Large earthquakes along this branch are mostly caused by strike slip faulting, often on transverse faults, by bookshelf-type kinematics. Thus, the HFF and GOR pose a significant seismic hazard to the nearby coastal communities and past earthquakes have caused considerable damage. A third of the full transform motion is accommodated by the HFF, while the rest is focused on the GOR. On the other hand, the

South Iceland Seismic Zone and the Reykjanes Peninsula Oblique Rift is one of the seismically most active and highest risk regions in Iceland.

Use case	Rationale	Summary
Northern Iceland	The TFZ is currently under active geophysical investigation: Various complex models of the TFZ fault system have been presented, and of its activity, aided by the detailed mapping of its bathymetry. Moreover, PSHA has been carried out using a simplified model of the TFZ fault system, the sensitivity of the PSHA for the TFZ due to different empirical GMMs the preliminary modeling and analysis of dynamic rupture on the HFF the constraining of earthquake source parameters from the Icelandic strong-motion dataset the detailed site-effect investigation in Húsavík.	We provide a framework of data-integrated and data-verified multiple forward simulations of varying degree of complexity: ranging from simple logic-tree approaches to kinematic source modeling to multi-physics dynamic rupture scenarios towards physics-based PSHA.
Southern Iceland	Southern Iceland Seismic Zone is a well recorded site. The study region is focusing on the transform belt of RPOR and SISZ and aligns with the study region of PD1 (see PD1 for more details).	We provide a framework for high resolution kinematic and dynamic wave propagation simulations in a synergetic effort with the SCEC Cybershake framework, exploring increasing complexity as topography, 3D velocity, and non-linear source effects and urgent simulations in PD1.

Functional Requirements

Functional Requirement	Rationale	Summary
CPU time	Approximated minimum cost in terms of resolution	One typical forward simulation of intermediate complexity and resolution costs ~2k CPU hours
Resolution of ground motion in the engineering frequency band	Each forward simulation may be scaled up to full-machine size if increasing resolution / physical complexity (cf. Heinecke et al., SC'14 or Uphoff et al., SC'17)	Computational cost is dictated by: 1) the resolved frequency content of the seismic Wavefield 2) the resolution of multi-physics rupture dynamics on and off complex faults 3) consideration of non-elastic and non-isotropic components such as viscoelastic attenuation, anisotropy, etc.

Parallel Input/Output	Asynchronous output, local time stepping enabled	Fully parallel I/O in form of hdf5/NetCDF data formats
Automatic check pointing	Resilience crucial on Exascale machines	
Statistical workflow	Large ensemble size required for extracting probabilistic hazard information	Hundreds- thousands of forward simulations: data handling, storage, automatic pre- and post-processing

Non-Functional Requirements

Non-Functional Requirement	Rationale	Summary
User support	The workflows are designed for manual user interaction with respect to the target use cases.	Documentation, examples, user support are provided.
Licensing	The policy of the development team is to release software and workflows open-source	Licensing issues will be discussed with stakeholders (open-source might be non-optimal for private companies)

Environment

PD5 expects TRL 6-7, thus it is assumed that the Demonstrator will be tested in a relevant environment making the Service potentially available to a broader pool of experts and in operational environments.

Services will be validated by the PD5 working group. Specifically, the service for both use cases will run at the LRZ, BSC and Swiss Supercomputing Center. In the scope of ChEESE WP2 the SeisSol flagship code is ported to GPU environments.

Computational Resources

The more resources the better, since scalability will be one of our main design and optimization criteria in the scope of PD5. For operational usage of PD5 services we expect to require computational resources of at least $O(10^6)$ CPU h for the generation of a specific hazard map designed for a specific target region and accounting for a specific degree of complexity/data-set.

PD6. Probabilistic volcanic hazard assessment

PD6	Probabilistic volcanic hazard assessment
Leader	INGV
Participants	BSC, IMO
Codes	FALL3D, ASHEE
TRL initial	3
TRL target	6-7

Service definition

Service type. We are proposing two prototype services for short-term (days to weeks before an eruption) and long-term (years to decades) Probabilistic Volcanic Hazard Assessment (PVHA).

Objectives. The objective of this twofold service is to overcome the current limits of PVHA imposed so far by the high computational cost required to adequately simulate complex volcanic phenomena (such as tephra dispersal) while fully exploring the natural variability associated to such volcanic phenomena, on a country-size domain (~thousands of km) at a high resolution (few km resolution). In the case of tephra hazard assessment, to achieve unbiased PVHA it is mandatory to account for the natural variability linked to the eruption scale or type, the consequent Eruptive Source Parameters (ESP), the position of eruptive vents, and the wind field.

Scientific and technological advancements. In more detail, we explain the two main points of the objective of this service:

1) *Full exploration of the ESP and wind field:*

So far, the most advanced PVHA for tephra hazard assessment are based on the concept of "representative eruptive scenarios" (a multifaceted term indicating a broad set of eruptive conditions): a few (or one) eruptive scenarios are subjectively identified by experts, and the hazard is computed by simulating tephra dispersal under the assumption of their "representativeness", in terms of hazard. This main assumption can be translated into that the variability due to combinations of parameters characterizing eruptions "near" a representative scenario is negligible when compared to the variability between the discretized, few representative scenarios. The effect of this assumption has been quantified previously for two specific volcanoes (Campi Flegrei and Vesuvius), on a domain of about 200×200 km, by exploring how hazard results change when considering a continuum of possible combinations of ESP, which translates into exploring a large set of simulations on this relatively small-scale domain.

2) *Country-size (or larger) domain:*

Distal (>few hundred km) tephra fallout, and/or distal airborne ash concentration, may represent a source of hazard only in case of extremely large and/or under extreme wind eruptions. In particular, distal tephra fallout may pose a risk on crucial assets such as water storage system, power lines or transport network (road, airports, railways, etc.), whereas distal airborne ash concentration may pose a risk to the aviation. Because of this, if we want to take into account also low-probability but high-consequence events in PVHA for tephra fallout, we need to account for large-scale (~thousands of km) but high-resolution (~1 to few km) domain (see e.g. right panel in Figure 3).

HPC added value. HPC will allow larger simulation ensembles. This translates into a large number of numerical simulations for tephra transport and deposition, on a large-scale domain and at high resolution, that we propose to perform with the flagship code FALL3D optimized in ChEESE. In this way, the service will allow the user to exploit HPC Exascale resources to explore the natural variability due to unknown ESP, vent position and climatological wind field variability, over a large-scale and high-resolution domain. To iterate, the added value from the exploitation of Exascale computing facilities allows us to increase the number of simulated eruptions with different ESP, on a large-scale and high-resolution domain.

Potential societal/economic impact. The two proposed services (short- and long-term PVHA) explore two different time scales of the forecast of tephra, and thus rely on different scientific evidence and computational needs. The type of final products is also thought to be exploited in different phases of a volcano life-time. A short-term PVHA is intended to be adopted during a volcanic crisis and when there is the urgency to take decision upon scientific data and observations. A long-term PVHA is instead intended to provide decisional support for a long-term assessment that might include land-planning, definition of mitigation measures.

As regards, the short-term PVHA (ST-PVHA), the pioneer approach adopted in PD6 is based on an Event Tree approach that combines: *(i)* evidence from real-time monitoring network to compute the probability of unrest, magmatic unrest, eruption and position of the vent, and *(ii)* results from on-the-fly simulations of tephra dispersal coupled with up-to-date wind field forecast. In this respect, this service is similar to an Urgent Computing one, because it must be scheduled to run routinely (once or twice a day), as soon as the new weather (wind) forecasts are available, which might occur at slightly different time every day. The number of simulations foreseen for this service are of the order of tens to hundreds per scheduling, in order to explore ESP uncertainty significantly, without having to wait too long for the result of the forecast, that needs to be available in minutes to few hours, in order to be usable by decision makers.

A similar service prototype was preliminary implemented in 2017 as an operational tool at the Osservatorio Vesuviano (Napoli) control room on a small-scale domain (200×200 km, see left panel in Figure 3). However, due to the computational cost of simulating twice a day the eruptive variability on local servers, it is currently based on a few representative scenarios approach.

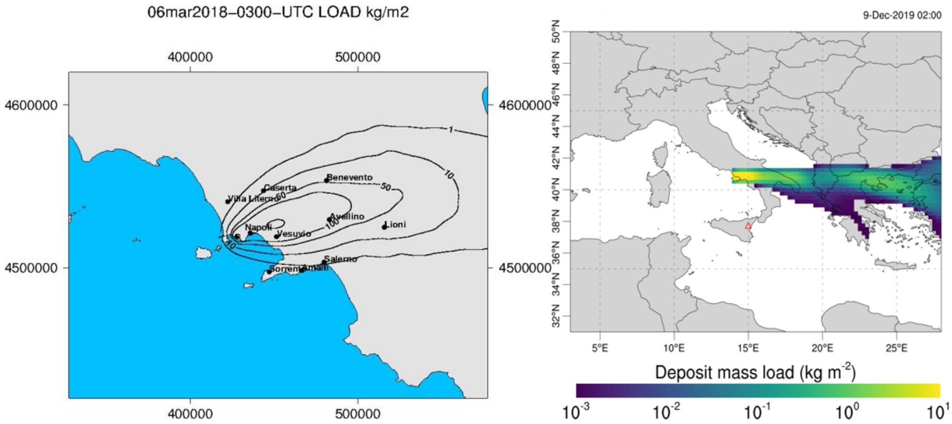

Figure 3. Left panel shows the current domain for the current prototypal service running twice a day Fall3D simulations for 3 specific eruption scenarios in Osservatorio Vesuviano control room. Right panel shows the domain explored by the proposed HPC service and LT-PVHA.

Regarding the Long-Term PVHA (LT-PVHA), we propose a service for which the user provides the ESP range for the volcano of interest, and a large number (several thousands) of FALL3D simulations is performed, based on a stratified sampling procedure for the ESP values, possible vent position, and wind field in each simulation. The wind field is sampled from the ERA5 database in the last 30 years, in order to sample the wind variability expected for a long-term PVHA (years to decades).

Potential target user

The most direct end-users of our services are Civil Protection Agencies. Previous contacts and ongoing collaborations between INGV, IMO, BSC and the Italian and Icelandic Civil Protections have been essential to design the Service functionality. However, their direct engagement in ChEESE will be developed in the second half of the project.

For the ST-PVHA, Services already in place for the Italian Civil protection provide the daily forecast of tephra dispersal from Campi Flegrei, or conditional to an eruption by Etna in Sicily, relying on a few selected scenarios. In the proposed ChEESE use cases, the ST-PVHA service we propose would represent a relevant step ahead for the Italian civil protection, and in general for civil protections acting in populated areas threatened by more than one volcano, such as the Neapolitan area and Reykjanes peninsula. In such cases, a full service could be able to combine hazard from all the potential sources (e.g., Vesuvius and Campi Flegrei in the Neapolitan area). The relevance of such a use-case was agreed by ***IAVCEI*** secretary (from IUB), whose suggestion has been to focus on densely inhabited areas to prioritize the application of these services.

A similar background exists for Iceland where the presence of volcanic systems (i.e. a volcanic fissure associated or not to a central volcano) make the anticipation of the possible vent opening very challenging. Over the last years, several cases of volcanic unrests in Iceland would have benefitted greatly by the availability of ST-PVHA. A couple of examples in the last few years are the Öræfajökull volcano and the Reykjanes-Svartsengi volcanic system for which a quick scenario-based hazard assessment was provided without the necessary consideration of the spatial uncertainty in the vent location. The Reykjanes-Svartsengi case includes both tephra and SO2 dispersal as the eruption scenarios cover the full range from purely effusive eruption to phreatomagmatic events.

In more detail, the strategy foreseen to develop the ST-PVHA service is based on the following steps:

1. The PD team should provide scientific support to set the initial conditions for the workflow functioning (definition of the ESP ranges and probability density function, of the probability of vent opening, and interpretation and inclusion in the workflow of monitoring data);
2. Civil protection should be directly involved, as well as consultants of civil protection (e.g. PLINIVS for the Italian case), in the definition of what output of the service is relevant to the user, for example:
    - hazard curves and/or hazard maps and/or probability maps;
    - thresholds in the corresponding exceedance probability or intensity measure, for hazard maps or probability maps respectively;
    - uncertainty measure to be provided, such as the mean or the median or some percentile of the output distribution;
    - time window of forecast (for example [00-24], [24-48] and [48-72] hours after each service launch).
3. The PD team runs the service and share the products with civil protection (and possibly consultants). This is the current view, but in the validation phase we will re-discuss with Civil protections their degree of involvement.

A ST-PVHA service developed for any volcano is also relevant to ARISTOTLE, as expressed in our contact during ChEESE General Assembly in March 2020. Their level of involvement expected with ARISTOTLE however accounts only for the access to the output information provided by the service, to exploit it in case of volcanic crises at the use-case volcanoes. In this view, the output from the service run for (or by) the civil protection will be shared with ARISTOTLE.

Scientific interest has been expressed in LT-PVHA by ***IGN*** in the contact we have had during the GA. They are interested in an application to Teide volcano; however, little is known on ESP for Teide, so they need to calibrate this scientific information on other volcanoes that are analogue to Teide. In this case, the strategy foreseen is that the user runs the service, supported in the initial stages by the PD team. ***Italian*** and ***Icelandic Civil protection*** are interested in the same use cases for the STPVHA, and the development strategy for LTPVHA does not differ from the one for STPHVA service.

System design and architecture

Use Cases

Use case	Rationale	Summary
Campi Flegrei (ST-PVHA) Operational User:	This is an improvement over the current activity at Osservatorio Vesuviano (see first section of this report). There is a convergence here between scientific interest and	- Domain 2000km×2000 km, 1km resolution, 14 flight levels - 100 Fall3D simulations per size and per time interval of

Italian Civil Protection	end user (Italian Civil Protection) needs	24hours ([00-24], [24-48], [48-72] hours), per day This allows exploring uncertainty on the occurrence of an eruption in a given time window, on the eruption size and related ESP, and on the vent position.
Teide (LT-PVHA) User: IGN	IGN expressed scientific interest in this service	It implies defining analogue volcanoes to Teide, define eruptive scales, their relative probability distribution, and using this new scientific information into the LTPVHA service.
Reykjanes (ST-PVHA) Semi-operational User: Icelandic Civil Protection	This tool does not currently exist in Iceland. Reykjanes is one of the region most exposed to volcanic hazards in the country due to the presence of key infrastructure (including the main international airport, two geothermal power plants, the main road connection to the capital area). The unrest occurred in January 2020 revealed the need for a tool in place.	- Domain 1000 km×1000 km, 0.5 km resolution, 14 flight levels - 100 Fall3D simulations per size and per time interval of 24hours ([00-24], [24-48], [48-72] hours), per day - two contaminants: tephra and SO_2 (only ground concentration) This allows exploring uncertainty on the occurrence of an eruption in a given time window, on the eruption size and related ESP, and on the vent position.
Jan Mayen (LT-PVHA) User: Icelandic Civil Protection	A LT-PVHA for Jan Mayen is foreseen for the success of the project. The main expected usage is for aviation purposes in order to evaluate the potential impact on air-routes flying north of Iceland.	The work needed includes the analysis of ESP distribution as not so much historical activity produced a relatively poor dataset to constrain eruptive scenarios. A 2000×2000 km domain is selected to investigate the potential effect on both air traffic over Iceland as well as over the UK.

Functional Requirements

There will be a dedicated meeting to define functional requirements before the validation phase. A preliminary assessment of functional requirements is reported in the following table.

Functional requirement	Rationale	Summary
FALL3D on a large-scale and high-resolution domain	This enables to explore the impact of tephra fallout and airborne ash concentration at distal locations, to evaluate the hazard posed by low-probability but high-consequence events	Domain approximately 2000x2000 km, 1km resolution, 14 flight levels (ST and LT service)
Number of FALL3D simulations (LT-PVHA)	To explore the variability in eruptive scale or type (and thus in related ESP), vent position and wind climatic statistics, we need a large number of simulations	Ideally, we need to run 10000 FALL3D simulations per eruptive size. Possibly 1000 is enough if sufficient computational resources are not found.
Number of FALL3D simulations per service launch (ST-PVHA)	To explore the variability in eruptive scale or type (and thus in related ESP) and vent position, we need a large number of simulations. Here, the wind field is taken from the most recent weather forecast.	Ideally, we need to run 100 FALL3D simulations per day, per time interval of forecast (tomorrow, in two days, in 3 days and so on) and per eruptive size.
FALL3D simulations in less than 3 (?) hours after each service launch (ST-PVHA)	The ST-PVHA needs to be fast enough to enable the usability of the resulting PVHA. If it takes too long, it is useless.	The maximum delay after the service launch is a user need and must be specified by the user
Workflow runtime less than a few hours after each service launch (ST-PVHA)	The ST-PVHA needs to be fast enough to enable the usability of the resulting PVHA. If it takes too long, it is useless.	The maximum delay after the service launch is a user need and must be specified by the user
Automatic and timely access to weather forecast (both runs at 00 and 12) (ST-PVHA)	FALL3D simulations in ST-PVHA will be run assuming the most recent wind forecast as the actual wind field for the next time window	This must be set up by the PD team in cooperation with the weather service providing forecasts in the domain of simulations, in agreement with the Civil Protection user that may guide the cooperation between PD team and the weather service.
Temporary storage and manipulation of a large set of FALL3D simulations (LT-PVHA)	The workflow needs to analyse a large number of simulations	Defined in the validation phase?
Storage of ESP probability density functions	These data represent the information to form a new set of starting conditions for the FALL3D simulations; they need to be stored for future scientific reanalysis of the results	This should not represent a problem, but it must be kept in mind as a to-do task.

Storage of the ESP and vent position samples used in input to FALL3D simulations (LT-PVHA)	These data represent the starting conditions of the FALL3D simulations and need to be stored both for workflow processing and for future scientific reanalysis of the results	This should not represent a problem, but it must be kept in mind as a to-do task.
Real-time access to Volcano Observatory database of monitoring data (ST-PVHA)	To evaluate to probability of unrest, magmatic unrest and eruption, the workflow relies on a Bayesian Event Tree (BET); the workflow must automatically fetch the up-to-date values of the relevant monitoring parameters, translate them into a degree of anomaly, and run the BET model to compute the aforementioned probabilities.	This must be set up by the PD team in cooperation with the Volcano Observatory, in agreement with the Civil Protection user that may guide the cooperation between PD team and Volcano Observatory.
Storage of BET settings	These data represent the information used to compute the probability of unrest, magmatic unrest and eruption; they need to be stored for future scientific reanalysis of the results	To be done.
Access to ERA5 reanalysis dataset (LT-PVHA)	To explore the variability in wind climatic statistics, we need to access a meteorological database and sample wind fields for the FALL3D simulations	To be done.

Environment

The service will be tested on realistic use cases, in an operational environment, i.e., in similar conditions to those expected for an operational service.

The ST-PVHA Pilot Demonstrator (at Campi Flegrei, IT) will be tested and validated in the observatory monitoring room at INGV-Osservatorio Vesuviano.

The ST-PVHA Pilot Demonstrator (at Reykjanes, IS) will be tested and validated in the observatory monitoring room at the IMO and shared with the Icelandic Civil Protection.

The LT-PVHA Pilot Demonstrator (at Campi Flegrei, I; and Jan Mayen, NO) will be tested and validated by INGV and IMO, and will be shared with the National Civil Protections. Ideally, IGN could take part in the testing and validation, as an interested scientific end user.

The service is tested on realistic use cases, in relevant HPC environments: Fall3D simulations for PD6 (use cases listed above) will be run in CEA-TGCC (France) on the Joliot Curie machine, for which a 2-year PRACE application (application number is

2019215114) has been funded starting on 1st April, 2020. The total awarded for tasks related to PD6 are about 5.4 Mhours, on two partitions (Irene Rome and SKL Irene).

Computational Resources

The estimated required computational budget for the target Use Cases is:

Use case	# runs	# CPU cores/run	#GPU /run	Core hours / run	Total CPU M hours	Total memory (TiB)	Total disk space (TiB)
Campi Flegrei (LTPVHA)	3650	1024	n/a	425	1.55	Fits within nodes	7.30
Jan Mayen (LTPVHA)	3650	1024	n/a	425	1.55	Fits within nodes	7.30
Campi Flegrei (STPVHA)	4320	1024	n/a	425	1.84	Fits within nodes	8.64

PD7. Probabilistic tsunami hazard assessment

PD7	Probabilistic tsunami hazard assessment
Leader	NGI
Participants	INGV, UMA
Codes	Tsunami-HySEA and Landslide-HySEA
TRL initial	3
TRL target	5-7

Service definition

Service type: The expected service prototype is a Probabilistic Hazard Assessment. Probabilistic Tsunami Hazard Analysis (PTHA) addresses the probability of exceeding a given metric of tsunami intensity at a given location within a given time interval. This metric can be, for example, a tsunami inundation height, velocity, or momentum flux.

Objectives: Develop a PTHA service that can provide local tsunami hazard maps onshore at high resolution over a local high-resolution topography dataset provided by the user.

Scientific and technological advancements: The service will facilitate a unique combination of providing 1) the hazard for a very large set of sources for an accurate representation of the hazard and a rich representation of the uncertainty and 2) the tsunami hazard at a very high spatial resolution. A service with such a capability does not exist anywhere else today. The application makes use of the ChEESE flagship code Tsunami-HySEA and uses telescopic grids to combine regional tsunami simulations at oceanic scale with local inundation simulations resolved to a few meters at city scale, combining tens to hundreds of thousands of simulations to aggregate the hazard.

HPC added value: The availability of large computational resources is instrumental in fulfilling the objectives of this services through: (a) providing sufficient resolution and coverage at the local site of interest, (b) making it possible to simulate enough scenarios to represent the tsunami hazard with sufficient accuracy while rigorously resolving the epistemic uncertainty due to source variability. A successful Exascale and pre-Exascale HPC PTHA application will solve this issue.

Potential societal/economic impact: Users need accurate high-resolution tsunami hazard maps for emergency planning and coastal engineering considerations. Most presently available PTHA software with data of similar complexity is limited to representing the hazard offshore, and the high-resolution coastal representation is missing. Similarly, PTHA applied to producing local high-resolution inundation maps can currently only cover a limited set of sources. The PD7 based service is unique in the sense that the user can upload their own local topography and map the tsunami hazard in this region. This will provide the following impact: 1) a tool to predict the consequences of existing offshore hazards on high resolution onshore maps that can be used in practical hazard and evacuation mapping 2) reduced uncertainty and land use restrictions through higher accuracy.

Impact on scientific community. The results of the analysis will be useful for society, but the use of the service may be equally important for the scientific community. Practices of simulating tsunami inundation using local telescopic grids have dominated for a long time, but only for single scenario analysis. The new service will bring this step further, through embedding hazard and uncertainty into local tsunami inundation modelling.

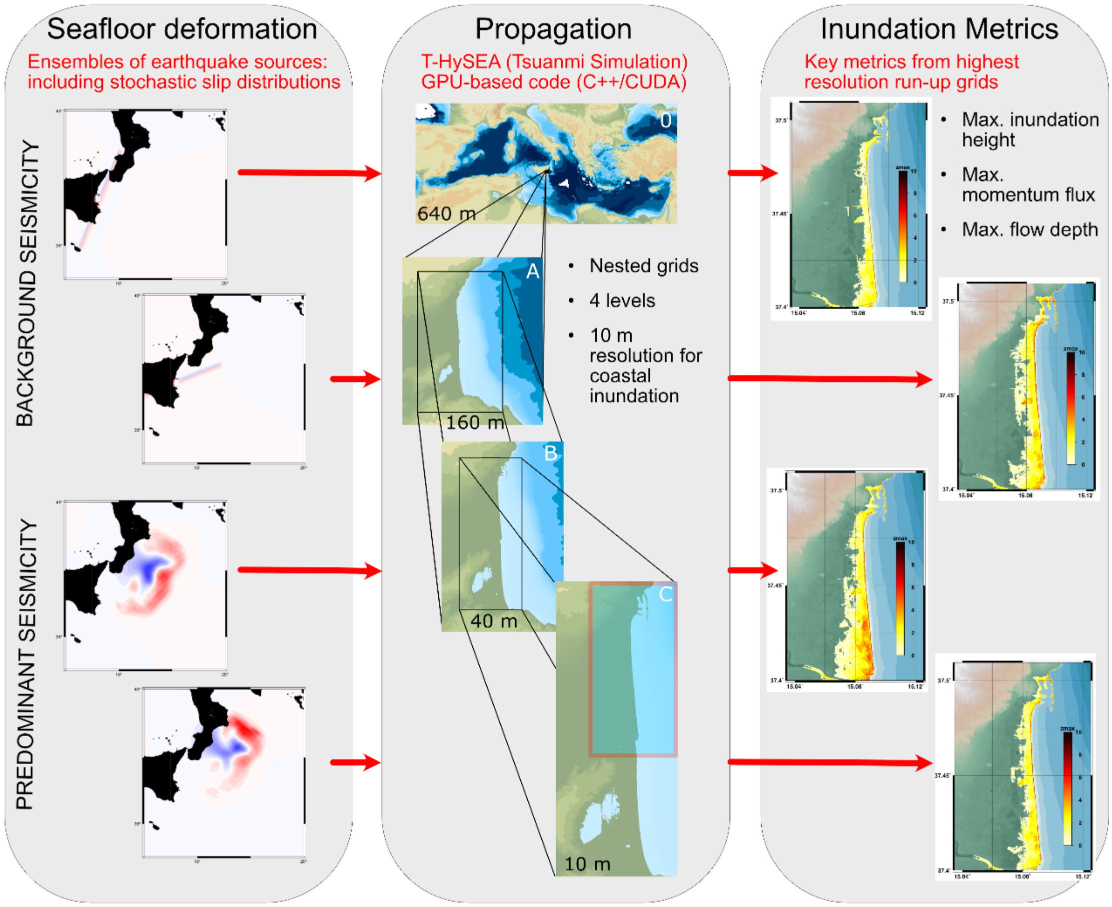

Figure 4. The core of the PTHA workflow. A high number of earthquake scenarios are specified and the tsunami propagation and inundation from the corresponding seafloor deformations is calculated using the GPU-optimized Tsunami-HySEA code. Finally, high-resolution tsunami hazard maps are calculated aggregating the hazard posed in the different scenarios.

Potential target users

Given the nature of the hazard maps generated, the potential target users of PTHA fall into four main categories:

1. National stakeholders and local stakeholders as well as Civil Protection. High-resolution hazard maps will facilitate emergency and evacuation planning.
2. Coastal Engineers. Site-specific hazard probability will guide consideration for construction standards and viability.
3. Insurance. Accurate models for hazard probability for a given location will help in categorization of property for insurance premiums.
4. The scientific community. A service in which rapid and accurate probabilistic hazard assessment can be conducted will create a platform on which researchers

can conduct robust investigations both into source physics and tsunamigenic processes and into factors affecting inundation. Both aspects are likely to improve future PTHA implementations.

System design and architecture

Use Cases

Use case	**Rationale**	**Summary**
Catania Hazard study	Test the functionality of the PD7 workflow on a realistic use-case (relevant environment)	Provide detail tsunami hazard map for pre-Exascale computations and show practical examples of usability. Link to potential operational requirements (early warning, land-use planning)

Functional requirements

Functional requirement	**Rationale**	**Summary**
Single scenario 8 hours wave propagation and inundation solved within 1 GPU hour	Manageable overall run time given the large number of events	Order (O) of O(10^5-10^6) scenarios may be foreseen, with O(10^3) GPU cores this can be accommodated with a few days to a few weeks' time.
Required spatial resolution of the order 10 m for inner grid for single runs	Need fine resolution for converged solution	Tests will be run at 10 m inner grid resolution
Inner grid should cover a city-scale areal extent	Sufficient coverage for application	Typical range would cover a 10-25 km coastal stretch
Required spatial resolution of maximum 30" (~900 m) global grid for single runs	Need fine resolution for converged solution	Tests will be run at 640 m global grid resolution covering the Mediterranean Sea
Telescopic grid refinement ratio 4	Compatibility requirements in Tsunami-HySEA	Respective resolutions, 640 m, 160 m, 40 m, 10 m.
Four levels of topo-bathymetric maps in netCDF format for telescopic grids required as input	Necessary input data	Inundation computed in the local higher resolution grids
Hazard level reproduced to 95% accuracy of original assessment	Must reproduce hazard sufficiently accurately from original assessment	For a pre-exascale test we expect this to require O(10^5-10^6) scenarios
Must be based on TSUMAPS-NEAM assessment	All sources and probabilities originate from this assessment	
Provide the hazard maps at a spatial resolution of 10 m for full set of exceedance probabilities for all wet areas	Needed for integration to tsunami risk	Different hazard metrics, flow depth, tsunami height, momentum fluxes.

Non-functional requirements

Non-functional requirement	Rationale	Summary
Interface for integration of user-specified topo-bathymetric grids and visualization of output.	Users are likely to want to apply the system to a specific region of relevance to them and their clients and are likely to want to specify their own topo-bathymetric maps. Output needs to be compatible with the users' GIS solutions.	A tool is being development which will ensure that user-specified topo-bathymetric input is valid input for the tsunami simulations. It will check the dimensions, units, and orientation. It will interpolate and output HySEA-compatible input when possible and inform the user if not possible. The hazard maps will be in formats compatible with GIS platforms.

Environment

Expected TRL for this service will be in the range of 5-7. The service will be tested primarily in a relevant HPC environment, producing hazard maps for at least one test site exposed to tsunami hazards. A typical environment is defined through the Catania used case above, where a proper high-resolution grid is available to test the full functionality of the service. The service prototype will be run on GPU clusters running Linux. The process has already been test-run on Piz Daint at CSCS.

The validation on the selected use cases aims to prove the feasibility of the workflow, but for the time being the service is to be operated by experts. During the course of the validation the results will be shared with the potential users in order to better define final outputs and products.

Computational Resources

We provide typical costs for a probabilistic tsunami hazard analysis in the Mediterranean Sea for the Sicilian towns of Catania and Siracusa. We consider cases calculating for each town separately and also for calculations are performed for both towns simultaneously. (The Tsunami-HySEA code can run with multiple nested grids.)

We consider cases where all of the tsunami sources are those calculated in the TSUMAPS study. When the two towns are considered together, the 99% hazard contains many common scenarios. The number of runs required for calculating the hazard at both towns in a single simulation is higher than the number of runs required for only one of the towns but is less than the sum of the number of runs required for both towns in separate calculations.

The total disk space volumes are based on saving three parameters at all points on the finest resolution grid for every run. Reductions in disk usage can be achieved by storing the values at a significantly reduced number of grid nodes or by performing hazard aggregation and visualization on the HPC resources and returning only aggregated data.

The estimated required computational budget for the target Use Cases are reported in the following table (please note that the numbers for total disc space usage are without compression. We will apply extensive file compression which is expected to reduce these numbers by at least one order of magnitude):

Use case	# runs	# CPU cores/run	#GPU /run	Core hours / run	Total CPU M hours	Total memory (TiB)	Total disk space (TiB)
Catania 99% hazard study (TSUMAPS)	32443	0	1	0.8	0.259 M GPU hours	10G VIRT 1.5G RES per run.	1.62
Siracusa 99% hazard study (TSUMAPS)	32514	0	1	0.8	0.260 M GPU hours	10G VIRT 1.5G RES per run.	0.97
Catania and Siracusa 99% joint study (TSUMAPS)	42720	0	1	1.2	0.513 M GPU hours	10G VIRT 3.0G RES per run.	3.4
Catania 99% hazard study (Refinement)	973290	0	1	0.8	0.779 M GPU hours	10G VIRT 1.5G RES per run.	48.6
Siracusa 99% hazard study (Refinement)	975420	0	1	0.8	0.780 M GPU hours	10G VIRT 1.5G RES per run.	29.3
Catania and Siracusa 99% joint study (Refinement)	1281600	0	1	1.2	1.538 M GPU hours	10G VIRT 3.0G RES per run.	102.5

PD8: Probabilistic Tsunami Forecast (PTF) for early warning and rapid post event assessment

PD8	Probabilistic Tsunami Forecast (PTF) for early warning and rapid post event assessment
Leader	INGV
Participants	UMA, NGI
Codes	Tsunami-HySEA
TRL initial	3
TRL target	6 - 8

Service definition

Service type: the service prototype is to be run immediately after a seismic event occurs. It provides Probabilistic Hazard Assessment (PHA) for an impending tsunami event. It may be used both in the context of Early Warning (EW) operations and in that of *Rapid Post-Event Assessment*; in both cases, Urgent Computing (UC) resources must be available.

Objectives: This pilot is aimed to provide rapid tsunami impact probability distribution along the coasts. Given the uncertain parameters of an earthquake, the ensuing uncertain tsunami intensity is estimated. The probabilistic tsunami forecasting can be updated as additional data come in. The geographic coverage is in principle global but some regional data are necessary for that. Hence, the first version will be dedicated to several use-cases at specific locations.

Scientific and technological advancements: Tsunamis may strike a coastal population within a very short amount of time. To effectively forecast and warn, extremely fast tsunami simulations are needed. Until recently, such urgent tsunami simulations were practically infeasible. However, Graphical Processing Units (GPUs) have opened a new avenue for performing Faster Than Real Time (FTRT) tsunami simulations (see PD2). The tsunami impact distribution is always complex and spatially heterogeneous, because of many factors related to complexity both of the earthquake source and of the hydrodynamic processes. Main uncertainty drivers include tsunami energy directivity related both to fault-to-coast orientation and bathymetric and topographic variations, earthquake slip direction (faulting mechanism) and heterogeneity, nonlinear inundation stage dynamics. In the case of distant trans-oceanic tsunami early warning, these uncertainties may be greatly reduced and perhaps implicitly treated as negligible via real-time assimilation and inversion of deep-sea tsunami data. For local tsunamis and/or if these data are not available, as in most of the cases worldwide, the tsunami forecast must be rapidly formed based on limited seismic data only. A prototype implementation of the PTF workflow for the Mediterranean Sea already exists and it is currently being tested at INGV in view of operational deployment for tsunami warning. This prototype version is based on a database of pre-calculated tsunami scenarios. In the new ChEESE workflow, the database will be replaced by FTRT tsunami simulations, increasing accuracy in the

Mediterranean and allowing for its applicability to any possible event worldwide. This will allow testing against past tsunami events to ensure PTF calibration and validation.

HPC added value: The earlier the tsunami forecasting, the lesser the available information, the higher the uncertainty. The inherently large uncertainties, due to limited seismic and tsunami data availability soon after an earthquake occurrence, especially for local tsunamis, can be then quantified through Probabilistic Tsunami Forecasting (PTF), using a very large number of FTRT simulations. Exascale capacity is then crucial to perform the high number of FTRT numerical simulations needed to obtain accurate results from PTF accounting for existing uncertainty, in order to rapidly assess the tsunami impact at different spatial scales (global, regional and local). A PRACE Project (TSU-CAST) was awarded for this PD8 PTF workflow testing. TSU-CAST is led by UMA while INGV, NGI, CINECA are co-PIs. TSU-CAST will allow us about 300,000 GPU hours on MARCONI100 at CINECA, Italy. The deployment of MARCONI100 is completed and then the project is about to start.

Potential societal/economic impact: PTF allow quickly forecasting coastal tsunami impact accounting for the uncertainty at the time of the estimate. It allows producing probabilistic hazard maps at regional scale and, potentially, a high-resolution inundation maps at specific sites. If sufficient urgent computing resources are promptly available, it may support decision making by informing estimates of location and size of the potential impact, then allowing for a rational approach to rapid post event assessment and disaster risk reduction. PTF can also help improving tsunami warning services, allowing for a rational definition of forecasting capabilities and safety levels to be adopted. In fact, forecasting and warning reliability would be improved by quantifying uncertainties in a PTF, rather than using, for example, a worst-case based approach. PTF can be then automatically translated into alert levels based on conversion rules established in advance. This allows a clear distinction between science and decision-making and a rational approach to disaster risk reduction.

Potential impact on Scientific Community: PTF provides the scientific ground for the quantification of uncertainty on real-time forecasts, producing results statistically testable against real observations. A worldwide extension of PTF provides the possibility to quantitatively test the forecasting capability of seismic monitoring systems and of tsunami modelling procedures, leading to future improvements of both forecasts and the underlying science. A paradigm-shift will be encouraged in this way towards explicit uncertainty quantification in tsunami warning operations, which are presently based on a worst-case oriented deterministic forecast in view (or perhaps we may say despite) of the high-uncertainty regime.

Potential target users

As just mentioned, an almost operational version of the PTF exists at ***CAT-INGV***, i.e. the upstream component of the Italian tsunami early warning center. Hence, the very first end-user is the CAT-INGV itself, which provides tsunami messages to:

1. the ***Italian Civil Protection*** (represented in the ChEESE IUB) for distribution to the national territory, and to
2. the ***NEAMTWS*** (the North-eastern Atlantic, the Mediterranean and connected seas Tsunami Early Warning and Mitigation System), which serves

IOC/UNESCO Member States and other Institutions in the region based on a subscription mechanism.

The PD8 responsible is a member of the CAT-INGV Governing Board, so CAT-INGV is directly involved in the definition of the service. Other tsunami warning centers belonging to the constellation of the Intergovernmental Coordination Groups (ICG) for the different world oceans coordinated by IOC/UNESCO to which the ICG/NEAMTWS refers might be interested to PTF, at least as a training tool for exploration of tsunami forecasting uncertainty.

PTF is also of specific interest for the ARISTOTLE ENHSP (European Natural Hazard Scientific Partnership), which is also represented in the ChEESE IUB. ARISTOTLE ENHSP in fact deliver multi-hazard advice capability to the Emergency Response Coordination Centre (ERCC) of the EU DG-ECHO (General Direction of European Civil Protection and Umanitarian Aid Operations), which is used in turn by ERCC for planning post-disaster intervention. Improving the tsunami impact rapid forecast capability is one of the primary goals of ARISTOTLE. INGV scientists involved in ChEESE are involved as well in ARISTOTLE service planning and delivery. Moreover, a synergy between the ChEESE tsunami component (UMA, NGI, INGV) is already in place and a PD2 prototype is already implemented in ARISTOTLE service. The “upgraded” service which PD8 would offer is being addressed in close cooperation between ChEESE partners and ARISTOTLE management.

The tsunami scenario uncertainty analysis capabilities offered by this PD are moreover of interest for the *scientific community*. PTF is indeed being considered for service provision in a future *EPOS Thematic Core Service* dedicated to tsunami science and tsunami disaster risk reduction. Several ChEESE partners are among the core group of European tsunami scientists which is being organized around the idea of founding the Tsunami TCS, and EPOS is part of the ChEESE IUB. Hence, the scene is fully set for a synergic service definition.

System design and architecture

PTF combines i) available data on seismic parameters (e.g. location and magnitude), considering uncertainty, with ii) geological and past seismicity information to constrain the unavailable source parameters (strike, dip, rake, slip distribution), and with iii) large ensembles of tsunami numerical simulations to be performed on the fly (Figure 5).

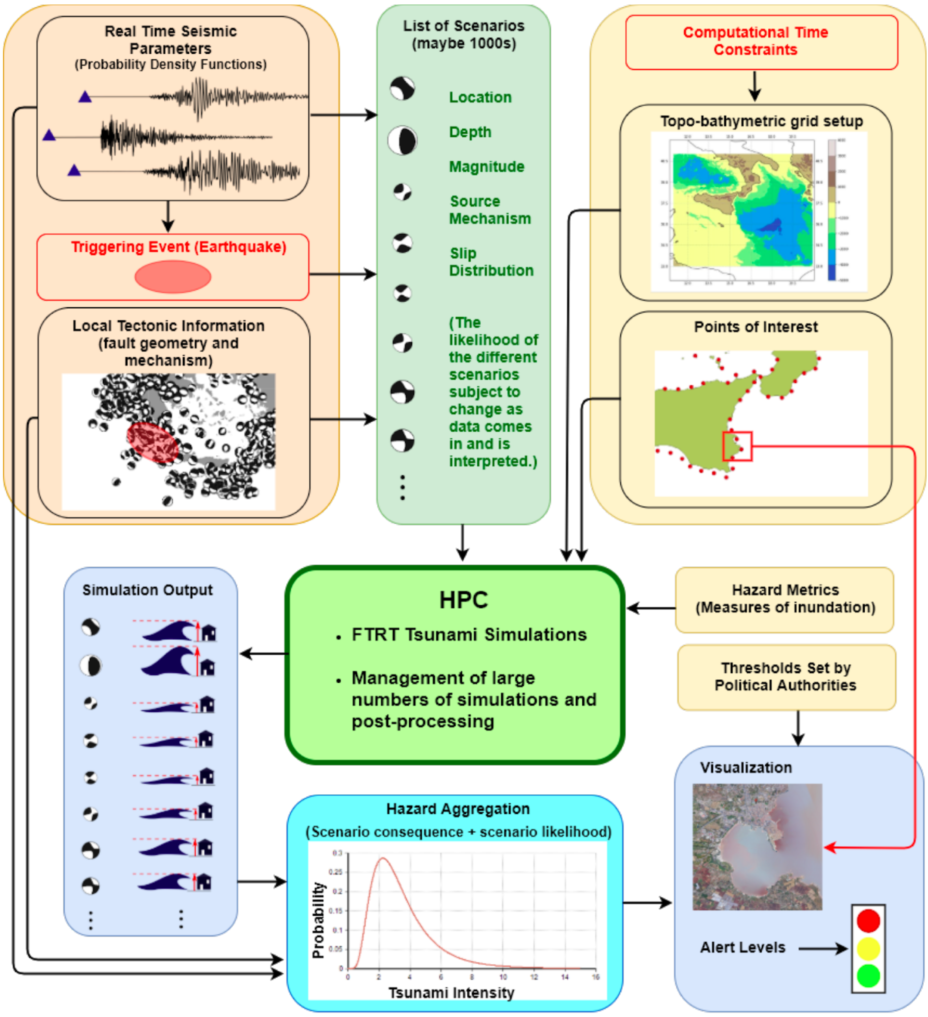

Figure 5. Schematic representation of the PTF workflow.

When an earthquake occurs, probability density functions of some earthquake parameters are estimated by the seismic analysis software. The probability of the different potential source realizations is then assigned according to their degree of consistency with the available data. The likelihood for some source parameters, which are not provided soon enough by the real-time monitoring system, can be estimated from the local historical and geophysical knowledge, such as the proximity to a subduction zone and local seismicity. This information can be used to constrain the likelihood of the faulting mechanism given the location and the magnitude of the earthquake. This process ends up with a weighted list of scenarios expressed by means of earthquake parameters suitable for initializing a tsunami simulation. Their weights represent the likelihood of each of them to actually represent and reproduce the earthquake which has just occurred. This part does not require HPC resources and it is used to generate the list of scenarios to be run on the three-level grid setup built up within the FTRT workflow. Resolution and size of computational grids are automatically tuned, using openly available bathymetric data. Source and grid inputs are then fed into the GPU Tsunami-HySEA simulations for a number of different scenarios; hence, the HPC part of the workflow needs to manage hundreds to tens of hundreds of concurrent and urgent runs. The outputs from the FTRT at all Points of

Interest is post-processed using approximate amplification factor method, to estimate inundation statistics. Results are aggregated and weighted according to scenario probabilities. The resulting tsunami probabilities are converted into alert levels based on pre-defined selection rules.

Use Cases

Regional information about faults and seismicity are necessary for defining the probabilistic seismic model used to integrate the real time seismic determinations. We then start with defining use cased for three regions, which are the Mediterranean Sea, East Japan, and a stretch of the Chilean subduction zone/coast. Part of the just awarded PRACE resources will be exploited for running these use-cases.

Use case	Rationale	Summary
Mediterranean Sea	Test the feasibility and the functionality of the PD8 prototype based on FTRT simulations and compare with the existing one based on precalculated scenarios	A hindcasting of a past event will be performed in the area; probability maps and tsunami coastal impact will be evaluated
Pacific Ocean (Chile/Japan)	Test the PD8 at global scale in an area hosting some significant past tsunamigenic earthquakes	A hindcasting of one (possibly two) past event will be performed in the area; probability maps and tsunami coast impact will be evaluated

Validation Procedures

These use-cases are part of a larger validation procedure to be accomplished within the TSU-CAST PRACE Project. TSU-CAST indeed foresees a 4-stage optimization, testing and validation exercise, that is:

1. To further optimize configuration of FTRT simulations and PTF workflows to enable many (e.g. several tens of thousands) concurrent FTRT simulations;

2. To enable the comparison between pre-operational in-house implementations of PTF based on pre-computed scenarios, obtained by combinations of unit-source simulations in the Mediterranean, with PTF supported by on-the-fly FTRT simulations;

3. To compare PTF through hind-casting on past events and prepare it to proper statistical testing based on its application to world-wide seismicity;

4. To test the sensitivity of the PTF results to compromises in terms of reduced and coarser- grained uncertainty exploration, due to potentially limited computational resources in day-by- day operations.

Functional requirements

Functional requirement	Rationale	Summary
Interface with real-time seismic/tsunami monitoring system	Needed to retrieve real-time seismic parameters (Hypocenter and Magnitude)	After a seismic event, the earthquake parameters will be

		used to trigger the ensemble of simulations
Access to TSUMAPS-NEAM database	Needed to retrieve long-term probability data concerning the source mechanism and the amplification factors	Long-term probability data are used to compose the ensemble of scenarios to be simulated; amplification factors are used to estimate tsunami maximum inundation heights from the simulated offshore values
Global regionalization, subduction models and probability distribution of earthquake mechanisms	Needed for global applications (outside the NEAM area)	
Global amplification factors	Needed for global applications (outside the NEAM area)	Amplification factors are used to estimate tsunami maximum inundation heights from the simulated offshore values
Global tide-gauge/DART locations, and data stream	Needed for comparison with tsunami observation for real events	
Topo-Bathymetric data	Needed for the numerical tsunami propagation and (potentially) inundation	Depending on the magnitude of the tsunamigenic earthquake, one-to-three domains at different scales and resolution will be defined to run the tsunami simulations
Ensemble of 10^4 to 10^6 simulations performed in parallel within 5-10 min up to 1 hour, depending on the end user	In the framework of urgent computing, the tsunami coastal impact probabilities must be computed soon after the seismic event	10^4 to 10^6 tsunami numerical simulations will be performed in the area of the seismic event at different spatial resolutions

Non-functional requirements

Non-functional requirement	Rationale	Summary
Interface for integration in the EW systems	Data need to be exchanged and visualized, to form the warning messages to be disseminated and to allow interoperability with other operational tools used by the early warning services	A suite of exchange/visualization functionalities are being implemented in the CAT-INGV and ARISTOTLE Decision and Alert Support Systems

Environments

Expected TRL for this service will be in the range of 6-8, because PTF PD8 is being tested and possibly deployed in a ***relevant environment*** in view of its application in an ***operational*** environment. As already mentioned in the section devoted to potential users, the service will be tested primarily at INGV where is sited the Italian Tsunami Warning Center - CAT (Centro Allerta Tsunami) - that acts i) as Tsunami Service Provider in the

NEAMTWS (since 2016) and ii) is operational at national level. It will be also considered for inclusion in ARISTOTLE operations. Its full exploitation is subject to availability of significant (exa-scale) HPC resources. In the absence of these resources, “lighter” versions are implemented. There is however a trade-off between uncertainty quantification capabilities - and possibly uncertainty size - and amount of available resources. Testing in the framework of the PRACE resources will allow to better address the extent of this trade-off, which is in turn necessary to address the effectiveness of downscaled PTF versions.

Access policies and IP issues

A Data Management Plan (DMP) is being prepared for PD8. Some main aspects of this DMP are reported below. Some features are yet To Be Defined (TBD).

Target TRL	Target user	Interface type	Software open access	Workflow open access	Products open access
TRL 3-4	Developers	Scripts and codes to be used both in non-HPC and HPC environments	YES	YES	YES
TRL 5-6	Field Specialists	Scripts and codes to be used both in non-HPC and HPC environments	YES	YES	YES
TRL7-8	Stakeholders	Integration into existing Decision and Alert Support Systems	TBD	TBD	TBD

Potential Ethical and/or Liability issues

This service is a tool for disaster risk reduction and for disaster consequence mitigation. It may be moreover used for alerting operations related to an impending disaster. Hence, ethical and liability issues are potentially all over the place. These ones should be carefully analysed for each specific **operational environment** where the Service may have the potential to become a TRL 9 one. This is the reason why some of the boxes in the table just above are in the TBD status.

Exploitation plan outlook

The exploitation plan was already sketched in the previous section. Two main end-user categories, the CAT-INGV representing early warning and the ARISTOTLE ENHSP representing expert support to post-disaster recovery, were identified. The testbed for tuning and validation is the PRACE Project TSU-CAST. This is expected to offer a clear view on the trade-off between available resources and uncertainty, which in turn will allow hardware design optimisation.

Computational resources

For PTF testing in the hind-casting mode, we have identified in TSU-CAST 6 cases of past tsunamis in the Mediterranean (with magnitude <=7.1), including the 1908 Messina and Reggio Calabria earthquake and tsunami, one of the most damaging tsunamis ever. We have estimated, with preliminary tests performed using the CAT-INGV PTF prototype, that dealing with the uncertainty on source parameters (location, magnitude, strike, dip, rake, fault size, slip distribution) for one of these events needs at least ~80,000 numerical simulations. We consider 3 calculation domains of progressively coarser spatial resolution but larger geographical extension (local, regional, and basin-wide), for a total of ~6x3x80,000=~1,440,000 simulations. We have also identified 4 cases in the Pacific Ocean with tsunamis caused by earthquakes of magnitude 8-9. Each of them requires ~600,000 simulations, repeated on the 3 domains, for a total of ~7,200,000 runs. We end up with ~8,640,000 simulations, which would require approximately 1,080,000 GPU hours if we assume a mean value of 450 sec / simul in a single P100 card. Part of these resources were allocated with TSU-CAST. We may downscale in different ways some of the above cases to try and achieve enough testing also with these reduced resources.

PD9. Seismic tomography

PD9	Seismic tomography
Leader	CNRS
Participants	INGV, ETH
Codes	SPECFEM3D, Salvus
TRL initial	4
TRL target	6

Service definition

The general objectives of PD9 is subsurface imaging by using seismic wave recorded at the surface. HPC allows to use more data, more precise modelling at higher frequency. This leads to better resolved tomographic images.

Different services are proposed, mostly addressed to the scientific community:

1. Provide an HPC workflow for subsurface imaging: from selection of data to computation on HPC cluster and also post-processing the results.
2. Publish computational results (i.e., the obtained subsurface models) and make them available in accessible repositories for:
    - visualization
    - scientific applications
3. Provide users with tools for reading and converting model in several format (e.g. to perform simulation inside the model, use for earthquake scenario simulation, source studies, interpretation).

Potential target users

For PD9, TRL 5-6 is targeted thus only experts are supposed to be able to use the software and data products. Both workflow and final results will potentially be made available. For workflow, experts are seismologist in academia and in industry (mining and oil companies), they are research engineers in the exploration or imaging departments. These experts will be able to run PD9 workflow on their own data set and manage to access HPC facilities. For results, the models should be of interest for expert geophysicist, volcanologist, geologist for further studies, like geodynamics interpretation, study of seismicity or using model for computing scenario for seismic hazard.

Scientific associations (IAVCEI and IASPEI) of the IUB have expressed their interest in PD9 outcomes. In particular, achieving high spatial resolution of subsurface tomography in volcanic areas would be of extreme interest for IAVCEI and require intense HPC effort. However, the issue of the density of receivers still limit the maximum achievable frequency.

Similarly, the importance of imaging of thin geologic structures such as faults has been raised by FAULTSHA, for evaluating seismic hazard for very rare events. This is of paramount importance of improving e.g. seismic hazards for Nuclear Power Plants in a broad range of frequencies and spatial scales.

Finally, accessible tomographic datasets can be of great interest for exploration companies, including some of those in the IUB (Repsol).

Service design and architecture

Use Cases

Use case	Rationale	Summary
Use case 1 : 3D Full wave form inversion with exploration geophysics OBS data set from site in north sea	- high quality OBS with 3 components - visco-elastic inversion - the size of computation is very big and only can be performed on bigger computer in the world	- we already have data and performed first tests. -We will perform hierarchical FWI from 4Hz to 12Hz
Use case 2 : Tomography of Japan	- high quality OBS + seismometers (buried instrument, very low noise) - high density of seismometers and high seismicity will lead to high resolution tomographic images	- We have acces to Japanese data. - we will choose the traget area.
Use case 3 : Tomography of Pyrenees	- several experiments where setup in Pyrenees, (PYROPE, OROGEN, IBERARRAY) with more than 200 seismic stations. - better understanding of geological structures of Pyrenees.	- we will perform teleseismic full waveform inversion. - we need to define the frequency to be used based on quality of data and station spacing.

Functional Requirements

Functional requirement	Rationale	Summary
Use case 1: Visco-elastic offshore exploration geophysics.	The state of the art is acoustic case only	The PD is developed in the visco-elastic case, a multiparameter inversion strategy is also developed
Use case 1: Offshore exploration geophysics case up to 12Hz	This frequency is reached in the state of the art but in the acoustic case only. For the visco-elastic case the necessary computational power increase by 1 or 2 order of magnitude.	The PD is optimized in order to be able to handle very big cases in GPU supercomputers.

Use case 1 : input raw data 1 To	3 components velocity dataset.	Using 3 components will allow us to resolve different physical parameters (Vp, Vs, density)
Check pointing	We store all iterations on disk.	By storing iterations (model and gradient) we are able to run the iterative process by several batches. This is useful if the wall time in computer is not enough.

Environment

The relevant environments are seismological laboratories and observatories for regional scale seismic tomography. Additionally, for exploration geophysics the oil and gas companies. The service will be tested on realistic use cases, in relevant HPC environments: on Marconi100 Tier0 computer in a PRACE project.

The validation on the selected use cases aims to prove the feasibility of the workflow, but for the time being and at TRL 5-6 the service is to be operated by experts. During the course of the validation the results will be shared with the users (e.g., scientific associations) in order to better define final outputs and products.

Computational Resources

The estimated required computational budget for the target Use Cases is:

Use case	# runs	# CPU cores/run	#GPU /run	GPU hours / run	Equivalent Total CPU M hours (PRACE convention)	Total memory (TiB)	Total disk space (TiB)
Use case 1 at 5 Hz	10 iterations	n/a	276	195	0.6	4.4	2
Use case 1 at 12 Hz	10 iterations	n/a	2760	6000	21	60	4
Use case 2	10 iterations	n/a	200	300	0.6	4	1

PD12: High-resolution volcanic ash dispersal forecast

PD12	High-resolution volcanic ash dispersal forecast
Leader	BSC
Participants	INGV, IMO
Codes	FALL3D
TRL initial	3
TRL target	6 - 7

Service definition

Objectives. The service objective is an Early Warning and high-resolution (space and time) forecast for atmospheric volcanic ash dispersal and ash fallout on the ground.

Scientific and technological advancements. Progress with respect the current state-of-the-art includes an ensemble-based forecast system with (satellite) data assimilation:

- Probabilistic forecast with uncertainty quantification.
- Use the Parallel Data Assimilation Framework (PDAF; http://pdaf.awi.de/trac/wiki). PDAF is an open-source software environment for ensemble data assimilation providing fully implemented and optimized data assimilation algorithms, in particular ensemble-based Kalman filters like LETKF and LSEIK.
- Potentially, model outputs can be crossed with flight data to anticipate flight disruption and/or rerouting.
- Volcanic ash and SO_2 detection algorithms applied to infrared observations taken by the SEVIRI sensor onboard Meteosat, which views the European region.

HPC added value

- Large number of ensemble members (related to better uncertainty quantification).
- High spatial model resolution on a continental scale (e.g. 4 km for a domain including continental Europe and Iceland).
- High frequency of results (15min as opposed to the 6h of current VAAC products).

Potential societal/economic impact

- Cost reduction to airlines and aviation stakeholders
- For ash fallout, prompt reaction during a crisis

Potential target users

Mitiga Solutions. It is already commercializing an early warning system that will be improved from the ChEESE developments. As stated in the GA; BSC foresees potential technology transfer to the spin-off Mitiga Solutions S.L. MITIGA Solutions S.L. (Mitiga; http://www.mitigasolutions.com) is a spin-off from BSC that develops and commercializes HPC solutions capable to evaluate and manage volcanic hazards and impacts on civil aviation, including operational volcanic ash dispersal forecasts. MITIGA Solutions is interested in providing services related to operational volcanic ash dispersal forecasts to clients (airlines, air navigation service providers) in Europe, Asia and South America

Icelandic Meteorological Office and Icelandic Civil Protection. Technology transfer is also expected towards IMO where is planned to install FALL3D on local machines and have the model running operationally as long as the other dispersal models currently available at IMO. FALL3D will be run daily for hypothetical eruptions and the module will be implemented in the current system (accessible at https://dispersion.vedur.is) to exploit a multi-model system. In case of eruptions FALL3D will be run at high resolution over Iceland to provide information to the meteorologists on duty for the identification of areas in the atmosphere that could be unsafe for aviation operations as well as on the ground (e.g. at the airports). FALL3D will use the high-resolution HARMONIE data as input and will be initialized with any observational data available from the monitoring network at IMO. Areas forecasted to receive high values of ash will be identified and used to prepare the Sigmets. Information provided by dispersal models run locally are essential to issue both Sigmet #1 (on average released after 15 minutes) and Sigmet #2 (on average released after 40 minutes). Safety Risk Assessment (SRA) becomes effective on the basis of the forecasted contaminated area as designed within Sigmet #2.

IGN. Interested in case of activity in Canary Islands.

Service design and architecture

Use Cases

Use case	Rationale	Summary
Icelandic volcanoes User: Icelandic Civil Protection	ICP is interested on a domain covering Northern Europe (regional) and a domain covering Iceland (local). The later one is intended for proximal ash fallout, whereas the other one is for long-range dispersal and affectation of civil aviation. At present, this is covered by the Volcanic Ash Advisory Centers (VAACs); i.e. London and Toulouse in the case of Europe. However, current resolution of operational products are not best-suited to civil aviation needs in terms of: model resolution, output frequency, and uncertainty quantification.	High spatial model resolution on a continental scale (e.g. 4 km for a domain including continental Europe and Iceland). Data assimilation from SEVIRI satellite data and measurements are provided at a nominal temporal resolution of 15 minutes (but can be provided at up to 5 minutes using EUMETSat's rapid-scan service) and offer the opportunity to track hazardous volcanic clouds in near-real time, 24-hours a day. Meteorological forecasts from HARMONIE (run operationally at IMO) and ECMWF (IFS ensemble forecast).

Italian volcanoes User: Italian Civil Protection	Domain covering Southern Europe. Same rational than above. For the Italian case, regional domain should cover southern Italy.	Same than above but using ARPA model for regional meteorological forecasts

Functional Requirements

Functional requirement	Rationale	Summary
Time to solution	Very relevant aspect to civil aviation. Updating of forecasts every 6-12h (as it is now) does not give prompt time to re-schedule or re-route flights.	Run forecasts within 1 hour whenever substantial changes occur in the eruption conditions. Forecast window up to 48 hours to cover tactical and pre-tactical flight design phases
Resolution	Current products from London VAAC Lagrangian model NAME are post-processed at 40 km resolution grid. Concentrations are determined by summing up the total mass of ash carried on model particles that are found in each grid box and dividing by the box volume. Outputs are produced every 6 hours for the forecast times 00:00, 06:00, 12:00 and 18:00 UTC, and at the start of a new eruption on the nearest hour, and for T+6, T+12, T+18 and T+24 after this. No data assimilation nor ensemble forecast to assess uncertainty.	We target at 10 km resolution for regional domains and 5 km for local domains. Output frequency 15 min.
Data Format	The official ICAO Volcanic Ash Advisory (VAA) and Volcanic Ash Graphic (VAG) products are produced by forecasters. These products are a human interpretation of available observations and the modelled output. Consequently they may not match the raw modelled areas.	Provide raw data (e.g. in netCDF format) to be further tailored to user needs. For example, this could be automatically crossed with flight plans to individually assess disruption based on some concentration threshold or engine dose. This approach is already been done by the IUB Mitiga Solutions.
Checkpointing	Results for individual ensemble members will be stored	Workflow will need to check (and resubmit) individual ensemble members on failure
Dimension of the statistical ensemble (number of simulations)	At present, no product exists based on ensembles. This prevents to "attach" an uncertainty index to deterministic forecasts. Ensembles would also allow for	Service using > 30 members for UQ and probabilistic products. Ensemble members obtaining from perturbation either the initial condition (e.g. data inserted from a satellite

	probabilistic forecast, although this requires of a good understanding from the final user; used to think in binary terms (e.g. fly/no fly).	retrieval), wind field and poorly-constrained Eruption Source Parameters (typically column height and eruption duration).

Environment

The relevant environments are volcano observatories and research Institutes. The service is tested on realistic use cases, in relevant HPC environments: on MareNostrum-4, Joliot-Courie (in the context of a call 20 PRACE project), and IMO and INGV clusters (reduced resolutions due to computational constrains).

The validation on the selected use cases aims to prove the feasibility of the workflow, but for the time being and the service is to be operated by experts. During the course of the validation the results will be shared with the users (e.g., Civil Protection authorities) in order to better define final outputs and products.

Computational Resources

The estimated required computational budget for the target Use Cases are:

Use case	# runs	# CPU cores/run	#GPU /run	Core hours / run	Total CPU k hours	Total memory (TiB)	Total disk space (TiB)
Iceland and Italy	30 per forecast cycle	1024x30 (30 members)	-	0.5	15	Fits in nodes	~100 GiB

A1. Table 1. ChEESE Technology Readiness Levels

Level	TRL Definition (H2020)	TRL for ChEESE Pilot Demonstrator	Target	WP
TRL 1	Basic principles observed	Equations and Quantities of Interest (QoI) defined. HPC parallel numerical code deployed.	Developers	n.a.
TRL 2	Technology concept formulated	Efficiency metrics defined and strategy towards *exascale* preliminarily assessed.	Developers	2,3
TRL 3	Experimental proof of concept	Proof of scalability and efficiency of individual components. Validation performed on benchmark problems.	Developers	2,4
TRL 4	Technology validated in lab	Proof of models/data interoperability in workflows. Validation performed on benchmark problems.	Developers	3,4
TRL 5	Technology validated in relevant environment (industrially relevant environment in the case of key enabling technologies)	Individual components of the prototype service demonstrated in relevant environments. Error checking.	Field specialists	2,4,5
TRL 6	Technology demonstrated in relevant environment (industrially relevant environment in the case of key enabling technologies)	Component integration and interoperability. Use-case tests demonstrated in relevant environments. Assessment of required resources.	Field specialists	3,4,5
TRL 7	System prototype demonstration in operational environment	Individual components of the prototype service demonstrated in operational environments. Training of users.	End-users	5
TRL 8	System complete and qualified	Component integration and interoperability. Use-case tests demonstrated in operational environments. Assessment of Ethical aspects.	End-users	5
TRL 9	Actual system proven in operational environment (competitive manufacturing in the case of key enabling technologies; or in space)	The service is deployed and operation-ready for different scenarios. Agreement on Ethical and Liability issues.	End-users	n.a.

A2. Table 2. IUB interest in ChEESE Services

IUB member	PD1	PD2	PD5	PD6	PD7	PD8	PD9	PD12
Italian Civil Protection Department	X	X	X	X	X	X		X
Icelandic Civil Protection	X		X	X				X
Spanish Geographic National Institute		X		X				X
IASPEI	X		X				X	
IAVCEI				X			X	X
EPOS	X		X		X	X		
GW4 ISAMBARD								
GEO-GSNL	X		X	X				X
REPSOL							X	
Schlumberger			X				X	
INTEL			X				X	X
MITIGA Solutions	X	X	X	X	X	X		X
Global Parametrics	X		X					
ARM	X		X				X	
Fault2SHA			X				X	
Modeling Collaboratory for Subduction		X	X	X	X	X	X	X
VECMA								
ARISTOTLE (ERCC)	X	X		X		X		X
Norwegian Water Resources and Energy Directorate		X			X	X		
AuScope								
National Computational Infrastructure Australia								

color	IUB type	Target PD
	Scientific association	TRL5-6
	EU project	
	Consortium	
	Operational governmental agency	TRL7-8
	IT private company	
	Risk/Insurance private company	
	Oil & gas private company	