Ref. Ares(2022)2394477 - 31/03/2022

D5.2 Documentation and User's Guides on Pilot Demonstrators

Version 1.2

Document Information

Contract Number	823844
Project Website	https://cheese-coe.eu/
Contractual Deadline	31/03/2022
Dissemination Level	Confidential
Nature	Other (Demonstrator)
Coordination and Editing	Tomaso Esposti-Ongaro (INGV)
Contributors	Marta Pienkowska (ETH), Laura Sandri and Beatriz Martinez Montesinos (INGV), Finn Lovholt and Steve Gibbons (NGI), Arnau Folch (CSIC), Leonardo Mingari (BSC), Jorge Macías and Carlos Sánchez-Linares (UMA), Vadim Monteiller (CNRS), Alice Gabriel (LMU), Josep de la Puente and Marisol Monterrubio (BSC), Stefano Lorito and Manuela Volpe (INGV)
Reviewers	Marisol Monterrubio (BSC), Joan Farnos (BSC)

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
1.0	Initial draft
1.1	Version revised by all authors
1.2	Includes reviewer's comments

Table of Contents

Introduction and structure of the report

ChEESE Pilot Demonstrators (PDs) have been developed in WP5 as potential services. The purpose of this Deliverable is to collect the documentation necessary to apply ChEESE PD's workflows to other use-cases. Application to different use-cases can be subject to constraints depending on the achieved TRL and the accessibility options available for the codes and workflows.

Target of this documentation are scientific experts and specialists in the field.

For each Pilot Demonstrator in WP5, we report in the following sections:

- The problem under study and the type of application, including scientific motivation, objectives and potential societal impact;
- The workflow, limiting to its basic components, potentially referring to documentation publicly available on the Zenodo ChEESE Community Repository.
- The required input data, allowing users to evaluate their availability for the use-case of interest.
- Output data, and their Level according to the EPOS data taxonomy:
    - **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
- The modality and policy to access the codes/workflows, having in mind the potential users (scientific community or stakeholders); in particular, specification if the workflows will be accessible via Virtual or TransNational Access.
- An estimate of the computational resources required to run a use-case for different target problems, in order to facilitate the planning of a new application.

Access to ChEESE HPC codes and workflows as a services

The purpose of ChEESE WP5 is to demonstrate the maturity of some HPC Pilot Demonstrators to be provided as a service to the scientific community (TRL 5-6) or to public stakeholders for operational services (TRL 7-8-9). Such an effort has already been documented in Deliverables 5.3 and 5.4.
Access to the developed services requires the definition of the policies, including the Intellectual Property rights and the definition of a plan for the long-term sustainability, which go beyond the scopes of the ChEESE project. However, with the development of Pilot Demonstrators and service prototypes, ChEESE candidates to establish a European e-infrastructure giving access to the newest and most advanced Solid Earth HPC codes and workflows in Europe as services. Virtual Access (VA; i.e., access through the web without any selection) will be provided to open-source codes and numerical engines through the ChEESE repository documented in Deliverable D4.13 and hosted on Zenodo (www.zenodo.org/communities/cheese-coe). TransNationalAccess (TNA) to the workflows will be in part provided through the Geo-INQUIRE project recently funded by the EC. Other modality of access will be defined upon the needs and requests by the end-users and stakeholders.

Summary

The aim of this Deliverable, in conjunction with D4.13 (Repository of numerical codes), is to improve and facilitate the access to computing applications and expertise that enables researchers and industry to be more productive, leading to scientific excellence and economic and social benefit.
The documentation will be made publicly available at least for those Pilot Demonstrators aiming at providing Trans National Access to the workflows as services.

PD 1. Urgent Seismic Simulations

PD 1	Urgent Seismic Simulations
Leader	Marta Pienkowska (ETH)
Participants	Juan Esteban Rodriguez (BSC), Marisol Monterrubio Velasco (BSC), Josep de la Puente (BSC), Andreas Fichtner (ETH)
Workflow	UCIS4EQ
Numerical Engine	Salvus
TRL achieved	6
Access Mode	developers only

Description of the pilot use-case

PD1 addresses the scientific problem of near real-time seismic simulations that require exascale computing resources.

Deadly earthquakes are unpredictable and relatively rare but have a high socioeconomic impact. Estimating this impact is crucial for hazard preparedness and mitigation, yet doing so from historic events alone remains difficult. With large earthquakes being so rare we often lack sufficient data to constrain any empirical relations, rendering extrapolations to future events imprecise. Physics-based numerical simulations and synthetic data analysis, on the other hand, are powerful tools that can provide detailed shaking information and contribute to outlining the impacts of such events.

Seismic wave propagation, however, is currently computationally prohibitive at high frequencies relevant for earthquake engineering or for civil protection purposes (up to 10 Hz). Time to solution for local near-field simulations accounting for frequencies above 1 Hz, as well as availability of substantial computational resources pose significant challenges that are incompatible with the requirements of decision makers. 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. Estimating such uncertainties is non-trivial from a scientific standpoint.

Seismic simulations for impact assessment therefore meet the three principal requirements of Urgent Computing:

1. The computations need to be performed within a specific time frame required by decision makers. The results are only valuable if achieved within the given strict time constraint.
2. The earthquake onset is unpredictable, and thus so is the computation trigger.
3. The simulations are highly computationally intensive and require significant high performance computing (HPC) resources.

Developments of computational HPC infrastructures will render routine executions of high frequency simulations possible, enabling new approaches to assess seismic hazard - such as Urgent Seismic Computing. Urgent seismic simulations within hours after an event could potentially deliver accurate synthetic short-time reports of the consequences of moderate to large earthquakes, including reliable error estimates.

Within PD1 we have developed a workflow prototype, Urgent Computing Integrated Services for Earthquakes (UCIS4EQ), in order to explore how deterministic modelling of ground motions can contribute to the short-term assessment of seismic hazard in the immediate aftermath of an earthquake. The workflow performs full end-to-end runs, from identifying an earthquake as an urgent computing candidate, through the definition of input parameters, job submissions and HPC simulations, all the way to post-processing and the visualisation of results. At TRL 6 the first working version has been tested in the relevant environment, but it remains in an extensive testing phase and thus is available only to the developers.

Workflow structure

The UCIS4EQ workflow consists of containerised microservices - sub-pieces of code that each tackle a different task - guaranteeing portability and adjustability. The interactions between the tasks are currently orchestrated by a micro-service that emulates a workflow manager, with the implementation of a workflow manager under way. UCIS4EQ is designed to ultimately become a hybrid in-production service that runs in the cloud and on HPC clusters.

The current release of the UCIS4EQ includes the following building blocks:

- **AlertService:** The alert service passively listens to information coming in from FDSN databases in real time and decides which (if any) information needs to be calculated. It triggers the workflow should the event be classified as an urgent computing candidate.
- **EventDomains:** The domain around the earthquake is defined for the simulation
- **ComputeResources:** The computational resources are estimated.
- **SourceParameters:** The computation of the source parameters is subdivided into a few steps. First, the workflow rapidly computes a set of CMT parameters in the region given information databases from past earthquakes in that same area (**CMTCalculation**). Then, it estimates the fault size and asperities size (**SourceType**) and populates the fault with a kinematic description of the slip via the Graves Pitarka rupture generator (**SlipGenGP**).
- **InputParametersBuilder**: All information is collected to generate a single YAML file with input parameters.
- **SalvusPrepare**: The inputs generated by the workflow are processed for a simulation with Salvus, locating sources and receivers on the specified mesh, attaching absorbing boundaries, and defining output types, among others.
- **SalvusRun**: The HPC wave propagation simulations are performed on the cluster.
- **SalvusPost**: The post-processing module collates results, generates netCDF files with ground motion proxies, and then automatically generates plots on the cluster. The plots are then compressed and immediately uploaded to B2Drop at the BSC and can be viewed in the User Portal.

Required input data

Input parameters need to be integrated for a given region in order to enable the service for that region. The inputs need to be ready on the designated HPC cluster.

Input Data	Format	Specific requirements	Accessibility
3D velocity model	HDF5 or netCDF	The current workflow implementation accepts models only in lat-lon coordinates, not UTM. Required for the mesh.	Via the Collaborative Seismic Earth Model (see https://cos.ethz.ch/research/CSEM.html)
Topography and bathymetry	HDF5 or netCDF	The current workflow implementation accepts topobathymetric data only in lat-lon coordinates, not	e.g. Earth2014 global topography (relief) model https://www.bgu.tum.de/iapg/forschung/topographie/earth2014/

		UTM. Required for the mesh.	(not limited to a single source)
FDSN databases	QuakeML (XML like format)	The workflow automatically monitors data from FDSN databases to eventually trigger simulations.	Available via https://www.fdsn.org/services/
Catalogs of past earthquakes	csv, XML	Required for statistical rapid CMT estimates for a given region.	https://www.globalcmt.org/CMTsearch.html

Table 1.1. Input data required for Seismic Urgent Computing simulation workflow.

Output data

Final processed data - such as final ground motion parameters and maps that gather information from the entire simulation ensemble - can be accessed via B2DROP. Raw time series or individual ground motion parameters (per-simulation processing) require access to the HPC cluster.

Output Data	Format	Number of files	File size	Level
Raw time series	HDF5	1 per simulation	~ 1 GB per file	Level 0
Ground motion parameters	netCDF	1 per simulation and 1 gather	~ 3 MB per file	Level 1
Maps of ground motion parameters	PNG	~20 (specific ground motion parameters can be requested by the user)	~ 2 MB per file	Level 1

Table 1.2. Output data produced by Seismic Urgent Computing simulation workflow.

Modality of accessing the codes/workflows

At TRL 6 the first working version has been tested in the relevant environment, but UCIS4EQ remains in an extensive testing phase and thus is available only to the developers. Access and further developments are ensured under collaborative actions and via further European projects (e-Flows4HPC and the upcoming DT-GEO). An example of post-processed results in netCDF can be found on the Zenodo ChEESE Community repository (https://doi.org/10.5281/zenodo.6381801).

The future service is designed to be deployed for the users via ready docker images and access to the code will not be required. A User Portal developed as a dashboard running in the browser allows users to interact with the service and monitor the progression of jobs on the cluster, while the post-processed results can be accessed via B2DROP (or another data server) once the service execution finishes. The User Portal and the automatic B2DROP uploads have been showcased in the PD1 live demonstration (see D5.4) where we have shown a full end-to-end execution that does not require the end users to access the code or to directly use the HPC cluster.

Computational resources

For the urgent seismic computing use case, the target size of the problem depends both on the frequency resolution (as indicated in the table) as well as on the defined region size. The number of required runs is also region dependent, so the 100 stated in the table is just a rough indication. In the current implementation the number of CMT solutions that define the number of runs varies according to the outputs of the CMT estimator which is region-dependent (as it is based on the historic events catalogues, see ‘Required input data’ above).

Target architecture	Node specs	Target size	Total number of nodes per use-case	Nodes per run	Number of runs	Duration of a single run	Data storage needs (outputs)	Total wallclock time GPUs
Piz Daint	Intel Xeon CPUs plus NVIDIA Tesla P100 GPUs	Low resolution (fmax = 1Hz)	200	2	~100	25 min	~ 1 GB	~ 41.5 h
		Medium resolution (fmax = 2Hz)	800	8	~100	34 min	~ 10 GB	~ 56.5 h
		High resolution (fmax = 5Hz)	~9000	90	~100	90 min	~ 100 GB	~ 150 h

PD2 Faster Than Real Time Tsunami Simulations (FTRT)

PD 2	Faster Than Real Time Tsunami Simulations (FTRT)
Leader	UMA
Participants	INGV, NGI
Workflow	FTRT Workflow
Numerical Engine	Tsunami-HySEA
TRL achieved	8
Access Mode	Trans National Access within Geo-INQUIRE project

Description of the pilot use-case

PD2 aims at producing much faster than real time simulations in the context of the Early Warning and Urgent Computing frameworks. On the 22nd of November 2021, a Live Demo for ChEESE PD2 was performed. This Live Demo addressed a use-case dealing with Early Warning, in particular in a set-up proposed by the IGN, the Spanish National Tsunami Warning Center. The aim of this section is to describe all the required data, inputs and resources needed to perform an exercise as the one we presented on the 22nd of November 2021 as a Live Demo for the Pilot Demonstrator 2 on Faster Than Real Time (FTRT) tsunami simulations in the framework of TEWS (Tsunami Early Warning Systems).

Different implementations of computational systems in real time using the Tsunami-HySEA code are already working in operational or pre-operational mode in several TEWS around the world. In particular, at the CAT-INGV (Italian TEWS) or in the Spanish IGN, but also in other tsunami systems as UNESCO GDACS (Global Disaster Alert Coordination System) at JRC or the Tsunami System for the Aristotle project. Currently, we have the project to integrate ChEESE flagship code Tsunami-HySEA in the SIFT (Short-term Inundation Forecasting for Tsunamis) system of PMEL/NOAA. But, these systems are implemented for producing a single kind of most-probable scenario simulation or, at most, few simulations limited by the computational resources available and the time restriction.

PD2 Live Demo tried to go a step forward and aims to include part of the variability associated with the uncertainty in the seismic source model parameters, magnitude and initial estimation in location . For this purpose, a reference scenario is considered and from this reference scenario aset of 135 associated scenarios is considered, varying 5 hypocenter locations, and 3 magnitudes, strike and dip angles. Several computational domains are also considered with different numerical resolutions, in order to provide different responses (in terms of accuracy) at different times, with the first level of alert being provided in one-minute time. The aim is to provide alert levels along all the Spanish coasts affected by the generated tsunami.
To perform this Live Demo within the timeframe required, suitable computational resources must be used. In our case 270 GPUs NVIDIA V100 in dedicated mode were used, and the computational resources provided by CINECA.

Workflow structure

Currently, when the Spanish National Seismic Network locates an earthquake with the capacity to generate a tsunami, a simulation of propagation of the tsunami is run, and, in less than 5 minutes, the arrival times and wave heights at forecast points are obtained. The present use-case proposes producing hundreds of FTRT tsunami simulations within minutes, i.e., remaining fast enough to use the numerical results for Early Warning. This requires quick access to suitable supercomputing resources, as the ones mentioned above. The general workflow for the PD FTRT Tsunami Simulations is shown in Figure 1 below. It comprises four main steps: (1) Scenario definition; (2) Performing the numerical simulations using Tsunami-HySEA code; (3) Generation of the raw outputs; and (4) Quick on-the-fly analysis of the numerical results and generation of the final output, in this case alert levels in coastal segments.

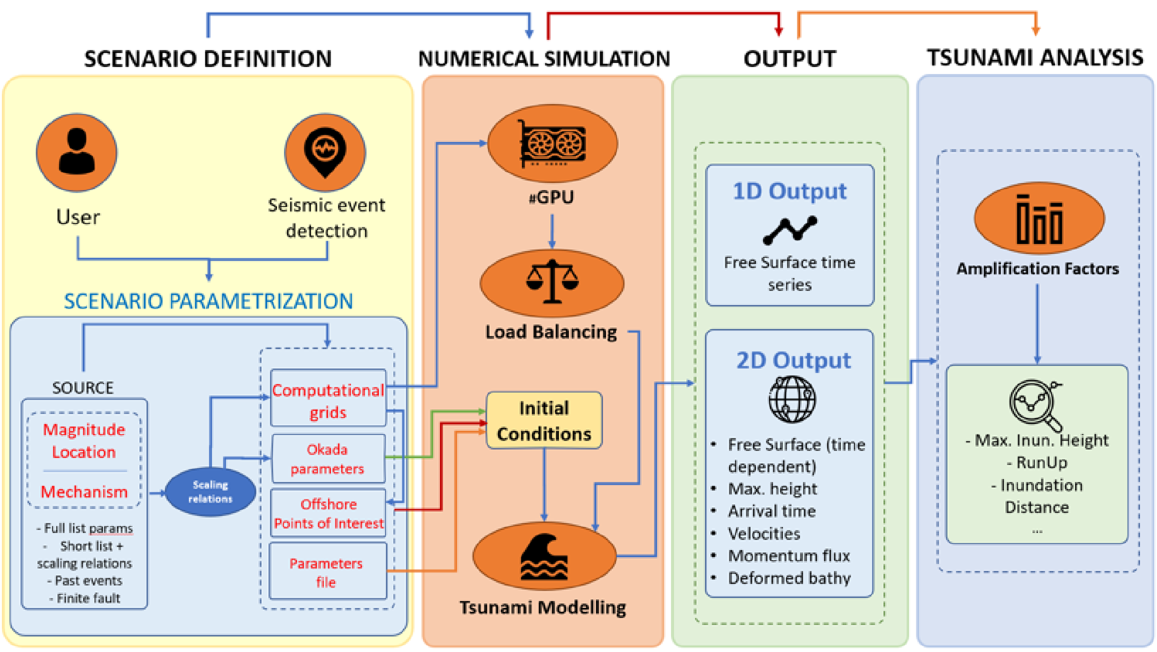


Fig 2.1. FTRT workflow scheme

The proposed approach in the Live Demo, allows addressing the problem of quantifying the uncertainty in the parameters that define the reference scenario, in opposition to the methodology that is based on considering a single deterministic (worst-case or not) scenario. In the way proposed here, hundreds of equally probable events are simulated, considering the variability in some parameters defining the seismic source, including magnitude and hypocenter location of the earthquake. After the simultaneous simulation of hundreds of scenarios, a post-process is carried out, also in real time, which analyses the information from all the simulations and offers a broader perspective on the worst possible scenario given that variability taken on the scenario reference.

The service implemented at IGN consists of the following parts (as shown in Figure 2):

- In case of a real earthquake with the potential of generating a tsunami, the former is detected, located and its magnitude is estimated by the Seismological Service at IGN. According to their dataset, a catalogued fault is associated with the undergoing event and the corresponding Okada parameters for the reference scenario are set. To address the uncertainty in the seismic tsunami source, a variability in parameters such as location, magnitude, strike and dip is considered, which will depend on the area in which they occur. This generates a list that can comprise hundreds of events that will be considered equally probable.

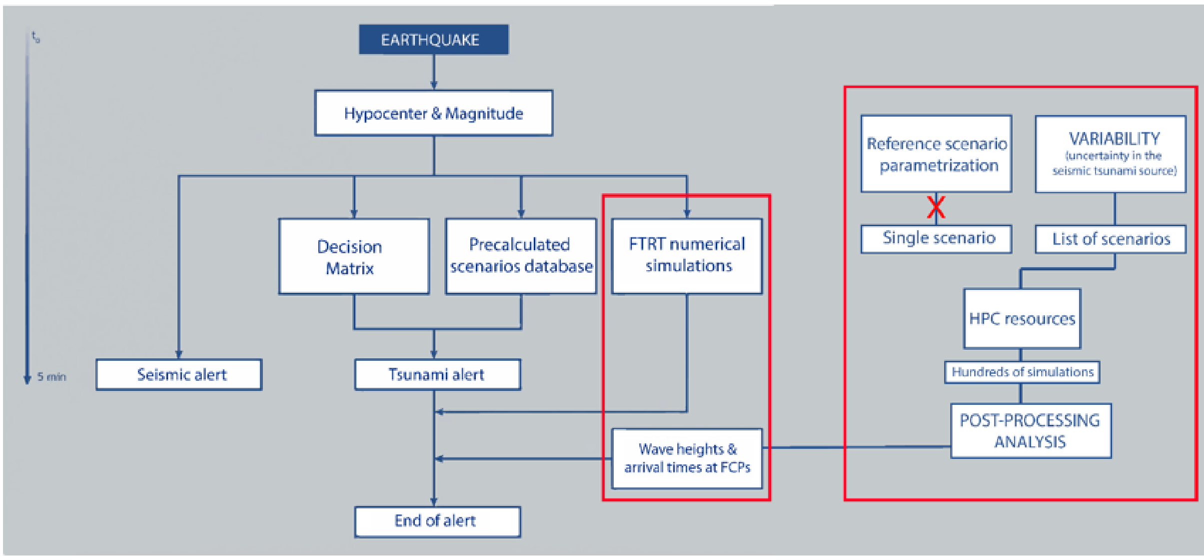


Fig 2. Tsunami simulation component (in red) in the Spanish TEWS.

This task corresponds to the stakeholders (in this case, IGN) who have the commitment to provide the list of scenarios to be simulated. In the same way, it is the joint task of the involved parties to contemplate the necessary agreements for the reservation of high-computing resources, in order to integrate the next step of the workflow.

- The simulations will be carried out simultaneously. Different areas of interest will be used as well as different resolutions. This is intended to give a first response in just 1 minute for low resolution domains and in just 7 minutes for high resolution domains. This allows the alert messages to be updated at the same time as the higher resolution simulations are completed. Runtimes are determined based on the area. For the Spanish Warning System in the Atlantic area, it has been tested in the 22nd of November 2021 Live Demo.

An important aspect in this branch of the workflow is the prior coordination between the stakeholders and the tsunami simulation provider, since it is necessary to define the configuration of the problems to be dealt with, which will depend on the geographical area, the available data, etc.

- Finally, as each batch of simulations concludes, the execution script sends a message to the user indicating that the job has finished, and that the post-processing of the results is carried out. A few seconds later, the summary of the analysis of the results are obtained, and they are sent back to the IGN, providing alerts levels in coastal segments and automatic message generation to include in the Tsunami Warning system.

With this implementation, we provide a quick response for tsunami action, incorporating not only deterministic simulations of a single event but also including variability in the parameters that define the source of the tsunami, showing that the use of supercomputing resources allows the realisation of hundreds of tsunami simulations in just minutes, allowing the analysis of the results to be incorporated into the Tsunami Early Warning System.

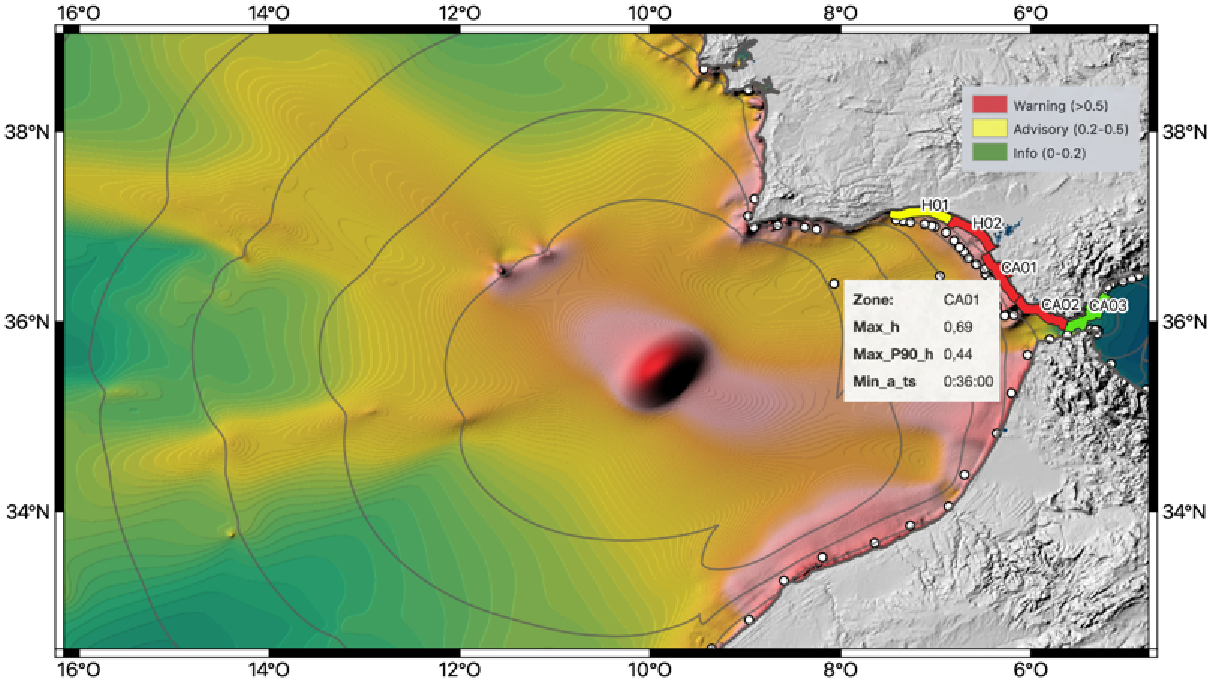


Fig 3. Tsunami warning map obtained from the results of the 135 simulations performed by including variability in the parameters of the seismic source model from the reference scenario. Forecast points and target areas are colored by alert level.

General requirements

The massive launch of simulations with the aim of carrying out either Early Warning tasks, Urgent Computing simulations or PTHA will necessarily require supercomputing resources, access to a large number of GPU units to be able to launch the simulations simultaneously in short computing times.

1. ***Software installation***

   The service runs on a Linux environment where OpenMPI, parallel NetCDF and CUDA libraries are necessary. From the hardware side the only requirement is a NVidia GPU.

   We will provide the source code of Tsunami-HySEA MonteCarlo through the Zenodo ChEESE repository (www.zenodo.org/communities/cheese-coe).

1. ***Input data requirements***

   Tsunami numerical simulations using Tsunami-HySEA require the inputs that we detail below:

   - *Bathymetry files*

     This file contains the points (x,y) of a given region along with the elevation z (either water or land) of each of those points.

     The format of this file must be .grd. This is a raster-type format used in the main Geographic Information Systems. The data type is Float64 - 64-bit floating point number,

and the description of the GDAL driver is Network Common Data Format (netCDF3 CLASSIC (32-bit offset) storage format).
These input bathymetry files must have longitude and latitude variables in double numerical precision, and the z variable (the bathymetry) in single precision.

- *Parameters input files*

This file contains the input data to run the model with the initial condition of the user choice. Within this file the characteristics of the event that produces the tsunami will be established, as well as the simulation time, the border conditions, file with points for saving time series, the different coefficients intrinsic to the model, among others.

- *File for managing the number of GPUs used*

This file .bin is generated by running the script get_load_balancing (included in the service application) applied to the parameter file, indicating how to divide the computational domain, thats depends directly on the bathymetry, into the available GPUs and the number of processes to carry out per domain.

Required input data

The idea here is to make users aware of the requirements for running a use-case, in terms of input data. Take as reference the use-cases described in D5.3 and/or D5.4

Input Data	Format	Specific requirements	Accessibility
Grids for computing in the different domains considered	grd	As indication, we used 4 grids with 3 numerical resolutions (½, ¼ and ⅛ arcmin)	https://data.mendeley.com/datasets/cpgf3c5dtx/1
List of scenarios	Text file in a given format (see Mendeley data)	Okada parameters for each scenario	https://data.mendeley.com/datasets/cpgf3c5dtx/1
File for managing the number of GPUs used	Binary	Generated by get_load_balancing	
Launching script			

Modality of accessing the codes/workflows

- Trans National Access to the full workflow will be granted by the Geo-INQUIRE project.
- Tsunami-HySEA model description can be found in the web page https://edanya.uma.es/hysea/index.php/models/tsunami-hysea

Scaling of computational resources

Target architecture	Node specs	Target size	Total number of nodes per use-case	Nodes per run	Number of runs	Duration of a single run

Marconi 100 @CINECA	Nodes equipped with NVIDIA Ampère GPU	small	16 nodes x 4 GPUs 64 GPUs	Each simulation is run in one GPU	135	
		Medium (use-case)	68 nodes x 4 GPUs Total 272 GPUs	idem	540 (135 at 4 different domains and resolutions	1 to 7 min
		large	272 nodes x 4 GPUs Total 1088 GPUs	idem	2160	

References

B. Gaite, J. Macías, J.V. Cantavella, C. Sánchez-Linares, C. González, L.C. Puertas (2022). Analysis of Faster-Than-Real-Time (FTRT) tsunami simulations for the Spanish Tsunami Warning System for the Atlantic. Submitted to GeoHazards.

NB: All the data generated and analysed in the previous study can be found at Mendeley repository: https://data.mendeley.com/datasets/cpgf3c5dtx/1.

Live Demo recording. https://www.youtube.com/watch?v=rkruUHAaleA

The Live Demo in the ChEESE project web page:
https://cheese-coe.eu/media/news/cheese-conducts-live-demo-faster-real-time-tsunami-simulations

PD 5. Physics-Based Probabilistic Seismic Hazard Assessment

PD 5	Physics-Based Probabilistic Seismic Hazard Assessment
Leader	LMU
Participants	IMO, BSC, TUM
Workflow	PSHA_WF
Numerical Engine	SeisSol, Cybershake
TRL achieved	6
Access Mode	SeisSol (open access), Cybershake (by request), Trans National Access within Geo-Inquire project (starting Oct 2022)

Description of the pilot use-case

Probabilistic Seismic Hazard Assessment (PSHA) is the process of quantifying the rate (or probability) of exceeding various ground-motion levels at a site (or a collection of sites mapping a region) given all possible earthquakes. PSHA is a highly active field of research in the scientific community as well as a key component of operational risk and hazard assessment. Classical PSHA relies on seismic observations and empirical models: empirical ground motion models (GMMs) quantifying the probability of exceeding a threshold shaking intensity depending on distance to the earthquake fault and earthquake source parameters; local earthquake catalogues and fault slip rate proxies.

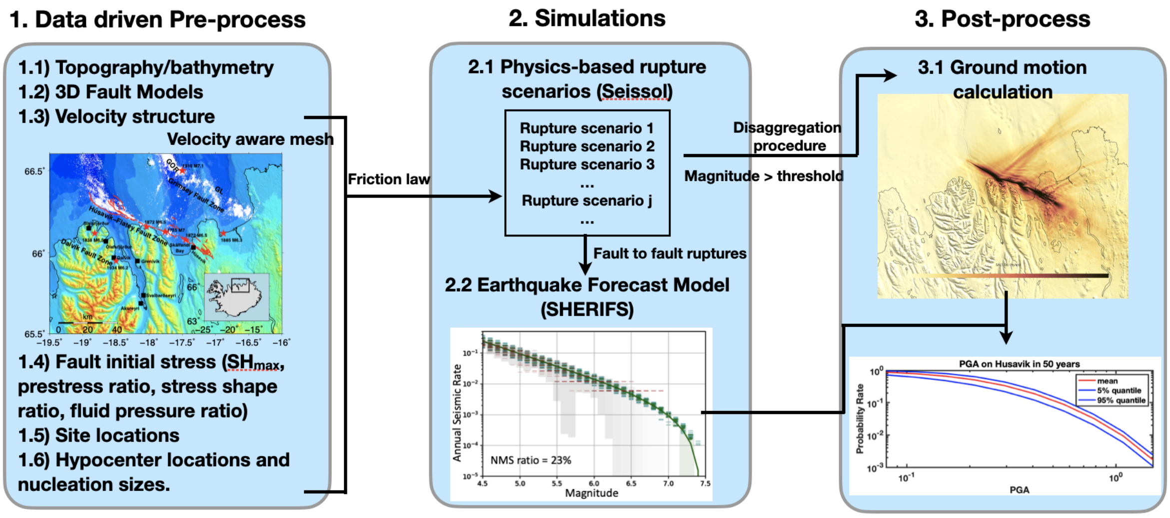
Physics-based and HPC empowered PSHA (p-PSHA) allows the analysis of multiple forward models encapsulating the non-linear coupling of source, path and site effects that conjunctively affect ground shaking. Realistic model setups may acknowledge topography, 3D geological structures, rheology, and fault geometries with appropriate stress and frictional parameters, all of which possibly affecting ground motions. In addition, such fully dynamic multi-physics earthquake simulations can, for example, constrain dynamic and static stress transfers by the naturally incorporated physical limit to “rupture jumping” and possible 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 established workflow of physics-based PSHA provides frameworks for validating and integrating multiple forward simulations of varying degree of complexity - ranging from simple logic-tree approaches to kinematic source modelling to multi-physics dynamic rupture scenarios. The simulations are analysed using a disaggregation procedure, since all earthquake scenario characteristics are not prescribed but evolve in each simulation based on first-order physical principles. To automatically detect scenarios which spontaneously develop into large, hazardous events and fulfil predefined criteria (specifically in terms of moment magnitude) for consideration in PSHA, we developed post-processing tools to automatically visualise outputs, calculate ground motion characteristics (peak ground displacement, velocity and acceleration, PGD, PGV, PGA and spectral acceleration at varying periods T, SA(T)) for selected sites, or generate shake maps for the simulation domain. Then scenarios based shaking intensity maps are combined with a logic-tree based estimation of seismic rates, yielding physics-based PSHA. In addition, the developed workflow also provides a mechanically possible worst scenario. Physics-based GMMs in the study region may also be derived.

There are two use-cases for which we applied the ChEESE physics-based PSHA workflows: 1. the Húsavík-Flatey Fault Zone (HFFZ) in Northern Iceland (SeisSol focused), and 2. the Reykjanes Peninsula Oblique Rift (RPOR) and Southern Iceland Seismic Zone (SISZ) (Cybershake+AWP-ODC focused). Detailed steps are documented in the following section.

Workflow structure for SeisSol p-PSHA

The SeisSol focused p-PSHA workflow presented below is based on a use case in the Húsavík-Flatey Fault Zone (HFFZ), in Northern Iceland. It can be easily adapted to other regions.



The workflow includes three main steps:

1. **Data-driven pre-processing**

   Using tectonic, geodetic, seismic and other available data, input files for the model simulations are prepared and constrained.

   1.1 Topography & bathymetry

   Incorporating local high resolution topography/bathymetry data is recommended. If not specified by the user, openly available data from GeoMapApp (http://www.geomapapp.org/), or GEBCO (https://download.gebco.net/) can be incorporated

   1.2 Fault Models

   Depending on the data available and its uncertainties, one well constrained model or multiple models considering different levels of uncertainties and complexities, can be used in the workflow. For the HFF case, four alternative fault models have been constructed:

   Model 0: the HFFZ is modelled as straight, linear faulting segments without the zones having distinct earthquake occurrence potential (Einarsson 1991).

   Model 1: Same as Model 0, except the zones now have specific and different seismic potentials, and the HFFZ is considered to be divided into three non-connected fault segments, each of different maximum magnitude potentials. Moreover, the seismicity rates on the segments are specified (Björnsson et al. 2007; Sigbjörnsson and Snæbjörnsson 2007).

   Model 2: The HFFZ is considered to be further segmented on the basis of fault mappings and seismicity distributions.

   Model 3: Same as Model 2, but with additional fault segments identified based on precisely relocated seismicity, including a total of 55 fault segments embedded in a recent 3D velocity model. Model 3 gives insight into a more comprehensive next generation fault model that would utilise the zones of concentrated relocated seismicity to reveal faulting structures in more detail.

In our use-case, we generate simulations based on Models 1, 2 and 3, to acknowledge the difference in the geometric complexity among models and their potential impact on p-PSHA.

1.3 Velocity structure

We incorporate in our models a 3D velocity model from the tomographic model by Abril et al. (2021), constrained by the Ocean-Bottom Seismometers temporarily deployed in Northern Iceland.

We build CAD models combining topography/bathymetry data, fault models and the velocity structure, and developed within the project and used a **velocity-aware** automatic and flexible computational mesh generation, yielding meshes of optimal element counts. An example configuration for this step is published here:

https://github.com/SeisSol/PUMGen/blob/master/XmlExample/meshAttributes.xml.

Based on numerical analysis presented in Käser et al. 2008, SeisSol requires approx. 2 elements per wavelength for accurate results using a numerical scheme with basis functions of polynomial degree 7. For example, assuming a shear wave speed of 5 km/s and a target frequency of 10 Hz, the shortest wavelength to be resolved is 500 m and the target geometrical mesh resolution is 250 m for the edge length of each tetrahedra of SeisSol's unstructured mesh.

1.4 Fault initial stresses

The HFFZ is exposed to a laterally homogeneous regional stress field constrained from seismo-tectonic observations, knowledge of fault fluid pressurisation, and the Mohr-Coulomb theory of frictional failure, extending the ideas developed in Ulrich et al. (2019). Such regional stress field is fully defined by four parameters: the orientation of the maximum horizontal compressive stress SH_{max}, the stress shape ratio $\boldsymbol{\nu}$= (s2−s3)/(s1−s3) with s1 > s2 > s3 principal stress magnitudes, the variations with depth of the intermediate principal stress magnitude (here assumed as a function of the confining stress times 1−γ with γ the ratio of the fluid pressure P_{fluid} to the background lithostatic stress $\varrho_{rock}gz$; The value $\gamma = \varrho_{water}/\varrho_{rock} = 0.37$ with 1D rock density 2670 kg/m^3 corresponds to a hydrostatic state; higher values γ > 0.37 correspond to overpressurized states) and the maximum pre-stress ratio R_0, which is the ratio of the potential co-seismic dynamic stress drop and the full frictional strength drop on a virtual, optimally oriented fault plane (Anochi & Madariaga, 2003). Ziegler et al. (2016) and Angelier et al. (2004), infer SHmax = 155+/-22 degrees clockwise from north and **ν** ~0.5, from borehole breakouts, drilling induced fractures, earthquake focal mechanism solutions, geological information and overcoring measurements. The choice of an Andersonian stress state, with s2 vertical, is also supported by the inference of a nearly vertical intermediate principal stress by Ziegler et al. (2016), and is consistent with the transform plate motion.

We vary SH_{max} between 135 and 170 degrees clockwise from the north, R_0 between 0.45 and 0.65, and the fluid pressure ratio between 0.55 and 0.70 for ensemble simulations considering the parameter uncertainty.

1.5 Site locations

In our use case, we select major towns along the North Iceland coast as site locations.

1.6 Hypocenter locations and nucleation size

We vary the source locations along the fault strike and dip distances, and different nucleation sizes yielding mechanically plausible rupture scenarios of different magnitudes.

2. **Simulations**

2.1 Physics-based rupture scenarios (SeisSol)

Using the inputs from step 1, we run dynamic rupture simulations using SeisSol. Detailed SeisSol documentation can be found in (www.seissol.org; https://github.com/SeisSol; https://zenodo.org/record/4899349#.YkDej25Bw-Q and https://readthedocs.org/projects/seissol/).

Considering the parameter uncertainties, different fault models, and various nucleation locations and sizes, rupture scenarios breaking different fault segments are generated, resulting in different rupture and slip patterns, and earthquake magnitudes.

2.2 **Earthquake Forecast Model (SHERIFS)**

In our workflow, we use the open source tool "seismic hazard and earthquake rates in fault systems - SHERIFS" (Chartier et al., 2019, https://github.com/tomchartier/SHERIFS), to estimate annual seismic rates using a computationally cheap logic-tree method. The slip rate on faults is constrained by geodetic studies in the study region. We use a scaling law from P. Mai & Beroza (2000), which best represents the scaling relationship in Iceland. Dynamic rupture scenarios allow constraining inputs as fault-to-fault rupture scenarios of the SHERIFS workflow.

3. **Post-processing**

   We developed python tools (https://github.com/SeisSol/SeisSol/tree/master/postprocessing/science/), including the disaggregation procedure, e.g. to filter out the spontaneous rupture scenarios that didn't break out the nucleation zone or generated small magnitudes, and calculate the ground motion maps utilizing routines from the GMPE Strong Motion Modeller's Toolkit (GMPE-SMTK,https://github.com/GEMScienceTools/gmpe-smtk).

   Combining all the ground motion outputs and the SHERIFs rupture probability outputs, we can calculate the hazard curves at the identified sites, or hazard maps for the whole simulation region.

Workflow structure for CyberShake p-PSHA

The Cybershake workflow has been developed by partners of the Southern California Earthquake Center (SCEC) and is well documented (e.g. at i https://strike.scec.org/scecpedia/CyberShake). Within ChEESE we have made an effort to port the Cybershake code base to both a European HPC system (BSC's MareNostrum4) and a European scenario relevant to ChEESE's PD5: The Southern Iceland Seismic Zone. The first goal has involved the development of an open-source workflow manager (UnifiedCSWFlow, see https://zenodo.org/record/6382176#.YkForltBwUE) which replaces the non-open counterparts used in the original CS Studies. In order to generate a new PSHA study we require

1. Velocity model and bounding box: Originally CS expects to enquiry the UCVM, which is specific to the original Studies area in California. In our case we replace the binary model with an ad-hoc file that must be consistent with a geolocalized bounding box. The model includes density, and P- and S- wave velocities (Vp and Vs) in our studies. The model must be compatible with the frequency desired. At present time there is no automation in the selection of the box and model properties.
2. Populate site list: Each location where an intensity measure needs to be analyzed must be included in a list. Typically sites include instrumented locations and a grid whereupon PSHA needs to be evaluated.
3. Earthquake rupture forecast
4. Setup UnifiedCSFlow: This includes selecting paths for inputs and outputs, frequency and a description of the computational resources. Performed by means of a json file.
5. Execution at an HPC site, currently only at BSC but portable. All of the following is automatically managed by UnifiedCSWflow:
   a. Generate SGT for each site: for each site a Green's tensor file is generated toward all potential source locations in model.
   b. Build seismograms from SGTs and time signatures: By means of convolution with time signatures and addition of SGTs all seismograms are built

6. Post-process seismograms to obtain intensity metrics: Preferred metrics are PSA at certain frequency thresholds, PGA, PGV and RotD50.

Required input data

Input parameters need to be stored in the correct format, and integrated for a given region in order to enable the simulation.

Input Data	Format	Specific requirements	Accessibility
SeisSol			
3D velocity model	netCDF	The velocity files needs to be projected from lat-lon to UTM	https://github.com/Bo-Li-LMU/ChEESE_PSHA
Topography and bathymetry	netCDF	The topography/bathymetry needs to be projected from lat-lon to UTM. Required for the mesh.	GeoMapApp (http://www.geomapapp.org/), or GEBCO (https://download.gebco.net/)
Mesh	HDF5 and xdmf		https://github.com/Bo-Li-LMU/ChEESE_PSHA
Fault, initial stress	yaml		https://github.com/Bo-Li-LMU/ChEESE_PSHA
Site locations	dat	The file needs to be projected from lat-lon to UTM	https://github.com/Bo-Li-LMU/ChEESE_PSHA
SHERIFS			
Fault, slip rate, and catalogues of past earthquakes, fault-to-fault ruptures	txt	Input files need to be in lat-lon coordinates	https://github.com/Bo-Li-LMU/ChEESE_PSHA

For Cybershake

Input Data	Format	Specific requirements	Accessibility
3D velocity model	AWP format binary and ASCII model box https://strike.scec.org/scecpedia/CyberShake_Code_Base#AWP_format	Both files need to be compatible. Originally extracted from database of UCVM, not the case in Iceland	https://zenodo.org/record/6382176#.YkHLqzyxVhE
Sites	Text file (csv files)	Coordinates in Lat-Lon of the sites where the intensity measures will be provided. The file depends on the simulation domain	https://zenodo.org/record/6382176#.YkHLqzyxVhE
Rupture forecast	ERF (csv files)	Fault system description including: magnitude, probabilities, rupture size	https://zenodo.org/record/6382176#.YkHLqzyxVhE

		and grid spacing, Initial-end spatial coordinates (lat-lon)	
UnifiedCSWflow parameter file	JSON	Study region, list of sites to process, model, CyberShake, database, simulations setup, number of parallel workers, computational resources and domain decomposition.	https://zenodo.org/record/6382176#.YkHLqzyxVhE
Sqlite database	sqlite	CyberShake mandatory tables	https://zenodo.org/record/6382176#.YkHLqzyxVhE

Output data

Output Data	Format	Number of files	File size	Level
Surface, fault and volume output	HDF5 and xdmf	3 for each	~80 Gb	Level 0
Receivers/sites output	dat	Number of receivers	1-1 Mb	Level 0
Ground shakings	HDF5 and xdmf	3	~500 Mb	Level 1
Rupture probability models	Xml, txt	1	20-100 Mb	Level 1
Hazard curve/map	png	1 for each site/region	~0.2 Mb	Level 2

For Cybershake

Output Data	Format	Number of files	File size	Level
Seismograms	Binary https://strike.scec.org/scecpedia/Accessing_CyberShake_Seismograms	One per each site and rupture ID including all the rupture variations.	~ 200 Mb	Level 0
Peak spectral accelerations, rotD values, and duration metrics.	Binary https://strike.scec.org/scecpedia/CyberShake_output_data_formats#Peak_Spectral_Acceleration_.28PSA.29_files	One per each site and rupture ID including all the rupture variations.	~ 15Mb	Level 0

Modality of accessing the codes/workflows

Trans National Access to the full p-PSHA workflows will be granted by the Geo-INQUIRE project. Virtual Access to the numerical engine are open access and can be found in (www.seissol.org; https://github.com/SeisSol; https://zenodo.org/record/4899349#.YkDej25Bw-Q, https://github.com/tomchartier/SHERIFS).
The code documentation is available at https://seissol.readthedocs.io/en/latest/.

Scaling of computational resources

The computational requirement is based on the size of target region, velocity model, and frequency needs to be resolved. Here we use the use case in HFF as an example.

Target architecture	Node specs	Target size	Total number of nodes per use-case	Nodes per run	Number of runs	Duration of a single run	Data storage needs	Total CPU hours
SuperMUC NG	Intel Skylake CPUs	4.6 million cells	20 nodes with 960 cores	20	>200	~2 hours	20-40 Tb	~420,000

For CyberShake the computational resources are based on specific tests, model domain, sites, and ruptures. In this table we include the use case of the CyberShake execution in the SISZ-RPOR region that contains a total of 144 sites located in a regular grid of 10 km and an ERF that contains 223 different ruptures.

Target architecture	Node specs	Target size	Total number of nodes per use-case	Nodes per run	Number of runs	Duration of a single run	Data storage needs	Total CPU hours
MareNostrum4	Intel Skylake CPUs	2880 millions cells	240 nodes with 11520 cores	24	15	~0.5 hours	7.2 Tb	~86400

References

Abril, C., Tryggvason, A., Gudmundsson, Ó., & Steffen, R. (2021). Local Earthquake Tomography in the Tjörnes Fracture Zone (North Iceland). *Journal of Geophysical Research: Solid Earth*, *126*(6), e2020JB020212.

Angelier, J., Slunga, R., Bergerat, F., Stefansson, R., & Homberg, C. (2004). Perturbation of stress and oceanic rift extension across transform faults shown by earthquake focal mechanisms in Iceland. *Earth and Planetary Science Letters*, *219*(3-4), 271-284.

Aochi, H., & Madariaga, R. (2003). The 1999 Izmit, Turkey, earthquake: Nonplanar fault structure, dynamic rupture process, and strong ground motion. *Bulletin of the Seismological Society of America*, *93*(3), 1249-1266.

Björnsson A, Sæmundsson K, Sigmundsson F, et al (2007) Geothermal projects in Iceland at Krafla, Bjarnarflag, Gjástykki and Theistareykir: assessment of geo-hazards affecting energy production and transmission systems emphasizing structural design criteria and mitigation of risk.

Chartier, T., Scotti, O., & Lyon-Caen, H. (2019). SHERIFS: Open-source code for computing earthquake rates in fault systems and constructing hazard models. *Seismological Research Letters*, *90*(4), 1678-1688.

Einarsson P (1991) Earthquakes and present-day tectonism in Iceland. Tectonophysics 189:261–279.

Käser, Martin and Verena Hermann, Josep de la Puente, Quantitative accuracy analysis of the discontinuous Galerkin method for seismic wave propagation, Geophysical Journal International, Volume 173, Issue 3, June 2008, Pages 990–999, https://doi.org/10.1111/j.1365-246X.2008.03781.x.

Mai, P. M., & Beroza, G. C. (2000). Source scaling properties from finite-fault-rupture models. *Bulletin of the Seismological Society of America*, *90*(3), 604-615.

Sigbjörnsson R, Snæbjörnsson JTh (2007) Earthquake hazard - Preliminary assessment for an industrial lot at Bakki near Húsavík. Earthquake Engineering Research Centre. University of Iceland, Selfoss.

Ulrich, T., Gabriel, A. A., Ampuero, J. P., & Xu, W. (2019). Dynamic viability of the 2016 Mw 7.8 Kaikōura earthquake cascade on weak crustal faults. *Nature communications*, *10*(1), 1-16.

Ziegler, M., Rajabi, M., Heidbach, O., Hersir, G. P., Ágústsson, K., Árnadóttir, S., & Zang, A. (2016). The stress pattern of Iceland. *Tectonophysics*, *674*, 101-113.

PD 6. Probabilistic Volcanic Hazard Assessment

PD 6	Probabilistic Volcanic Hazard Assessment
Leader	Laura Sandri (INGV)
Participants	Beatriz Martinez Montesinos, Antonio Costa, Giovanni, Macedonio (INGV) Manuel Titos, Sara Barsotti (IMO) Arnau Folch, Leonardo Mingari (CSIC)
Workflow	PVHA_WF
Numerical Engine	BET_Tephra, Fall3D
TRL achieved	7
Access Mode	Trans National Access within Geo-INQUIRE project

Description of the pilot use-case

During explosive eruptions volcanoes might inject large volumes of fragmented magmatic material into the atmosphere. Forecasting what will happen in the next few hours when a volcano erupts or quantifying the potential impacts of a future eruption are relevant issues for civil protection agencies, government agencies, aviation and other stakeholders. Since volcanic processes are governed by many degrees of freedom, their outcome is intrinsically unpredictable from a deterministic point of view, in terms of temporal occurrence, eruption scale and volcanic scenario, thus, the probabilistic approach is necessary.

In the case of tephra hazard assessment, to achieve a quality Probabilistic Volcanic Hazard Assessment (PVHA) it is necessary to simulate a large number of volcanic scenarios 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. Furthermore, modelling extremely large eruptions and/or eruptions occurring under extreme wind conditions that produce distal tephra fallout and/or distal or persisting airborne ash concentration demands large scale high-resolution time/space domains. These requirements have not been met so far due to the high computational cost involved.

The objective of PD6 is to first use HPC to overcome the current limits of PVHA in terms of computational resources, and then show the usefulness of PVHA for society in considering the impact of low probability but high impact events.

Specifically, to address the goal of PD6:

- we created the workflow PVHA_WF (https://doi.org/10.5072/zenodo.1036411) able to provide probability and hazard maps, with uncertainty, for tephra fallout at ground and airborne ash concentration and time-persistence at different flight levels,
    - exploring the natural variability in ESPs and wind conditions through a large number of ash dispersal simulations carried out with the Flagship model Fall3D (https://doi.org/10.5281/zenodo.6343786),
    - increasing the computational domain of current PVHA products such BET-Tephra (https://doi.org/10.5281/zenodo.4638667)
        - by one order of magnitude in terms of size and resolution, namely, at European country-size at 2 km x 2 km resolution,

        - also including airborne tephra in 8 vertical levels,
- we showcased PVHA_WF in three selected use-cases that span different time scales and regions:
    - Short-term (ST) PVHA at Campi Flegrei, to produce routine forecasts of ash in 0-24 and 0-48 hours from the forecasting time, accounting for the most recent wind field forecast in tephra simulations, and for the most recent monitoring data from the surveillance network to update the probability of eruption and to produce hazard, probability and arrival time maps for tephra fall, airborne ash concentration at different flight levels, and their persistence in time,
    - Long-term (LT) PVHA at Campi Flegrei and Jan Mayen (Titos et al., 2022), to produce hazard, probability and arrival time maps for tephra fall, airborne ash concentration at different flight levels, and their persistence in time, given an eruption from these volcanoes and considering the climatological statistics of winds,
- through a live exercise, we proved the feasibility and usefulness for end-users, such as the Centre of Competence of the Italian Civil Protection PLINIVS and ARISTOTLE (http://aristotle.ingv.it/tiki-index.php), of such hazard evaluations to produce useful short-term impact assessment of tephra ground load at the scale of a country, in particular over mobility networks (road, railways, seaports and airports) and electrical networks, in an operational environment, dealing with real-time performance-distributed workflow using the ChEESE's WMS-light (https://doi.org/10.5281/zenodo.6340199), in particular,
    - fetching real-time monitoring data from Osservatorio Vesuviano surveillance system, and processing them at INGV Bologna on the computer cluster ADA,
    - fetching weather forecast from GFS and processing them at INGV Bologna,
    - running large-scale and high-resolution tephra dispersal simulations with the Flagship code Fall3D on MareNostrum at BSC, and
    - handling the output from the above on ADA cluster at INGV-Bologna,

Workflow structure

The software consists of four main python components called BET (Bayesian Event Tree) modules.
BET_FETCH. It interrogates the monitoring system and consolidates monitoring data. Used only in case of ST assessment.
BET_EF. It calculates the Eruptive Forecasting (EF) by computing the probabilities of unrest, magmatic unrest and magmatic eruption and by finding the probabilities for vent opening locations. For the ST case these calculations are mainly based on the anomalies, if any, highlighted by the BET_FETCH module previously described.
BET_TEPHRA. It invokes the scripts to download the weather forecast and creates a set of volcanic scenarios (VS) by choosing at random the ESPs within their domain, and a date of occurrence within the period of time under study. For each VS this module launches the execution of an instance of Fall3D in order to calculate the tephra dispersal under the wind conditions downloaded for the date selected. Finally, Fall3D output is analysed and merged to calculate either exceedance probabilities and arrival times for each eruptive size and vent location, conditional to the occurrence of an eruption.
BET_VH. It combines the BET_EF and the BET_TEPHRA results to calculate the final Volcanic Hazard (VH) producing conditional and absolute probabilities matrices, hazard curves and maps.

More specifically, each of these components is divided into subcomponents and managed by the ChEESE's WMS-light, the details of which we do not show here, but can be found on zenodo (www.zenodo.org/communities/cheese.coe).

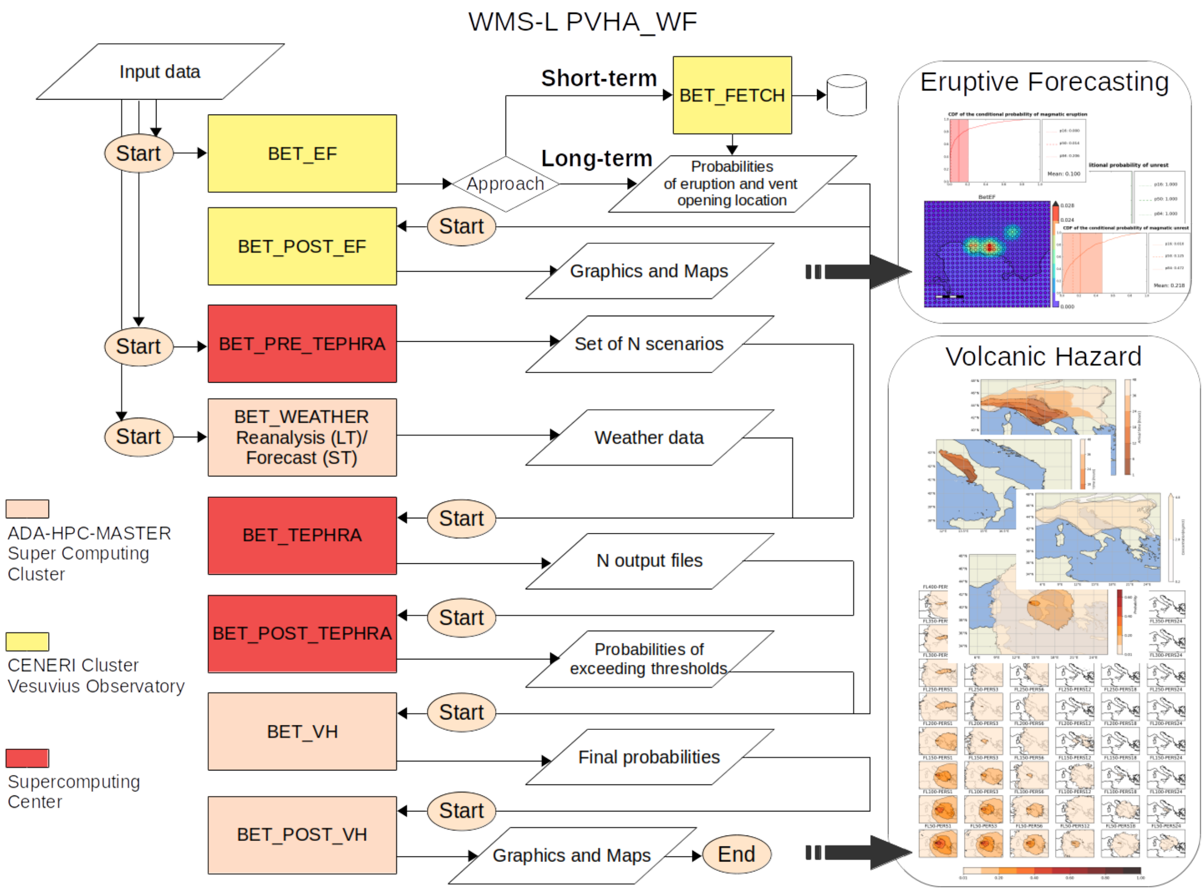


Figure 6.1 Esquema of the PVHA_WH managed by the WMS-light.

Required input data

The idea here is to make users aware of the requirements for running a use-case, in terms of input data. Take as reference the use-cases described in D5.3 and/or D5.4

Input Data	Format	Specific requirements	Accessibility
Probability density functions of each eruptive source parameter and long-term EF	Text file	Configuration file	https://github.com/pvha-wf/PVHA_WF
Location and prior probabilities of vents	Text file	Lat-lon coordinates	ttps://github.com/pvha-wf/PVHA_WF
Grid defining the geographical domain for the Fall3D simulations	Text file	Lat-lon coordinates	ttps://github.com/pvha-wf/PVHA_WF
Grid defining the geographical domain for the VH	Text file with latitude and longitude	Lat-lon coordinates	ttps://github.com/pvha-wf/PVHA_WF

Mapping between each vent location and the corresponding positions of its grid in the Fall3D grid	Numpy array	Array with dimension (number of vents x number of Fall3D-grid points)	ttps://github.com/pvha-wf/PVHA_WF
For the ST, in case of not having access to monitoring data, the ST Eruptive Forecasting is needed	Numpy arrays	probabilities of unrest, magmatic unrest, magmatic eruption and vent opening location, and samples for each of them	ttps://github.com/pvha-wf/PVHA_WF

Table 6.1. Input data required for Probabilistic Volcanic Hazard Assessment simulation workflow.

Output data

Output Data	Format	Number of files	File size	Level
Eruption forecasting figures	png	4	~40 kB	Level 1
Eruption forecasting	npy	7	~0.2 MB	Level 3
Probability of exceeding critical thresholds at ground for each vent position	npy	2	~8 GB	Level 3
Probability of exceeding critical thresholds at flight levels for each vent position	npy	96	~1 GB	Level 3
Mean absolute and conditional probabilities of exceeding critical thresholds at ground	npy	8	~47 MB	Level 3
Percentiles for absolute and conditional probabilities of exceeding critical thresholds at ground	npy	8	~188 MB	Level 3
Mean absolute and conditional probabilities of exceeding critical thresholds at flight levels	npy	192	~6 MB	Level 3
Percentiles for absolute and conditional probabilities of exceeding critical threshold at flight levels	npy	192	~25 MB	Level 3
Arrival times probabilities for a prefixed eruptive size	npy	Max 141	~146 MB	Level 3

Tephra load at ground for a prefixed exceedance absolute probability and a prefixed exceedance conditional probability	npy	8	~5 MB	Level 3
Tephra concentration at flight levels for a prefixed exceedance absolute probability and a prefixed exceedance conditional probability	npy	192	~4MB	Level 3
Images displaying the obtained information	png		~50 kB	Level 1
(*) This information corresponds to the case of Campi Flegrei for a computational domain of 2000 km x 2000 km approximately, at 0.025 degrees of resolution, 8 flight levels (50FL to 400FL), 6 time persistences (1, 3, 6, 12, 18 and 24), 40 potential vent opening locations, and two time periods ([0-24] and [0,48] hours)				

Modality of accessing the codes/workflows

Virtual Access to the numerical engine and documentation will be available via www.zenodo.org (https://doi.org/10.5072/zenodo.1036411) and www.github.org (tps://github.com/pvha-wf/PVHA_WF).

Trans National Access to the full workflow will be granted within the Geo-INQUIRE project.

Scaling of computational resources

Target architecture	Node specs	Target size	Total number of nodes per use-case	Nodes per run	Number of runs	Duration of a single run	Data storage needs
IRENE-SKYLAKE		**Large** This information corresponds to the case of ST Campi Flegrei for a computational domain of 2000 km x 2000 km approximately, at 0.025 degrees of resolution, 8 flight levels (50FL to 400FL), 6 time persistences (1, 3, 6, 12, 18 and 24), 40 potential vent opening locations, using 540 VS	serial	1	-	~10 min	~2 Mb
	Intel-skylake 2.7 GHz, 48 cores per node		8640	16	540	~3 hours	~2.5 Tb
			600	12	50	~15 min	50 Gb

INGV ADA cluster	4CPU Intel Xeon E5-Gold 6140 (Skylake) x 18 cores 24.75 MB di cache, 2.30 GHz		serial	1	-	~30 min	~1Gb
			2	2	1	~15 min	50 Gb
	CENERI - Debian GNU/Linux	**Medium** This information corresponds to the live exercise of ST tephra load assessment at Campi Flegrei for a computational domain of about 200 km x 200 km approximately, at 0.02 degrees of resolution, 40 potential vent opening locations, using 300 VS	serial	1	-	~10 min	~2 Mb
Mare Nostrum 4	2x E52670 SandyBridge -EP 2.6GHz cache 20MB 8-core		24	12 6 6 1	100 100 100 2	~20 min ~20 min ~5 min ~30 min	~112 Gb
INGV ADA cluster	4CPU Intel Xeon E5-Gold 6140 (Skylake) x 18 cores 24.75 MB di cache, 2.30 GHz		serial	1	-	~30 min	~100 Mb
			12	12	1	~90 min	~2 Gb

PD 7. Probabilistic tsunami hazard assessment

PD 7	Probabilistic tsunami hazard assessment (PTHA)
Leader	NGI
Participants	INGV, UMA
Workflow	Suite of bash and python codes for workflow execution, pre- and post-processing;
Numerical Engine	Tsunami-HySEA
TRL achieved	6
Access Mode	Trans National Access within Geo-INQUIRE project

Description of the pilot use-case

Probabilistic Tsunami Hazard Analysis (PTHA) quantifies the likelihood of exceeding a specified measure of tsunami inundation at a given location within a given time interval. It provides scientific guidance for decision making regarding coastal engineering and evacuation planning. PTHA considers a discretization of the total hazard into many potential scenarios together with an evaluation of the probability of each scenario. Site-specific PTHA, with an adequate discretization of source scenarios, combined with high-resolution inundation modelling, has been out of reach with existing models and computing capabilities with tens to hundreds of thousands of moderately intensive numerical simulations being required. In recent years, more efficient GPU-based High-Performance Computing (HPC) facilities, together with efficient GPU-optimised shallow water models for simulating tsunami inundation, have made regional and local long-term hazard assessment feasible. PTHA is Pilot Demonstrator 7 (PD7) in ChEESE. The PD7 workflow operates in three main stages: i) site-specific source selection and discretization, ii) efficient numerical inundation simulation for each scenario using the GPU-based Tsunami-HySEA numerical tsunami propagation model, and iii) hazard aggregation. The PD7 workflow enables integration of large batches of tsunami simulations (typically tens to hundreds of thousands) carried out on Tier-0 systems, for quantifying probabilistic hazard. By aggregating the hazard, a much more fine-grained assessment compared to the previously available regional assessments can be produced. Through the PRACE project TsuHazAP, the capability of the PD7 workflow has been demonstrated on the Marconi100 Supercomputer.

Workflow structure

The PTHA workflow consists of three primary parts: pre-processing, simulations, and post-processing (Figure 7.1). The Pilot Demonstrator developed in the ChEESE project builds upon the NEAM Tsunami Hazard Model 2018 or NEAMTHM18 (Basili et al., 2021) and, as such, is restricted in its current form to the NEAM (North-Eastern Atlantic, Mediterranean and connected seas) region. The pre-processing steps involve the specification of the coastal region of interest and hazard metrics and user requirements such as the target probability range, and provision of bathymetric grids. A series of scripts with the appropriate parameters then probe the NEAMTHM18 database and return an ensemble of earthquake scenarios which dominate the hazard at the coastline of interest. In the simulation step, a set of input files to the Tsunami-HySEA code is copied over to HPC resources and the simulations carried out for each scenario.

In the post-processing step, the output from each scenario (maximum height and/or maximum momentum flux and/or maximum modulus of velocity, and ground deformation, each evaluated on a high resolution grid, of the order of few meters to ten meters) is returned from the HPC resources and combined with the source probabilities to produce probability of exceedance curves and maps. A more complete description of a case-study for Catania, on the Eastern Coast of Sicily, is provided by Gibbons et al. (2020).

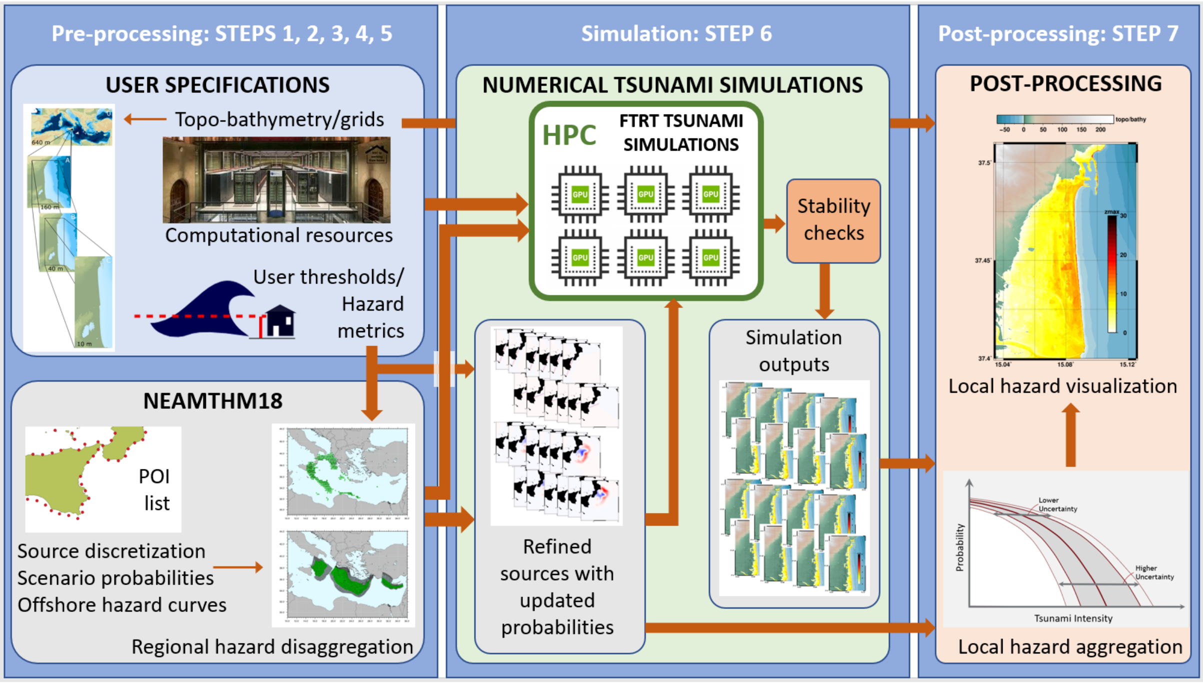


Figure 7.1: The PTHA workflow has three primary components: pre-processing, simulation, and post-processing. For practical purposes these are broken down into seven steps that are executed sequentially.

Figure 7.2 takes us through the sequence of steps needed to complete the PTHA workflow which are summarised below:

1. Selection of Points of Interest (POIs). This process, administered by a single python script, **select_tsumaps_pois.py**, specifies a rectangular geographical region containing the coastline of interest and returns those Points of Interest from the NEAMTHM18 Tsunami Hazard Model that fall within that region.
2. "Scenario Dumping" extracts all of the earthquake scenarios and corresponding metadata which can be associated with the tsunami hazard at the coastline of interest. This process is administered by a single python script, **poi_dumping.py**. The scenarios output may be PS (Predominant Seismicity) sources – subduction earthquakes with well-defined fault geometries – or BS (Background Seismicity) sources – crustal sources for regions of more uncertain focal mechanisms.
3. Scenario Disaggregation finds from the NEAMTHM18 database the scenarios that contribute most significantly to the tsunami hazard at the POI according to the user-specified thresholds and the range of potential Maximum Inundation Height (MIH) to be considered. For example,

we may want a list of those scenarios accounting for 95% of the total hazard within the 1-4 m range of MIH. This step is administered by a python script **poi_disaggregation.py**.

4. "Scenario Rates" interrogates the NEAMTHM18 database for the probabilities of occurrence of each of the scenarios extracted in Step 3. This is to say that, for any given earthquake scenario, we obtain a distribution of the expected annual rates of occurrence, with the epistemic uncertainty expressed through the sampling of 1000 alternative models. This step is managed by the python scripts **scenarioBS_probabilities.py** and **scenarioPS_probabilities.py**.
5. "Scenario Refinement" is an optional step which increases the resolution of the source discretization with respect to the NEAMTHM18 Hazard Model. It allows us to specify a denser discretization of the sources, both geographically, with regards to the fault centers, and with respect to the focal mechanism (strike angle) (for BS sources),, and with regards to the slip distributions for subduction earthquakes (PS sources).
6. Tsunami Simulations. Here we generate input files for the Tsunami-HySEA code from all of the requested scenarios and generate a sequence of recipes for executing the simulations on HPC resources, using telescopic nested grids up to the maximum resolution around the target site. At present, there are several different scripts for different HPC machines. As a service, this is one of the key points that would have to be addressed.
7. Hazard aggregation and visualization. Here, the output from each of the simulations is combined with the scenario rates calculated in Step 4. The output generated is a set of hazard curves, for each point of the high resolution grid and for each model of the epistemic uncertainty. Form the curves, hazard and probability maps within a specified time-frame can be easily obtained .

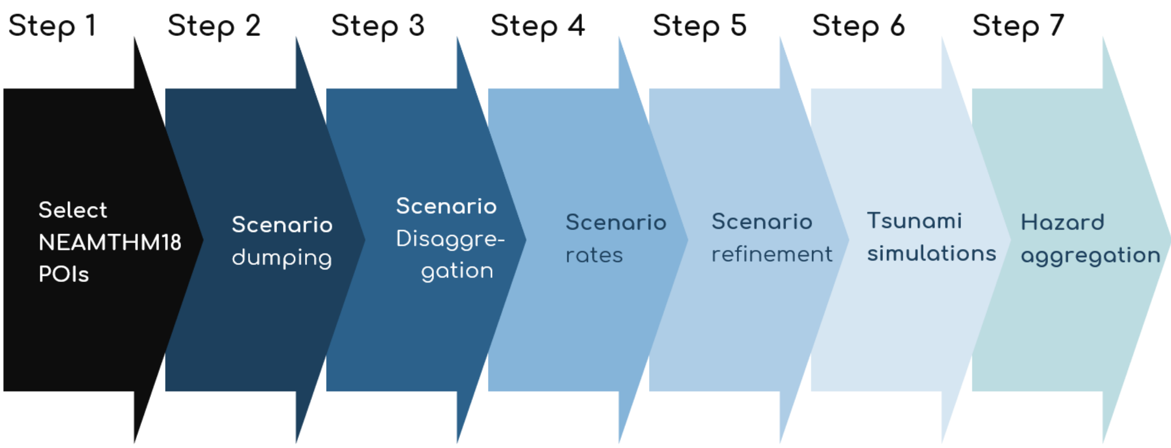


Figure 7.2 Seven sequential steps for implementing the PTHA workflow.

The NEAMTHM18 tsunami hazard model is located at
http://www.tsumaps-neam.eu/neamthm18/

The Workflow is documented on the Zenodo ChEESE Community Repository with
DOI:10.5281/zenodo.6376607
https://zenodo.org/record/6376607#.Yjwd4zUo9aQ

Required input data

Input Data	Format	Specific requirements	Accessibility
NEAMTHM18 Tsunami Hazard model	Various digital file formats.		Metadata and background at https://data.ingv.it/en/dataset/309#additional-metadata http://www.tsumaps-neam.eu/neamthm18/ Access will be available via Trans-National Access under GEO-INQUIRE.
Grid for calculating tsunami propagation. The Catania case study uses a set of nested grids with 10m by 10m resolution for the inundation calculation with a factor of 4 decrease in resolution with each layer with 640m by 640m for the coarsest grid.	netcdf	The only requirement from the Tsunami-HySEA code is that the scaling factor between the dimensions at each level is a factor of 2: in practice either 2 or 4.	The GEBCO (General Bathymetric Chart of the Oceans) publicly available grid (https://www.gebco.net/) is typically sufficient for the coarsest grid. The higher resolution grids may require proprietary data.

Table 7.1. Input data required for Probabilistic Tsunami Hazard Assessment simulation workflow.

Output data

Output Data	Format	Number of files	File size	Level
Single simulation output files	netcdf	3 output files per simulation.	5 MB per simulation.	Level 1
Aggregated hazard maps	netcdf	~50-100 files (depends on number of hazard metrics required by user)	5 MB per file.	Level 2

Table 7.2. Output data produced by Probabilistic Tsunami Hazard Assessment simulation workflow.

Modality of accessing the codes/workflows

- Trans National Access to the full workflow will be granted by the Geo-INQUIRE project.

Scaling of computational resources

Target architecture	Node specs	Target size	Total number of nodes per use-case	Nodes occupied by a single run	Number of runs	Duration of a single run	Data storage needs	Total GPU hours
Marconi100	4 GPU V100	Small (Order of magnitude 10^4 sim)	32 (128 simulations per job)	1	Order of magnitude 10^4 simulations	23 minutes	0.3 TB	Typical range 3000-30000
		Large (Order of magnitude 10^5 sim)	32 (128 simulations per job)	1	Order of magnitude 10^5 simulations	23 minutes	3.5 TB	Typical range 30000-300000

References

Basili, R., Brizuela, B., Herrero, A., Iqbal, S., Lorito, S., Maesano, F. E., et al. (2021). The Making of the NEAM Tsunami Hazard Model 2018 (NEAMTHM18). Front. Earth Sci. 8, 1–29. doi:10.3389/feart.2020.616594.

Gibbons, S. J., Lorito, S., Macías, J., Løvholt, F., Selva, J., Volpe, M., et al. (2020). Probabilistic Tsunami Hazard Analysis: High Performance Computing for Massive Scale Inundation Simulations. Front. Earth Sci. 8, 1–20. doi:10.3389/feart.2020.591549.

PD 8. Probabilistic Tsunami Forecasting (PTF) for early warning and rapid post event assessment

PD 8	PD NAME
Leader	Stefano Lorito (INGV)
Participants	Manuela Volpe, Jacopo Selva, Fabrizio Romano, Roberto Tonini, Fabrizio Bernardi, Maria Concetta Lorenzino (INGV Carlos Sanches-Linares, Jorge Macias, Marc de la Asuncion, Jose Manuel Gonzalez Vida, Manuel J. Castro (UMA) Steven Gibbons, Finn Løvholt, Malte Vöge, Sylfest Glimsdal (NGI) Silvia Giuliani, Isabella Baccarelli, Piero Lanucara (CINECA) Alexey Cheptsov (HLRS) External Contributors Antonio Scala (UniNA) Andrey Babeyko (GFZ)
Workflow	Suite of bash, python/Matlab, C, GMT codes for workflow execution, pre- and post-processing; HLRS WMS-light workflow
Numerical Engine	Tsunami-HySEA
TRL achieved	7-8 for the Urgent Computing mode PTF version, based on new simulations of large ensembles of tsunami scenarios 9 for the Early Warning mode PTF version, based on pre-calculated large ensembles of tsunami scenarios, which runs in the operational environment of the Italian Tsunami Service Provider (CAT-INGV)
Access Mode	Virtual Access (download from GitHub) for the basic Early Warning mode version Trans National Access (by request), for the multi-purpose version (Early Warning and Urgent Computing mode)

Description of the pilot use-case

Tsunamis may strike a coastal population within a very short amount of time, as happened in Palu bay, Sulawesi Island, Indonesia, in September 2018. For such a sudden and local phenomenon (Carvajal et al., 2019), the spontaneous reaction of a well-prepared population to natural warnings—such as intense and/or long shaking, sea level anomalies and "roars" from the sea—probably remains the best and only viable life-saver.

However, in less extreme situations, but still for local tsunamis that may hit in less than, say, 10 minutes, extremely fast decisions need to be based on very limited data. The uncertainty can be conveniently quantified through Probabilistic Tsunami Forecasting (PTF). A unique tsunami initial condition on which to base the forecasting of the tsunami coastal impact cannot be reasonably well constrained until some time after the earthquake occurrence. Therefore, and especially for tsunamis generated close to the target coastlines (local tsunamis), dedicated methods are needed to deal with the uncertainty due to many different possible source mechanisms. These include, for instance, modeling different candidate slip distributions used for the quantification of the tsunami impact uncertainty range. Poor source or tsunami characterization may in fact dramatically affect inundation forecasting (e.g. Satake et al, 2013). Uncertainty also stems from an incomplete understanding of the complex and nonlinear overall dynamic process, from generation to inundation and damage (Lynett, 2009; Geist, 2009; Nosov and Kolesov, 2011; Burbidge et al., 2015; Goda and Abilova, 2016); lack of base data such as imperfect knowledge of the near-coast bathymetry and coastal topography (Griffin et al., 2015; Song and Goda, 2019); necessary approximations in the source selection, physics, and resolution for numerical modeling, to increase speed on not-unlimited resources (e.g., Sarri et al., 2012; Glimsdal et al., 2013; Behrens and Dias, 2015, Lorito et al., 2015; Glimsdal et al., 2019; Volpe et al., 2019).

PD8 provides a rapid probabilistic forecast of tsunami inundation, following an earthquake offshore or close to the coast, before it actually occurs or before tsunami observations are available. The PTF concept is sketched in Figure 8.1. More details can be found in Selva et al. (2021 https://www.nature.com/articles/s41467-021-25815-w).

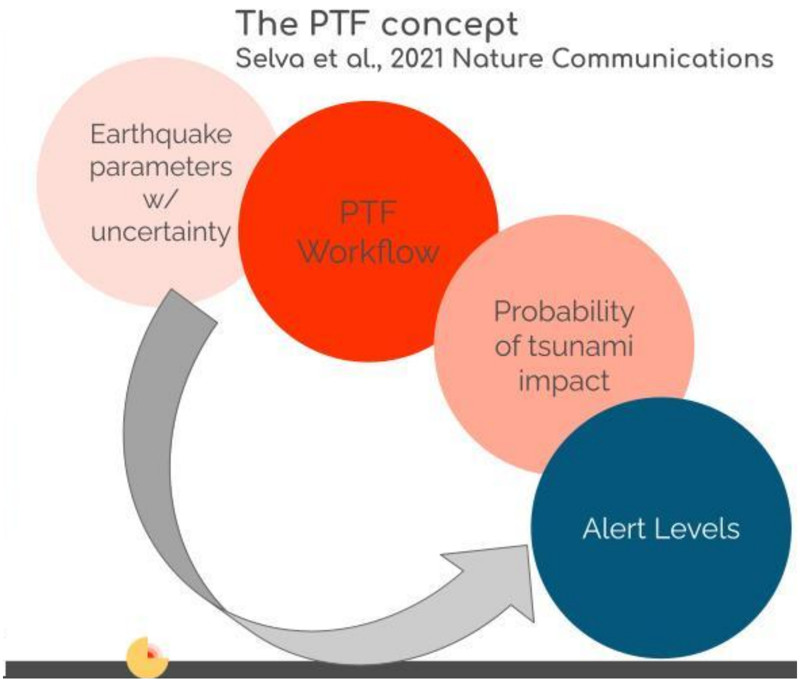
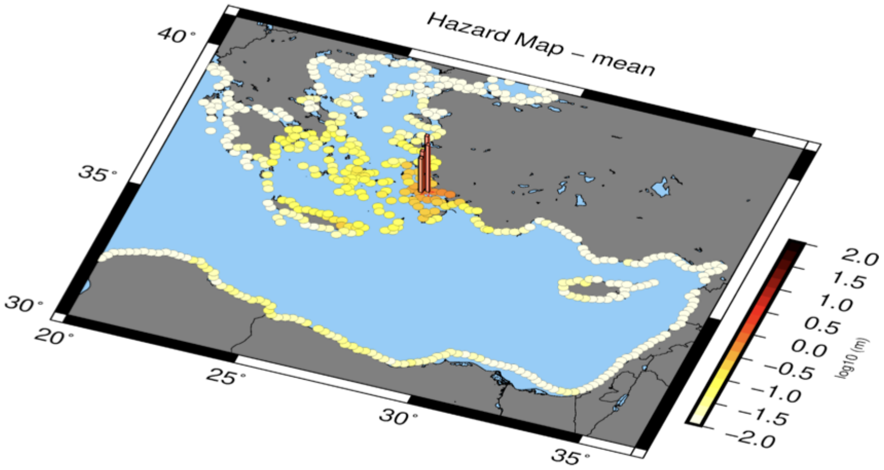


Figure 8.1. The PD8 PTF concept. The PTF Workflow block includes large ensembles of numerical simulations for uncertainty quantification. The simulation results can be queried by a pre-calculated database, or they can be run from scratch.

For near-field tsunami early warning (EW) purposes, the large uncertainty about earthquake location and magnitude, the only parameters available in the first minutes, is reflected into forecasting uncertainty (Figure 8.2). The probabilistic tsunami forecast can be translated into an alert level, based on which risk mitigation operations would start.

For the purpose of supporting rapid post-disaster intervention, for which more time is available, additional source and even tsunami observations in the subsequent phases can be exploited to eventually narrow down the tsunami forecast uncertainty.

As a by-product, the service also provides an early estimate of the not-yet-available earthquake parameters with their uncertainty, such as the fault and rupture orientation (strike, dip, rake).



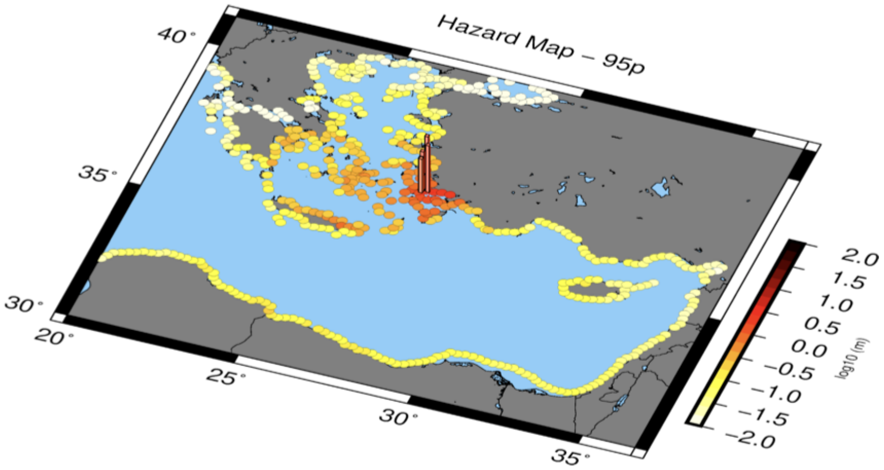


Figure 8.2. Hazard maps hind-casting for the Samos 30 October 2020 Mw 7 earthquake.

Workflow structure

The objective of PTF is to quantify this uncertainty and to project it on the tsunami impact prediction. To illustrate the workflow, in Figure 8.3 we use figures from the PTF applied to the Mw 8.6 NEAMWave17 Scenario (IOC, 2017).

To obtain the probabilities for all necessary source parameters (Figure 8.2C), PTF combines the

1. available, uncertain, real-time seismic event parameters (in the present implementation, the earthquake location, and magnitude, Figure 8.3A);
2. the "long-term" probabilities for unavailable parameters, as derived from local tectonics and past seismicity (Figure 8.3B);
3. stochastic simulations for predicting earthquake slip distributions.

Eventually, the probability of an ensemble of different potential source realizations is assigned according to the degree of consistency with the available real-time data, using the source discretization scheme associated with the long-term probability, and merging it with the underlying tsunami hazard model.

For a real event, the earthquake parameters would be provided a few minutes from earthquake occurrence by the Early-Est software used by CAT-INGV (e.g. Bernardi et al., 2015). Conversely, in the example qualitatively sketched in Figure 8.3, the probabilities for real-time earthquake parameters are artificial and correspond to a plausible estimation of location and magnitude which would be performed in real-time for a subduction earthquake of this size. The information that the input synthetic scenario is a subduction event, with a pure thrusting mechanism at the subduction interface, and not a crustal earthquake, is not passed along, though. The only inputs provided to the workflow are the tri-dimensional covariance matrix describing earthquake location uncertainty and the empirical probability density function for the earthquake magnitude. Based on the magnitude and the subduction interface proximity, the PTF sets the probability of being a crustal earthquake as almost negligible.

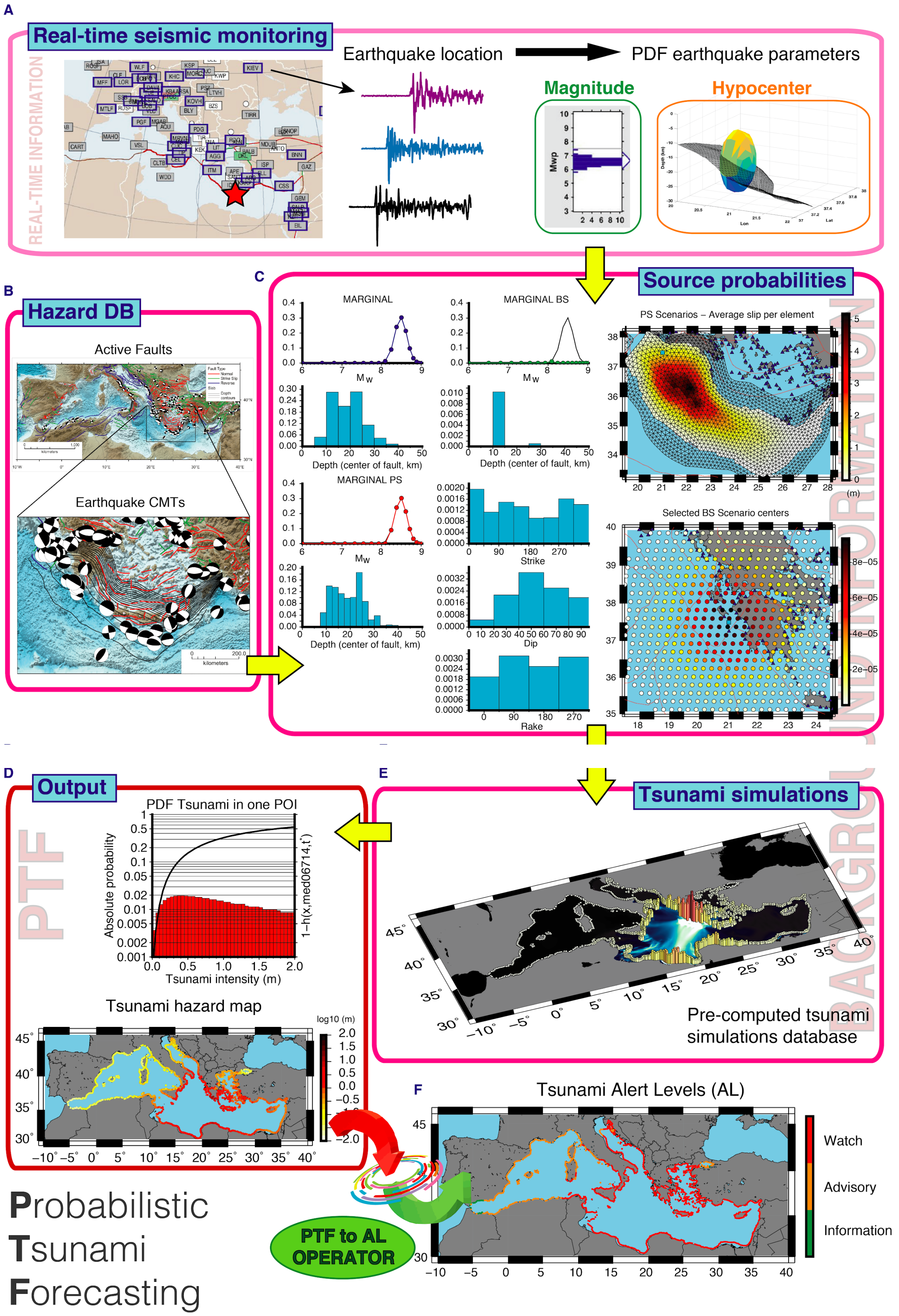


Figure 8.3. Schematic representation of the PTF workflow applied to the NEAMWave17 Mw 8.6 earthquake scenario on the Hellenic Arc.

In the version implemented for the Mediterranean Sea, the fault databases, the seismicity catalogs, and the likelihood of different faulting mechanisms given the location and the magnitude of the earthquake are derived from NEAMTHM18, the hazard model produced by the project TSUMAPS-NEAM (Basili et al., 2018; 2021). NEAMTHM18 is the first regional probabilistic tsunami hazard assessment (PTHA) for the NEAM region (North-eastern Atlantic, the Mediterranean and connected seas).

The details of NEAMTHM18 and the underlying databases and tools are available in the TSUMAPS-NEAM Documentation (Basili et al., 2019). It is worth noting that a series of rather innovative approaches, which constitute the model foundation, were introduced by the Project's partners, including source, tsunami and uncertainty characterization modeling (Basili et al., 2013; Lorito et al., 2015; Molinari et al., 2016; Murphy et al., 2016; Selva et al., 2016; Davies et al., 2018; Maesano et al., 2017; Herrero and Murphy, 2018; Glimsdal et al., 2019; Maesano et al., 2020; Scala et al., 2020).

In Figure 8.3E, a single tsunami numerical simulation output is sketched. This is a simulation of the tsunami which would be generated by a magnitude 8-class earthquake occurring in the Hellenic Arc subduction zone. However, to complete the PTF assessment, typically tens of thousands of simulations are necessary, one for each scenario included in the ensemble of sources. The latter is the result of sampling from the earthquake parameter space, resulting in a subset consistent with the available seismic data within a prescribed level of tolerance.

The simulations are performed with the Tsunami-HySEA model (see e.g., de la Asunción, 2013), enabling FTRT tsunami simulations on GPU and multi-GPU architectures, was developed by EDANYA group at the University of Malaga, Spain. Tsunami-HySEA was benchmarked and approved by the NTHMP (National Tsunami Hazard and Mitigation Program) to be used officially in risk assessment studies for tsunamis in the USA, with funding from this program (Macías et al., 2016; 2017). The great computational efficiency of this code compared to the traditional tsunami codes has opened the way to integrate it into Tsunami Early Warning Systems (TEWS). At present, the national warning systems of Spain and Italy implemented Tsunami-HySEA as computing software. Tsunami-HySEA is one of the ChEESE flagship codes and it was futher optimized during the project.

The ensembles are typically run using one GPU for each scenario, by sending groups of scenarios (e.g. 4 simulations/node on Marconi 100 @CINECA, having 4GPUs/node) on the available nodes using the MC version of Tsunami-HySEA which deals with chunks of scenarios.

There are two versions of the PTF tool. Both of them have been run already for various earthquakes and tsunamis in the Mediterranean Sea and in the Pacific Ocean.

Early Warning mode version - This version is based on pre-calculated numerical simulations of tsunami scenarios. It works in near-real time in the premises of the CAT-INGV Tsunami Warning Centre (https://www.ingv.it/cat/en/), which is a NEAMTWS Tsunami Service Provider (http://www.ioc-tsunami.org/). It deals with potentially tsunamigenic earthquakes anywhere in the entire Mediterranean Sea (offshore and inland close to the coast). Order of 10-100 k pre-calculated scenarios were run on a 30 arc-sec grid for 8 hours of simulation. The number of scenarios in the ensemble depends on the magnitude of the earthquake and on the desired level of accuracy, that is, roughly speaking, how many standard deviations are explored around the expected values of the seismic parameters. In this configuration, PTF output provides exceedance probabilities for tsunami heights just off the coastline for almost equally spaced points of interest every 20 km along the coasts of the Mediterranean Sea.

Multi-purpose - Early Warning and Urgent Computing mode version - The second version is based on simulation ensembles to be run from scratch on large enough HPC clusters in the Urgent Computing mode. Nonetheless, it can be also fed with pre-calculated scenarios, in the Early Warning mode. Using the Mw 7.0 Samos earthquake as an input, this version of the workflow has been tested with several large scale runs in urgent computing mode on Marconi100 at CINECA. A reservation of 800 nodes each equipped with 4 V100 GPUs has been used for testing the workflow, considering two standard deviations for the source uncertainty, which leads to almost 40 000 scenario simulations. From the first test several bottlenecks have been identified, and a new test at the same scale was successfully carried out. Consequently new scripts are now being integrated into the prototype WMS-light-based workflow developed in collaboration with HLRS.

In the Early Warning version running at CAT-INGV, the tsunami scenarios are obtained using a look-up table to retrieve them according to the seismic parameters. There are pre-calculated tsunami events calculated from linear combinations of “mareogram Green’s functions”, obtained by propagating Gaussian-shaped unit sources for the sea level elevation with the Tsunami-HySEA code (Molinari et al., 2016). Combination coefficients are those that best reproduce the tsunami initial condition as predicted by the Okada (1985) algorithm low-pass filtered through the water column (Kajiura, 1963). Maxima of tsunami elevation recorded at predefined locations along the 50 meters isobath are recorded. They are then treated in two possible ways. First and simpler is their extrapolation according to amplification foreseen by the basic version of the Green’s law (e.g. Kamigaichi, 2014); then a relatively simple treatment of uncertainties related to all phases of the process from generation to inundation is applied (Davies et al., 2018). The second approach is the same as in TSUMAPS-NEAM, using the amplification factors introduced by Glimsdal et al. (2019), which differ depending on the incident wave leading polarity and dominant period, and on the local offshore near-coast bathymetric profile.

In the ChEESE PD8 PTF workflow, the role of the pre-calculated scenarios is passed to on-the-fly simulations of scenario ensembles, performed with Tsunami-HySEA. This is the HPC component of the PD8-PTF workflow. This version allows to provide the tsunami forecast inside the Mediterranean Sea using spatially “densified” points of interest along the coasts, or to deal with new cases outside the Mediterranean Sea.

Both versions of PD8 progressed during ChEESE from a set of scripts previously run manually to a now fully automated workflow. The version based on pre-computed scenarios can be activated by a Rabbit-MQ message produced by CAT-INGV and can produce automatic alert messages. The urgent computing version needs to be triggered manually but then runs from the pre-processing, through the HPC part, to the post-processing and basic visualisation of the results.

The simulation results are aggregated with the source probabilities sketched in Figure 8.3C, to obtain, at each point of interest along the coast, outputs like those depicted in Figure 8.3D. The source probabilities quantify the resemblance between a member of the source scenario ensemble and the data. Hence, the ensemble of deterministic numerical simulations acts as an “uncertainty propagator” from the source to the tsunami impact. The latter is expressed as a probability of a chosen tsunami intensity provided by the simulations, conditioned to the occurrence of a seismic event as characterized, in turn, by the data probabilities in Figure 8.3A. Those are either sketched by a graphical representation of Probability and Cumulative Density Functions (PDF and CDF), for some specific target point along the coast, or as a “hazard map”, representing the exceedance probability curve for different values of the tsunami intensity, as obtained considering the PDF’s mean values along the coastal points.

Finally, the PTF workflow contains tools for converting the results to Alert Levels (ALs), which refer to specific actions such as the evacuation of a coastal zone. The ALs are determined by combining the tsunami-inundation-height probability with predefined conversion rules (provided by the decision-makers) at all the points of interest. This is a very important property of the PTF because it introduces a transparent approach to uncertainty management from the “upstream” component of a tsunami early warning system to the “last mile”. The conversion rules allow the explicit mapping of a given “acceptable hazard level” into the corresponding probability threshold. When the threshold is exceeded, the corresponding AL is activated. The conversion rules are to be established in advance but can be then automatically implemented for seamless real-time operations. The conversion rules may be for example based on the comparison between different statistics of the PDF at one specific location with tsunami intensity thresholds (Selva et al., 2021). Further documentation can be found in the repositories for each version of the workflow.

Required input data

The idea here is to make users aware of the requirements for running a use-case, in terms of input data. Take as reference the use-cases described in D5.3 and/or D5.4

Input Data	Format	Specific requirements	Accessibility
Mediterranean case	Digital format or standards	E.g., resolution, frequency	Public repository available?
Earthquake parameters	text	Lon, Lat, Depth, Magnitude, Uncertainties	Examples on https://github.com/INGV/matPTF
Regionalization	Binary (Matlab)	Coordinates, types of seismicity	http://www.tsumaps-neam.eu/documentation/
Event tree Discretization	Binary (Matlab)	Possible values for earthquake magnitude (BS, PS), earthquake hypocenter (BS), earthquake focal mechanism (BS), fault size	http://www.tsumaps-neam.eu/documentation/
Weights of alternative models (epistemic uncertainty)	Binary (Matlab)	Weights for earthquake magnitude (BS, PS), earthquake hypocenter (BS), earthquake focal mechanism (BS), fault size	http://www.tsumaps-neam.eu/documentation/
Discretization on Predominant Seismicity sources (barycenters)	Binary (Matlab)	Barycenters of PS sources	http://www.tsumaps-neam.eu/documentation/
Scenario List BS	Binary (Matlab)	Earthquake magnitude, hypocenter, strike, dip, rake, fault area, fault length, slip	http://www.tsumaps-neam.eu/documentation/
Scenario List PS	Binary (Matlab)	Id scenario, slip distribution, mesh element indices	http://www.tsumaps-neam.eu/documentation/

Triangular meshes of the subduction interface sources	text	Three files containing elements indices, vertices (lon, lat, depth) of the nodes, and coordinates (lon, lat) of the elements	http://www.tsumaps-neam.eu/documentation/
Probability of focal mechanisms (mean of the epistemic uncertainty)	Binary (Matlab)	Probability for evry a set of possible triplets strike/dip/rake	http://www.tsumaps-neam.eu/documentation/
List of POIs	text	Lon, Lat	https://zenodo.org/record/6376520
Thresholds for tsunami intensity for Hazard curves	text	Wave amplitude	https://zenodo.org/record/6376520
Amplified offshore wave amplitude	Binary (Matlab)	Wave amplitude	https://zenodo.org/record/6376520
Bathymetry[**]	Binary (netcdf)	Lon, Lat, Depth	https://zenodo.org/record/6376520
POI depths[**]	text	Depth	https://zenodo.org/record/6376520

[*]Input Data used only in the “on the fly” simulations mode

Input Data	**Format**	**Specific requirements**	**Accessibility**
Pacific ocean case	Digital format or standards	E.g., resolution, frequency	Public repository available?
Earthquake parameters	text	Lon, Lat, Depth, Magnitude, Uncertainties	Examples on https://github.com/INGV/matPTF
Regionalization	Binary (Matlab)	Coordinates, types of seismicity	http://www.tsumaps-neam.eu/documentation/
Event tree Discretization	Binary (Matlab)	Possible values for earthquake magnitude (BS, PS), earthquake hypocenter (BS), earthquake focal mechanism (BS), fault size	http://www.tsumaps-neam.eu/documentation/

Discretization on Predominant Seismicity sources (barycenters)	Binary (Matlab)	Barycenters of PS sources	http://www.tsumaps-neam.eu/documentation/
Triangular meshes of the subduction interface sources	text	Three files containing elements indices, vertices (lon, lat, depth) of the nodes, and coordinates (lon, lat) of the elements	http://www.tsumaps-neam.eu/documentation/
Probability of focal mechanisms (mean of the epistemic uncertainty)	Binary (Matlab)	Probability for evry a set of possible triplets strike/dip/rake	http://www.tsumaps-neam.eu/documentation/
List of POIs	text	Lon, Lat	https://zenodo.org/record/6376520
Thresholds for tsunami intensity for Hazard curves	text	Wave amplitude	https://zenodo.org/record/6376520
Amplified offshore wave amplitude	Binary (Matlab)	Wave amplitude	https://zenodo.org/record/6376520
Bathymetry*	Binary (netcdf)	Lon, Lat, Depth	https://zenodo.org/record/6376520
POI depths*	text	Depth	https://zenodo.org/record/6376520

*Input Data used only in the "on the fly" simulations mode

Output data (Mediterranean case)

Output Data	Format	Number of files	File size	Level
Scenario Lists	txt	1 (2 if both BS and PS are considered)	~10 MB	Level 3
Settings from Step1	Binary (Matlab)	1	<1KB	Level 3
Early-Est parameters	Binary (Matlab)	1	<1KB	Level 3

Long Term information	Binary (Matlab)	1	~10 MB	Level 3
Short Term information	Binary (Matlab)	1	~10 KB	Level 3
Scenario probabilities	Binary (Matlab)	1	~1 MB	Level 3
Used functions	Binary (Matlab)	1	~10 KB	Level 3
Amplified offshore wave amplitude[*]	Binary (Matlab)	1	~1-5 GB	Level 3
POI info	Binary (Matlab)	1	~100 KB	Level 3
Hazard Curves	CSV, netcdf, matlab	5	<100 MB	Level 3
Alert Levels	Binary (Matlab)	1	<1 MB	Level 3
Figure Map Alert Levels	png	1	~1 MB	Level 3
Figure Hazard Map (mean, 5, 50, 95 percentiles)	png	4	< 10 MB	

[*]Output Data obtained only in the "on the fly" simulations mode

Output data (Pacific ocean [Maule] case)

Output Data	Format	Number of files	File size	Level
Scenario Lists	txt	1 (2 if both BS and PS are considered)	~10 MB	Level 3
Settings from Step1	Binary (Matlab)	1	<1KB	Level 3
Early-Est parameters	Binary (Matlab)	1	<1KB	Level 3
Long Term information	Binary (Matlab)	1	~100 MB	Level 3
Short Term information	Binary (Matlab)	1	<100 KB	Level 3
Scenario probabilities	Binary (Matlab)	1	~<100 KB	Level 3
Used functions	Binary (Matlab)	1	<10 KB	Level 3
Amplified offshore wave amplitude[*]	Binary (Matlab)	1	< 100 MB	Level 3

POI info	Binary (Matlab)	1	~100 KB	Level 3
Hazard Curves	CSV, netcdf, matlab	5	<200 KB	Level 3
Alert Levels	Binary (Matlab)	1	<100 KB	Level 3
Figure Map Alert Levels	png	1	~1 MB	Level 3
Figure Hazard Map (mean, 5, 50, 95 percentiles)	png	4	< 10 MB	

*Output Data obtained only in the "on the fly" simulations mode

Modality of accessing the codes/workflows

Virtual Access to the PD8 - Early-Warning mode (https://doi.org/10.5281/zenodo.6390936).

A stand-alone Matlab-based version of the PD8 source code is available on a github repository (https://github.com/INGV/matPTF; release 1.0.0). It is part of the ChEESE-COE zenodo repository (https://zenodo.org/communities/cheese-coe/?page=1&size=20). This version is the accompanying material to the published scientific paper presenting the methodology (Selva et al., 2021; https://www.nature.com/articles/s41467-021-25815-w), which fully acknowledges the ChEESE contribution. It is based on pre-calculated scenarios. Some input data necessary to run the workflow can be retrieved by following the links provided on the github repository. They are hosted elsewhere because of the file sizes, namely on http://www.tsumaps-neam.eu/documentation/ and https://doi.org/10.6084/m9.figshare.15015132.

Trans National Access to the PD8 for selected users, also under collaborative actions - Multi-purpose version (Early Warning and Urgent Computing mode - https://doi.org/10.5281/zenodo.6376520).

The source code resides on a private GitLab repository (https://gitlab.com/manuela.volpe/cheese-ptf). It is a suite of bash, python/Matlab, C, GMT codes for workflow execution, pre- and post-processing. It also includes a branching embedded in the HLRS WMS-light workflow (https://doi.org/10.5281/zenodo.6340199). Trans National access can be granted to selected users, also in the framework of a collaboration by contacting manuela.volpe@ingv.it. Running this workflow on new cases is subject to the availability of adequate computational resources. For reduced-size use-cases and/or testing purposes, limited computational resources may be available to run Tsunami-HySEA on the Mercalli@INGV hybrid CPU-GPU cluster.

Furthermore, the access will be facilitated through a dedicated portal built by the candidate Thematic Core Service Tsunami (cTCS-Tsu) in the framework of EPOS-ERIC infrastructure, that will enable long term access sustainability and ensures standardization towards other hazard components. The services of the cTCS-Tsu have just started and will soon be available through the community portal https://tsunamidata.org/ (presently under construction) created and maintained by INGV.

Scaling of computational resources

Target architecture	Node specs	Target size	Total number of nodes per use-case (can be allocated at different times; uses the Tsunami-HySEA MC version)	Runs per node	Number of runs	Duration of a single run	Data storage needs	Total CPU hours
Marconi100 @CINECA	4 GPU V100	Small or medium	As available to run several 4- to 256-scenario chunks at the same time (805 Marconi100 nodes were allocated for the UC exercise)	4	40 000	30 sec	20 Mb per run	83,33
		large	As available	4	40 000	2 h	20 Mb per run	20 000

References

Basili, R., Brizuela, B., Herrero, A., Iqbal, S., Lorito, S., Maesano, F. E., et al., 2021. The Making of the NEAM Tsunami Hazard Model 2018 (NEAMTHM18). Front. Earth Sci. 8, 1–29. doi:10.3389/feart.2020.616594.

Basili, R., Brizuela, B., Herrero, A., Iqbal, S., Lorito, S., Maesano, F. E., et al., 2018. NEAM Tsunami Hazard Model 2018 (NEAMTHM18): online data of the Probabilistic Tsunami Hazard Model for the NEAM Region from the TSUMAPS-NEAM project. Istituto Nazionale di Geofisica e Vulcanologia (INGV). Available at http://doi.org/10.13127/tsunami/neamthm18

Basili, R., Brizuela, B., Herrero, A., Iqbal, S., Lorito, S., Maesano, F. E., et al., 2019. NEAMTHM18 Documentation: the making of the TSUMAPS-NEAM Tsunami Hazard Model 2018. Zenodo. https://doi.org/10.5281/zenodo.3406625

Basili, R., Tiberti, M.M., Kastelic, V., Romano, F., Piatanesi, A., Selva, J., Lorito, S., 2013. Integrating geologic fault data into tsunami hazard studies. Nat. Hazards Earth Syst. Sci. 13, 1025–1050. https://doi.org/10.5194/nhess-13-1025-2013

Behrens, J., Dias, F., 2015. New computational methods in tsunami science. Philos. Trans. R. Soc. Math. Phys. Eng. Sci. 373, 20140382. https://doi.org/10.1098/rsta.2014.0382

Bernardi, F., Lomax, A., Michelini, A., Lauciani, V., Piatanesi, A., Lorito, S., 2015. Appraising the

Early-est earthquake monitoring system for tsunami alerting at the Italian Candidate Tsunami Service Provider. Nat. Hazards Earth Syst. Sci. 15, 2019–2036. https://doi.org/10.5194/nhess-15-2019-2015

Burbidge, D., Mueller, C., Power, W., 2015. The effect of uncertainty in earthquake fault parameters on the maximum wave height from a tsunami propagation model. Nat. Hazards Earth Syst. Sci. 15, 2299–2312. https://doi.org/10.5194/nhess-15-2299-2015

Carvajal, M., Araya-Cornejo, C., Sepúlveda, I., Melnick, D., Haase, J.S., 2019. Nearly Instantaneous Tsunamis Following the Mw 7.5 2018 Palu Earthquake. Geophys. Res. Lett. 46, 5117–5126. https://doi.org/10.1029/2019GL082578

Davies, G., Griffin, J., Løvholt, F., Glimsdal, S., Harbitz, C., Thio, H.K., Lorito, S., Basili, R., Selva, J., Geist, E., Baptista, M.A. (2018). A global probabilistic tsunami hazard assessment from earthquake sources. Geol. Soc. Lond. Spec. Publ. 456, 219–244. https://doi.org/10.1144/SP456.5

de la Asunción, M., Castro, M.J., Fernández-Nieto, E.D., Mantas, J.M., Acosta, S.O., González-Vida, J.M., 2013. Efficient GPU implementation of a two waves TVD-WAF method for the two-dimensional one layer shallow water system on structured meshes. Comput. Fluids, Selected contributions of the 23rd International Conference on Parallel Fluid Dynamics ParCFD2011 80, 441–452. https://doi.org/10.1016/j.compfluid.2012.01.012

Geist, E.L., 2009. Chapter 3 - Phenomenology of Tsunamis: Statistical Properties from Generation to Runup, in: Advances in Geophysics, Advances in Geophysics. Elsevier, pp. 107–169. https://doi.org/10.1016/S0065-2687(09)05108-5

Glimsdal, S., Løvholt, F., Harbitz, C.B., Romano, F., Lorito, S., Orefice, S., Brizuela, B., Selva, J., Hoechner, A., Volpe, M., Babeyko, A., Tonini, R., Wronna, M., Omira, R., 2019. A New Approximate Method for Quantifying Tsunami Maximum Inundation Height Probability. Pure Appl. Geophys. 176, 3227–3246. https://doi.org/10.1007/s00024-019-02091-w

Glimsdal, S., Pedersen, G.K., Harbitz, C.B., Løvholt, F., 2013. Dispersion of tsunamis: does it really matter? Nat. Hazards Earth Syst. Sci. 13, 1507–1526. https://doi.org/10.5194/nhess-13-1507-2013

Goda, K., Abilova, K., 2016. Tsunami hazard warning and risk prediction based on inaccurate earthquake source parameters. Nat. Hazards Earth Syst. Sci. 16, 577–593. https://doi.org/10.5194/nhess-16-577-2016

Griffin, J., Latief, H., Kongko, W., Harig, S., Horspool, N., Hanung, R., Rojali, A., Maher, N., Fuchs, A., Hossen, J., Upi, S., Edi, D., Rakowsky, N., Cummins, P., 2015. An evaluation of onshore digital elevation models for modeling tsunami inundation zones. Front. Earth Sci. 3. https://doi.org/10.3389/feart.2015.00032

Herrero, A., Murphy, S., 2018. Self-similar slip distributions on irregular shaped faults. Geophys. J. Int. 213, 2060–2070. https://doi.org/10.1093/gji/ggy104

IOC, 2017b. Intergovernmental Oceanographic Commission (IOC), "Tsunami Exercise NEAMWave17 – A Tsunami Warning and Communication Exercise for the North-eastern Atlantic, the Mediterranean, and Connected Seas Region, 31 October – 3 November 2017: Exercise Instructions (IOC Technical Series No.134 Vol.1).

K. Kajiura. "The leading wave of a tsunami", Bull. Earthq. Res. Inst., 41:535–571, 1963.

Kamigaichi, O., 2014. Tsunami Forecasting and Warning, in: Meyers, R.A. (Ed.), Encyclopedia of Complexity and Systems Science. Springer, Berlin, Heidelberg, pp. 1–38. https://doi.org/10.1007/978-3-642-27737-5_568-3

Lorito, S., Selva, J., Basili, R., Romano, F., Tiberti, M.M., Piatanesi, A., 2015. Probabilistic hazard for seismically induced tsunamis: accuracy and feasibility of inundation maps. Geophys. J. Int. 200, 574–588. https://doi.org/10.1093/gji/ggu408

Lynett, P.J., 2009. Tsunami Inundation, Modeling of, in: Meyers, R.A. (Ed.), Encyclopedia of Complexity and Systems Science. Springer, New York, NY, pp. 9618–9631. https://doi.org/10.1007/978-0-387-30440-3_569

Macías, J., Castro, M.J., Ortega, S., Escalante, C., González-Vida, J.M., 2017. Performance Benchmarking of Tsunami-HySEA Model for NTHMP's Inundation Mapping Activities. Pure Appl. Geophys. 174, 3147–3183. https://doi.org/10.1007/s00024-017-1583-1

Macías, J., Mercado, A., González-Vida, J.M., Ortega, S., Castro, M.J., 2016. Comparison and Computational Performance of Tsunami-HySEA and MOST Models for LANTEX 2013 Scenario: Impact Assessment on Puerto Rico Coasts. Pure Appl. Geophys. 173, 3973–3997. https://doi.org/10.1007/s00024-016-1387-8

Maesano, F.E., Tiberti, M.M., Basili, R., 2020. Deformation and Fault Propagation at the Lateral Termination of a Subduction Zone: The Alfeo Fault System in the Calabrian Arc, Southern Italy. Front. Earth Sci. 8. https://doi.org/10.3389/feart.2020.00107

Maesano, F.E., Tiberti, M.M., Basili, R., 2017. The Calabrian Arc: three-dimensional modelling of the subduction interface. Sci. Rep. 7, 8887. https://doi.org/10.1038/s41598-017-09074-8

Molinari, I., Tonini, R., Lorito, S., Piatanesi, A., Romano, F., Melini, D., Hoechner, A., Gonzàlez Vida, J.M., Maciás, J., Castro, M.J., Asunción, M. de la, 2016. Fast evaluation of tsunami scenarios: uncertainty assessment for a Mediterranean Sea database. Nat. Hazards Earth Syst. Sci. 16, 2593–2602. https://doi.org/10.5194/nhess-16-2593-2016

Murphy, S., Scala, A., Herrero, A., Lorito, S., Festa, G., Trasatti, E., Tonini, R., Romano, F., Molinari, I., Nielsen, S., 2016. Shallow slip amplification and enhanced tsunami hazard unravelled by dynamic simulations of mega-thrust earthquakes. Sci. Rep. 6, 35007. https://doi.org/10.1038/srep35007

Nosov, M.A., Kolesov, S.V., 2011. Optimal Initial Conditions for Simulation of Seismotectonic Tsunamis. Pure Appl. Geophys. 168, 1223–1237. https://doi.org/10.1007/s00024-010-0226-6

Okada, Y., 1985. Surface deformation due to shear and tensile faults in a half-space. Bull. Seismol. Soc. Am. 75, 1135–1154.

Sarri, A., Guillas, S., Dias, F., 2012. Statistical emulation of a tsunami model for sensitivity analysis and uncertainty quantification. Nat. Hazards Earth Syst. Sci. 12, 2003–2018. https://doi.org/10.5194/nhess-12-2003-2012

Satake, K., Fujii, Y., Harada, T., Namegaya, Y., 2013. Time and Space Distribution of Coseismic Slip of the 2011 Tohoku Earthquake as Inferred from Tsunami Waveform DataTime and Space Distribution of Coseismic Slip of the 2011 Tohoku Earthquake. Bull. Seismol. Soc. Am. 103, 1473–1492. https://doi.org/10.1785/0120120122

Scala, A., Lorito, S., Romano, F., Murphy, S., Selva, J., Basili, R., Babeyko, A., Herrero, A., Hoechner, A., Løvholt, F., Maesano, F.E., Perfetti, P., Tiberti, M.M., Tonini, R., Volpe, M., Davies, G., Festa, G., Power, W., Piatanesi, A., Cirella, A., 2020. Effect of Shallow Slip Amplification Uncertainty on Probabilistic Tsunami Hazard Analysis in Subduction Zones: Use of Long-Term Balanced Stochastic Slip Models. Pure Appl. Geophys. 177, 1497–1520. https://doi.org/10.1007/s00024-019-02260-x

Selva, J., Tonini, R., Molinari, I., Tiberti, M.M., Romano, F., Grezio, A., Melini, D., Piatanesi, A., Basili, R., Lorito, S., 2016. Quantification of source uncertainties in Seismic Probabilistic Tsunami Hazard Analysis (SPTHA). Geophys. J. Int. 205, 1780–1803. https://doi.org/10.1093/gji/ggw107

Selva, J., Lorito, S., Volpe, M. et al. Probabilistic tsunami forecasting for early warning. Nat Commun 12, 5677 (2021). https://doi.org/10.1038/s41467-021-25815-w.

Song, J., Goda, K., 2019. Influence of Elevation Data Resolution on Tsunami Loss Estimation and Insurance Rate-Making. Front. Earth Sci. 7. https://doi.org/10.3389/feart.2019.00246

Volpe, M., Lorito, S., Selva, J., Tonini, R., Romano, F., Brizuela, B., 2019. From regional to local SPTHA: efficient computation of probabilistic tsunami inundation maps addressing near-field sources. Nat. Hazards Earth Syst. Sci. 19, 455–469. https://doi.org/10.5194/nhess-19-455-2019

PD 9. Seismic Tomography

PD 12	High-resolution volcanic ash dispersal forecast
Leader	CNRS
Workflow	specfem_fwi
Numerical Engine	SPECFEM3D, PySpecfem
TRL achieved	6
Access Mode	By request

Introduction

This deliverable on the use of PD9 comes after two other deliverables, D4.9 which presents the PD9 workflow, and deliverable D5.3 which presented two cases to validate PD9. Here we illustrate the tutorial with data from Japan that was not presented in D5.3, which is the last use case to be presented in ChEESE. We do not describe the workflow of PD9 here, for which we request the reader to refer to D4.9. Here we explain how to handle PD9 in practice.

Installation

PD9 is based on two tools, a python tool for pre and post processing of data and models PySpecfem and an HPC tool based on spectral elements specfem_fwi. The procedure for installing these two tools is described in this section.

PySpecfem

we recommend installing PySpecfem with python3.6 3.7 or 3.8, we have not tested with 3.9. We recommend to use pip in this case the installation is automatic, the dependencies will be automatically installed by pip. In the pyspecfem repository just type,

```
> pip install .
```

Some tools will be installed, including :

pyspecfem_setup

pyseseiswaves

pyspecfem_create_initial_smooth_model

pyspecfem_make_report_on_event_all_traces

pyspecfem_compute_pert_xdmf

Their use will be described in the following.

specfem_fwi

The installation of specfem_fwi requires fortran, c and nvcc compilers for GPU computing and a MPI library. A compiler help script is provided with specfem_fwi to configure the different compilers. It is necessary to inform the name of the various compilers and possibly the paths to the cuda and MPI libraries if the system does not have information on them. In addition, FFTW and hdf5 can be used, in which case the corresponding variables must be entered. Here is an example of the compilation script:

```
#!/bin/bash
####### CPU #######
fc=ifort
cc=icc
mpifc=mpif90

# mpi include diectory
mpi_inc=/usr/lib/x86_64-linux-gnu/openmpi/include

##### FFTW #######
FFTW_DIR=/usr/local
#or
FFTW_INC=
FFTW_LIB=

#### HDF5 #####
#-with-hdf5
mpif90=h5pfc

#####  GPU ######
# for compiling wih cuda specify the following option with the cuda flag decribed below

# nvidia GPU option
#
cuda_flag=cuda10

#   our options for compiling in differents GPU nvidia architecture :
#
# cuda10 Turing RTX500, RTX4000, RTX***
# cuda9  Volta V100,
# cuda8  Pascal P100, P***
# cuda(5 or 6)  Kepler K80
# cuda5 kepler K20
#

# configure to create Makefile
./configure --with-mpi CC=$cc FC=$fc MPIFC=$mpifc MPI_INC=$mpi_inc --with-fftw3-dir=$FFTW_DIR --with-cuda=$cuda_flag --with-hdf5

# compile the full package
make all
```

Data preprocess

The preprocessing of the seismic data is to select the traces and to eliminate noisy traces that could pollute the inversion. The traces from an event are stored in a separate directory which contains all the traces and the information about the seismic source. For this purpose, the user can refer to the well-known obspy (https://docs.obspy.org/) package to download traces and event information.

Data preparation

The seismological data can be stored in two formats sac ( ) or miniseed (A GUI tool, pyseiswaves, is used to select the data and save the resulting traces in the format used by specfem_fwi.

File format

The following files are produced by pyseiswaves and are needed for the inversion step based on specfem_fwi.

data.bin : binary file which contains the data

data.hdr : meta data for data.bin

STATIONS_geogr : stations information (network, name, position)

STATIONS_weight : weight on each trace (0 => not used, 1=> used)

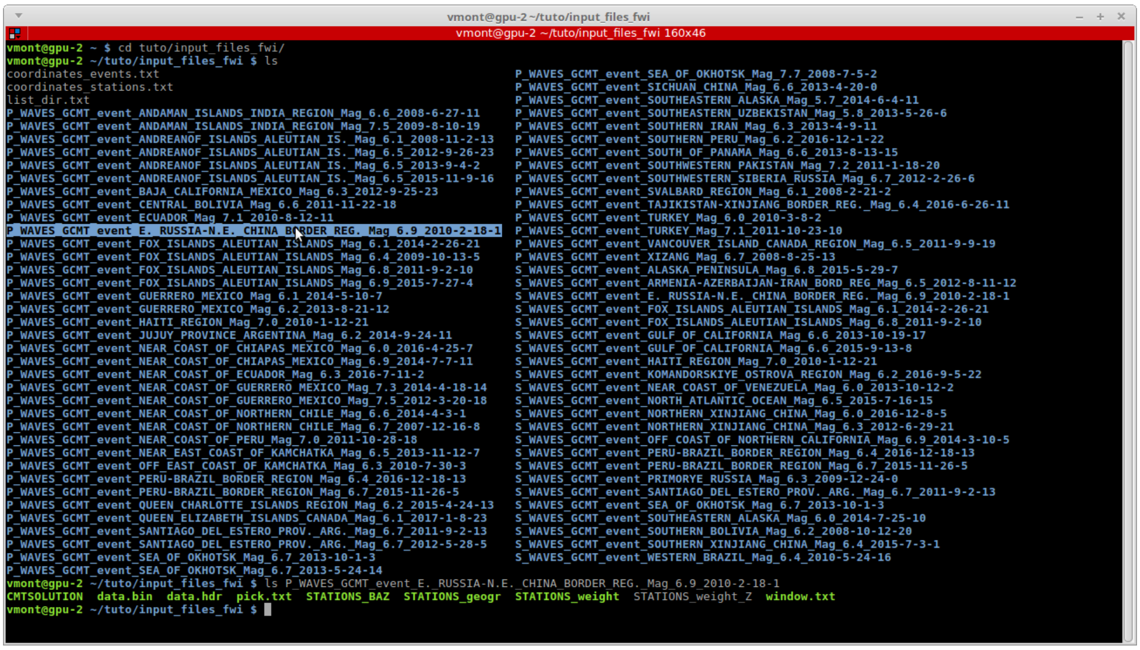
STATIONS_BAZ : back azimuth for each stations

pick.txt : arrival time of the seismological phase of interest (eg: P, S)

window.txt : time window to invert with respect to the pick

CMTSOLUTON : information on earthquake (position, moment tensor)

After this first step, all the data are put inside a directory called input_files_fwi/ which contains one directory per event and each directory contains the previous files :



pyseiswave usage

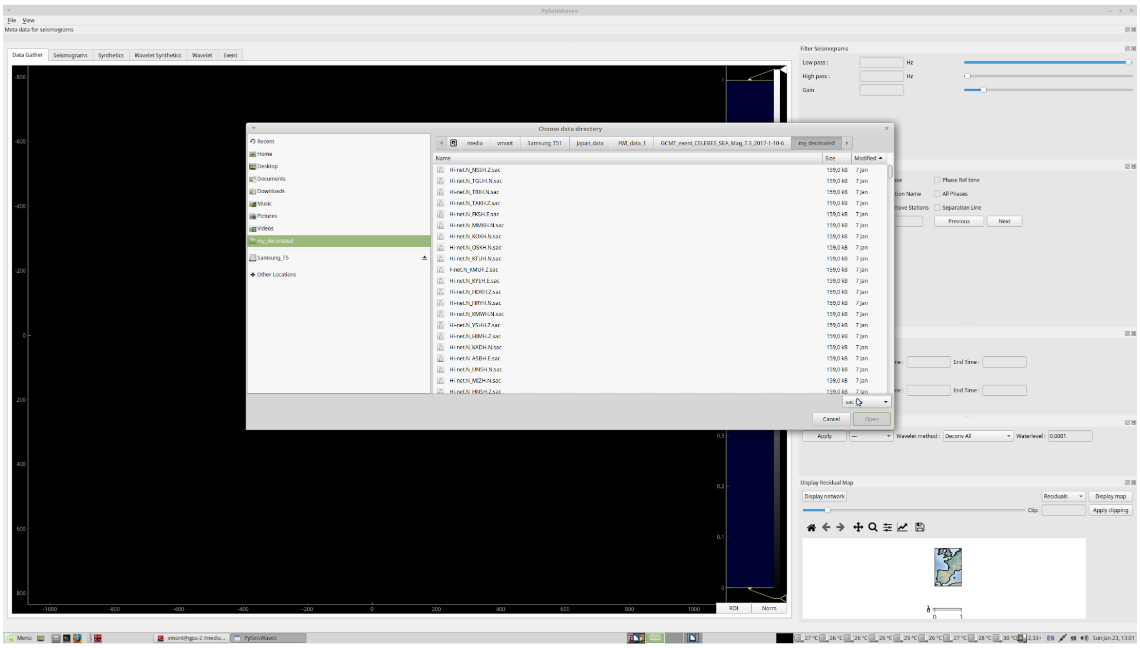
Displaying data

Launch pyseiswave by just typing in comand teminal.

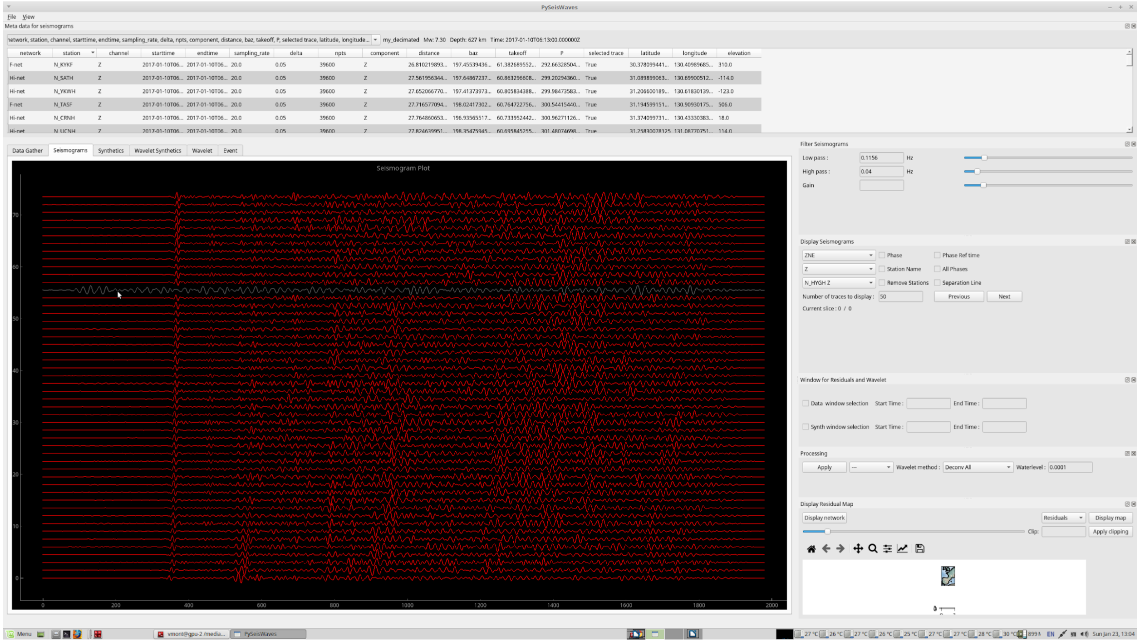
Reading data by using menu tool bar

“File”→”Open Mseed/SAC Data”

Then select the directory with data and select file format.



After loading all data, the traces are displayed in the “**Seismogram”** tab.

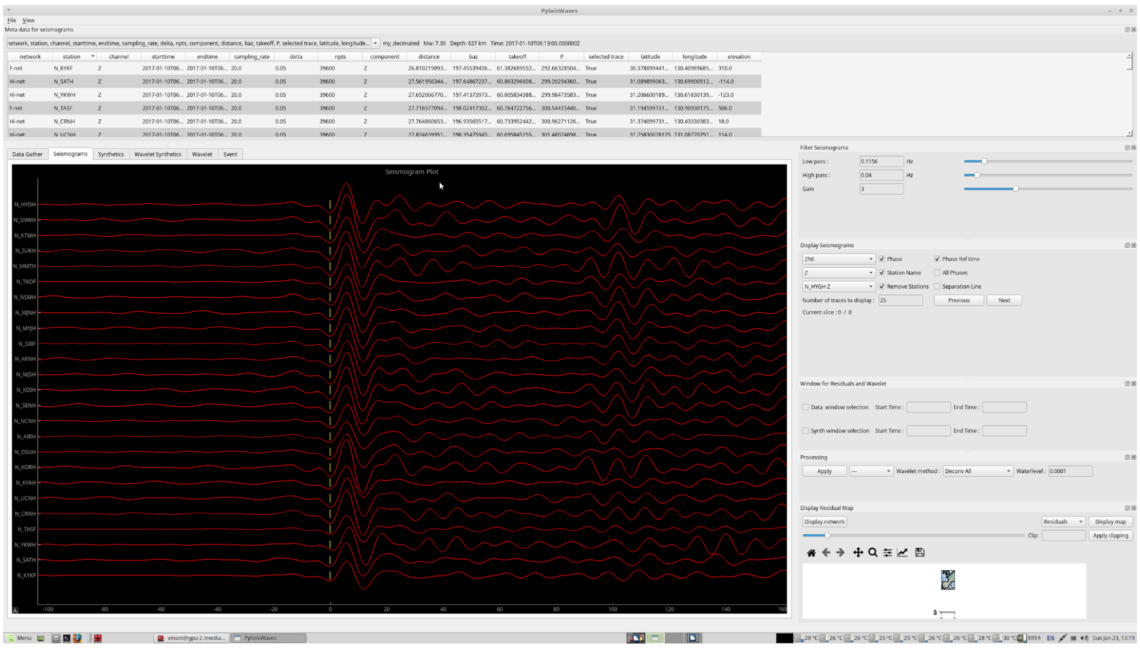


At the right the “**Filter Seismogram”** panel allows to filter the data and display with different gain.

A mouse click on a trace will select it for discarding, the trace then appears in grey. A first visual inspection of data allows the user to select and remove a spurious data by a simple mouse click.

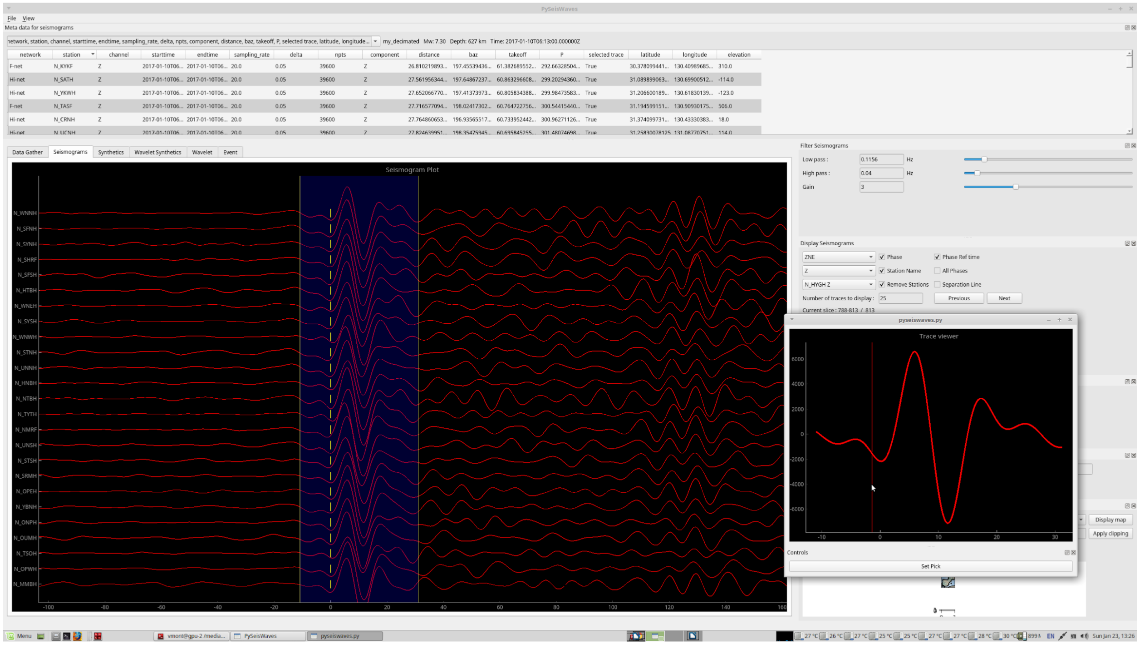
Quality control metric

For the seismic tomography purpose, the data must be carefully cleaned for this purpose after a first visual inspection, some quality control metric can be used to help find and remove the remaining bad data. First focus on phase of interest by checking in “**Display Seismograms”** panel : “**Phase”**, “**Phase Ref time”**, “**Remove Stations”**, “**Station Name”**.

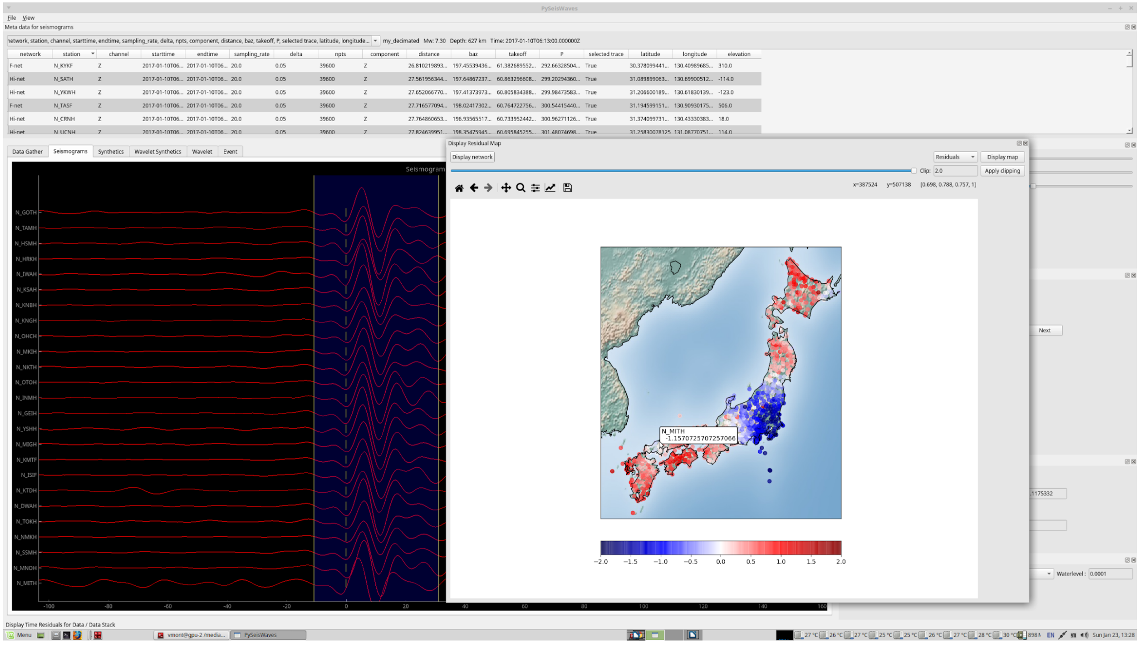


The theoretical arrival time of phase of interest (P teleseismic in the figure) is then plotted in each traces. A mouse scroll allow to zoom on the phase.

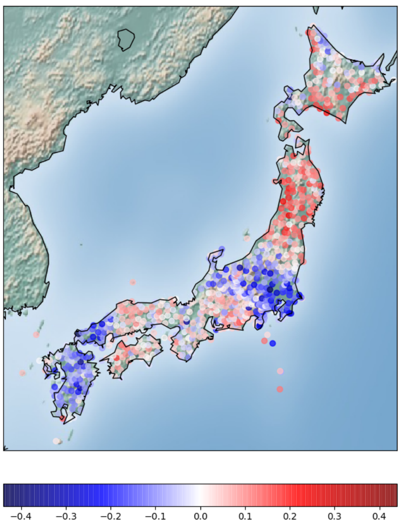
To compute the quality control metric, check “**Data window selection”** in “**Window for residual and wavelet”** panel and select the window by adjusting the selector tool that appears in the “**Seismograms”** tab.

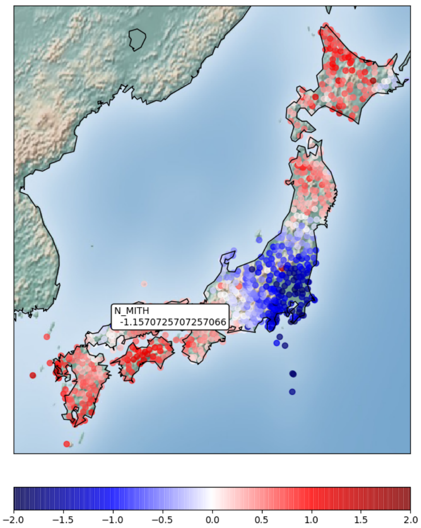


In processing panel select “**Data CC Align/stack pick”**, and push “**Apply”** button. A new figure is then displayed with the stack of all traces in the selected window. Move the vertical bar with mouse to select a reference time in the stacked wavelet and the push the “**Set Pick”** button. The tool will process some metric computation then by pushing the button “**Display network”** and “**Display map”** in the “**Display Residual Map”** panel. The map will be updated, each station position appears as a coloured point. This allow a visual inspection and some spurious colours, (eg : a blue spot in red patch, or red spot in blue patch) need to be inspected by clicking on the point. Then the corresponding trace is displayed at the bottom of “**Seismograms”** tab. The user can make a decision to remove or keep this data.



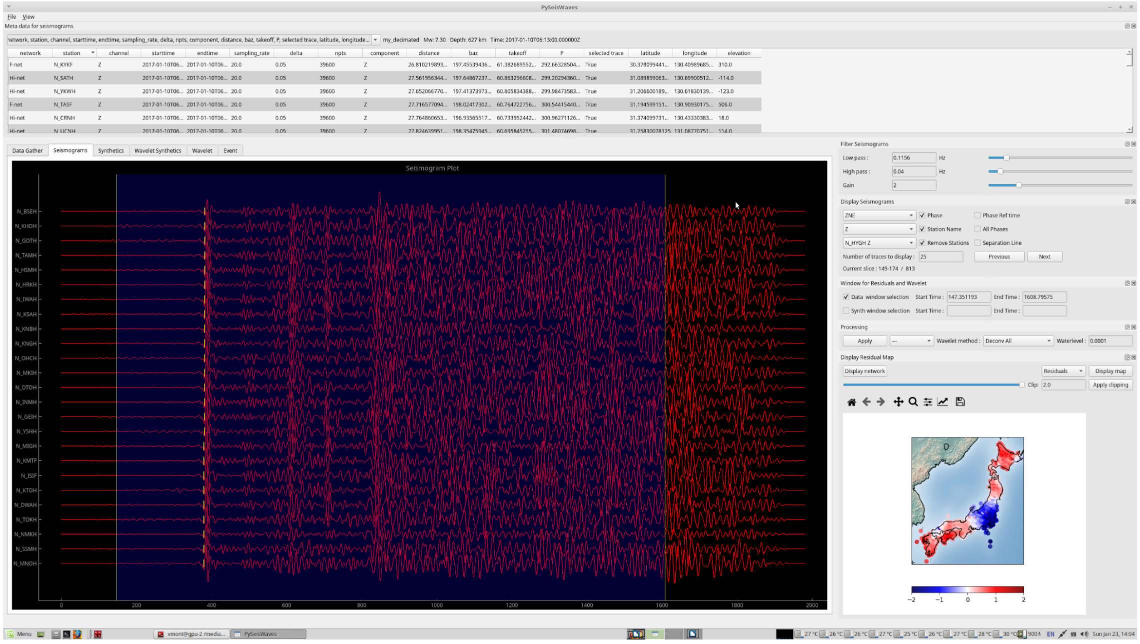
Four metrics are available, phase residual, amplitude residual, correlation and signal on noise ratio. In the following figures the phase residuals are displayed at the right and amplitude misfit at the left.





Saving input data for specfem_fwi

After cleaning data set unselect in "**Display Seismograms**" panel "**Phase Ref time**". In Processing panel select "**—**" and click on "**Apply**" button. Then select the time window to cut the data for the seismic tomography.



For saving the selected data set, click in the menu bar:

“File”→ “Save Data (binary)”

All the files needed for specfem_fwi will be writen in the chosen directory. After processing all events the data are ready for the first step of the inversion.

Seismic tomography

This section describe how to launch the inversion process. The first step is to upload the data on the cluster and prepare the setup file. This setup file is described in appendix A1. To prepare all the files and scripts for the process use the following command:

> pyspecfem_setup setup_file.in

where setup_file.in is the name of the setup file used for the case.

Synthetics computation

Preparing initial model

The initial guess model for seismic tomography is a smoothed version of a layered classical seismological model (eg PREM, IASP91, AK135, …) to produce it, type this command :

> pyspecfem_create_initial_smooth_model –options

where options can be defined by the user. The options available are the following (use –help):

usage: pyspecfem_create_initial_smooth_model [-h]

[--topo_file TOPO_FILE]

[--topo_type TOPO_TYPE]

[--lon_center LON_CENTER]

[--lat_center LAT_CENTER]

[--azi AZI]

[--lon_extension LON_EXTENSION]

[--lat_extension LAT_EXTENSION]

[--interval INTERVAL]

[--filter_size FILTER_SIZE]

[--nzkeep NZKEEP]

Create input smooth model on regular grid

optional arguments:

-h, --help show this help message and exit

--topo_file TOPO_FILE

file with topography [/gpfswork/rech/gkj/rukn006/topo/etopo1_bed_g_f4.flt]

--topo_type TOPO_TYPE

type of file with topography [etopo1]

--lon_center LON_CENTER

longitude center of chunk [-3.]

--lat_center LAT_CENTER

latitude center of chunk [39.]

--azi AZI

azimuth of chunk [0.]

--lon_extension LON_EXTENSION

extension of chunk in longitude [16.5]

--lat_extension LAT_EXTENSION

extension of chunk in latitude [23.5]

--interval INTERVAL

size of the grid size in meters [10000.]

--filter_size FILTER_SIZE

the number of points which is used for smoothing [40]

--nzkeep NZKEEP

number of the elements from the surface which keeps the value without smoothing [5]

for example,

> pyspecfem_create_initial_smooth_model --lon_center 136. --lat_center 37.5 --azi " -50." --lon_extension 24. --lat_extension 15.5

This command produces 4 files which contains all informations for the initial guess model :

fd_grid.txt   : desciption of the regular grid

rho_smooth.bin : values of density

vp_smooth.bin : values of P-wave velocity

vs_smooth.bin : values of S-wave velocity

Creating the mesh

To create the mesh a script which launch all the specfem_fwi meshers tools is used by typing:

> ./launch_mesher.sh

This script schedule the *xmeshfem3D* code and *xgenerate_databases* using the scheduler installed in the cluster.

forward simulation

The forward simulation is simply launched by the command :

> sbatch run_FWI_forward.batch

After the forward simulation run, for each event synthetics are computed and stored in same directory where are the data in a binary file : synth_smooth_ak135.bin, with the same format than data.

Source time function estimation

The source estimation is the first step of the inversion when the model is remain fixed and only the source time function of the event is estimated by a least square inversion comparing the data and previously computed synthetics. The psyseiswave GUI is used to perform this step. In the menu toolbar,

“File” → “Load binary data”

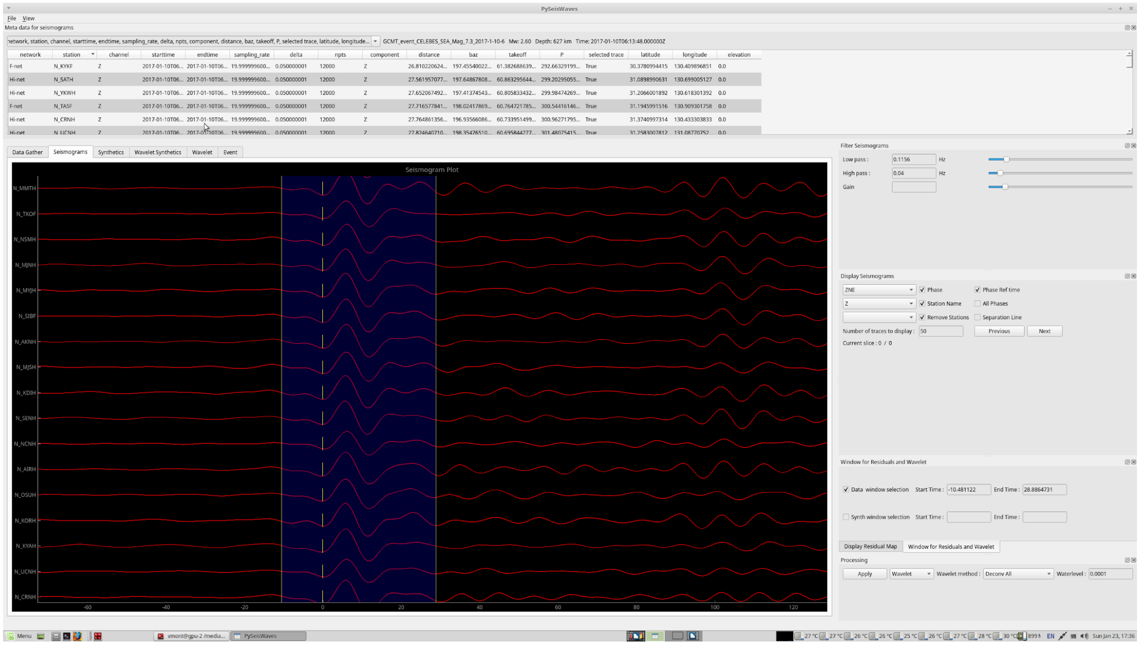
then select binary file data and the corresponding meta data file with *.hdr extension. After the complete loading of data, return in menu toolbar to load synthetics,

“**File**” → “**Load binary Synthetics**”

then select synthetics and the correspondig meta data file with *.hdr extension.

In the “**Seismograms”** tab the data are displayed and the synthetics are in the “**Synthetics”** tab.

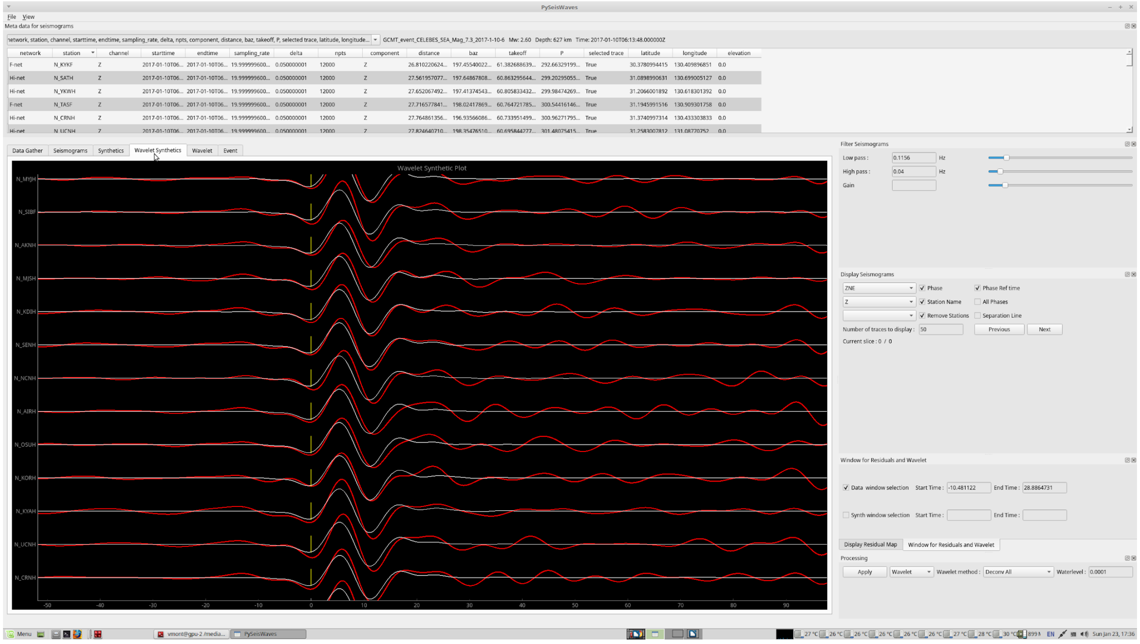
Now select a time window around the phase of interest on data, for that select first in “ “**Display Seismograms”** panel : “**Phase”**, “**Phase Ref time”**, “**Remove Stations”**, “**Station Name”**. Then select the window with the interactive data selection control.



Now the gui is ready to compute the source time function, to do that select “**Wavelet**” in “**Processing panel”** and push the “**Apply**” button. The source time function obtained is displayed in the “**Wavelet**” tab and synthetics convolved with the wavelet and data are displayed together in the the “**Wavelet Synthetics**” tab. The data appears in red and synthetics in white.   To save the source time function click on menu toolbar,

“File” → “Save Wavelet”

The wavelet is written in wavelet.txt file in the event directory.



Full Waveform Inversion

Launching inversion

The purpose of full waveform inversion is now to minimize the misfit between the data and synthetics by modifying the model. Before launching the inversion, the wavelet files computed before must be uploaded in event directory where the data are stored on the HPC cluster. The file inversion_fwi_par described in appendix A.2 is used to tune the different parameters involved in the inversion. The important parameters that have an impact on the solution are the frequency band used and the smoothing. It is recommended to start at the lowest possible frequency (0.1 Hz for the teleseismic case) and to increase the frequency gradually, 0.12, 0.15, 0.2 Hz. To do this, simply fill in the fields in the file inversion_fwi_par with several frequency bands. It is recommended to start by smoothing much more than 100 km and when the model is more accurate, this smoothing can be reduced by a factor of 2 or more depending on the level of noise in the data. The noisier the data, the larger the smoothing should be.

Monitoring iterations

The cost function during iterations is stored in the file output_iteration.txt.

iter | cost function | gradient norm | relat cost | relat grad| nb line search

```
 FREQUENCY GROUP :  3.9999999E-02  0.1000000

  0| 0.31801636E+04| 0.32137892E+10| 0.10000E+01| 0.10000E+01|
  1| 0.29109644E+04| 0.16411451E+10| 0.91535E+00| 0.51066E+00|  1
  2| 0.26785173E+04| 0.57882547E+09| 0.84226E+00| 0.18011E+00|  1
  3| 0.25527654E+04| 0.22598426E+09| 0.80272E+00| 0.70317E-01|  1
  4| 0.24700425E+04| 0.12514536E+09| 0.77670E+00| 0.38940E-01|  1
  5| 0.24021162E+04| 0.10127956E+09| 0.75534E+00| 0.31514E-01|  1
  6| 0.23394854E+04| 0.86354064E+08| 0.73565E+00| 0.26870E-01|  1
  7| 0.22826199E+04| 0.55500520E+08| 0.71777E+00| 0.17269E-01|  1
  8| 0.22333950E+04| 0.41608764E+08| 0.70229E+00| 0.12947E-01|  1
```

```
 9| 0.21949937E+04| 0.52968624E+08| 0.69021E+00| 0.16482E-01|  1
10| 0.21740005E+04| 0.30048956E+08| 0.68361E+00| 0.93500E-02|  1
11| 0.21537427E+04| 0.18997396E+08| 0.67724E+00| 0.59112E-02|  1
12| 0.21286589E+04| 0.15205215E+08| 0.66936E+00| 0.47312E-02|  1
13| 0.21027058E+04| 0.14745280E+08| 0.66119E+00| 0.45881E-02|  1
14| 0.20775898E+04| 0.30293150E+08| 0.65330E+00| 0.94260E-02|  1
15| 0.20580979E+04| 0.16897050E+08| 0.64717E+00| 0.52577E-02|  1
16| 0.20448235E+04| 0.12351874E+08| 0.64299E+00| 0.38434E-02|  1
17| 0.20261509E+04| 0.79018350E+07| 0.63712E+00| 0.24587E-02|  1
18| 0.20041323E+04| 0.55027610E+07| 0.63020E+00| 0.17122E-02|  1
19| 0.19845096E+04| 0.12265880E+08| 0.62403E+00| 0.38166E-02|  1
20| 0.19726245E+04| 0.97775460E+07| 0.62029E+00| 0.30424E-02|  1
21| 0.19592942E+04| 0.82504650E+07| 0.61610E+00| 0.25672E-02|  1
22| 0.19429418E+04| 0.13943988E+08| 0.61096E+00| 0.43388E-02|  1
23| 0.19287456E+04| 0.69805965E+07| 0.60649E+00| 0.21721E-02|  1
24| 0.19186587E+04| 0.51471280E+07| 0.60332E+00| 0.16016E-02|  1
25| 0.19051600E+04| 0.43084080E+07| 0.59908E+00| 0.13406E-02|  1
26| 0.18891104E+04| 0.48651890E+07| 0.59403E+00| 0.15138E-02|  1

FREQUENCY GROUP :  3.9999999E-02  0.1200000

 0| 0.16941924E+04| 0.10263220E+08| 0.10000E+01| 0.10000E+01|
 1| 0.16907869E+04| 0.18757286E+08| 0.99799E+00| 0.18276E+01|  5
 2| 0.16868927E+04| 0.11067442E+08| 0.99569E+00| 0.10784E+01|  5
 3| 0.16833455E+04| 0.41378188E+07| 0.99360E+00| 0.40317E+00|  1
 4| 0.16773461E+04| 0.13777090E+08| 0.99006E+00| 0.13424E+01|  1
 5| 0.16703187E+04| 0.13780261E+08| 0.98591E+00| 0.13427E+01|  1
 6| 0.16656694E+04| 0.38308768E+08| 0.98316E+00| 0.37326E+01|  1
 7| 0.16620652E+04| 0.16377900E+08| 0.98104E+00| 0.15958E+01|  1
 8| 0.16592399E+04| 0.21453802E+08| 0.97937E+00| 0.20904E+01|  1
 9| 0.16581714E+04| 0.17201162E+08| 0.97874E+00| 0.16760E+01|  1
10| 0.16560343E+04| 0.18342256E+08| 0.97748E+00| 0.17872E+01|  1
11| 0.16555096E+04| 0.17983568E+08| 0.97717E+00| 0.17522E+01|  1
12| 0.16537134E+04| 0.18645348E+08| 0.97611E+00| 0.18167E+01|  1
13| 0.16536510E+04| 0.20093512E+08| 0.97607E+00| 0.19578E+01|  1
14| 0.16522853E+04| 0.21830480E+08| 0.97526E+00| 0.21271E+01|  1
15| 0.16468512E+04| 0.95345762E+06| 0.97206E+00| 0.92900E-01|  2
16| 0.16448369E+04| 0.96949056E+06| 0.97087E+00| 0.94463E-01|  2
17| 0.16398049E+04| 0.15627898E+08| 0.96790E+00| 0.15227E+01|  1
18| 0.16283970E+04| 0.12537640E+08| 0.96116E+00| 0.12216E+01|  3
19| 0.16207343E+04| 0.42303805E+07| 0.95664E+00| 0.41219E+00|  1
20| 0.16170309E+04| 0.19001239E+07| 0.95446E+00| 0.18514E+00|  1
21| 0.16130270E+04| 0.95427300E+06| 0.95209E+00| 0.92980E-01|  1
22| 0.16083943E+04| 0.13204686E+07| 0.94936E+00| 0.12866E+00|  1
23| 0.16034520E+04| 0.37238735E+07| 0.94644E+00| 0.36284E+00|  1
24| 0.15980964E+04| 0.42648505E+07| 0.94328E+00| 0.41555E+00|  1
25| 0.15897390E+04| 0.54420290E+07| 0.93835E+00| 0.53025E+00|  1
26| 0.15828480E+04| 0.17829610E+08| 0.93428E+00| 0.17372E+01|  1
```

The left-hand column is the iteration number, then comes the value of the cost function, the norm of the gradient and the relative values of the cost and the norm of the gradient.

Post-process

This section describe how to visualize the inverted model and how to control the data fit.

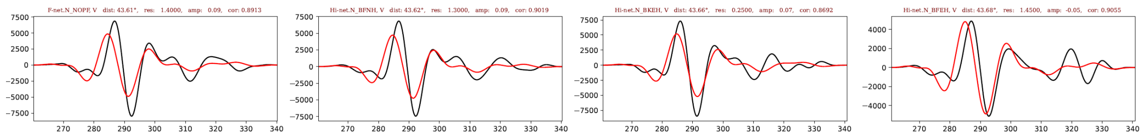
Data fit

The adjustment of the data can be controlled using a pyspecfem command pyspecfem_make_report_on_event_all_traces.

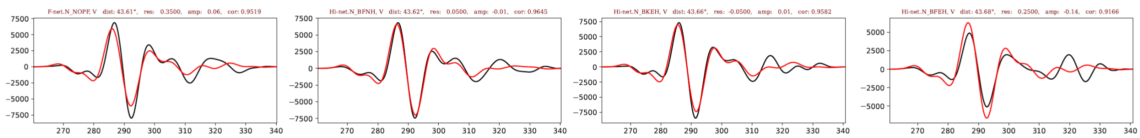
The option –help display all the options available. For example to produce a report on data fit,

> pyspecfem_make_report_on_event_all_traces --event_directory GCMT_event_CELEBES_SEA_Mag_7.3_2017-1-10-6/ --fname_syn data_for_fwi.bin1030_dir --use_FWI_window --frq_max 0.1

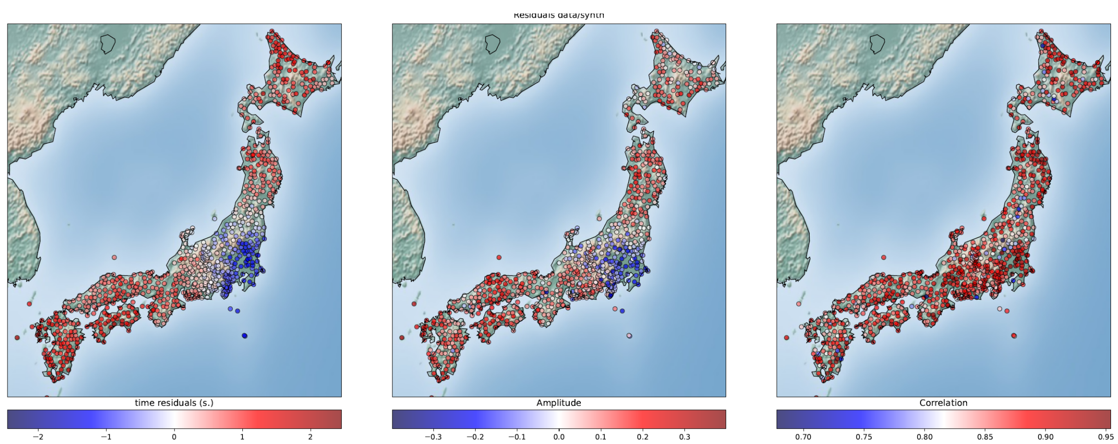
The result is a pdf file (event_report_all_traces.pdf) that contains the geometry network, the data fit by plotting each trace for a given component and quality metric maps. The following figure are extracted form the report produced, the synthetics traces (red) are compared with the data (black) in initial model :



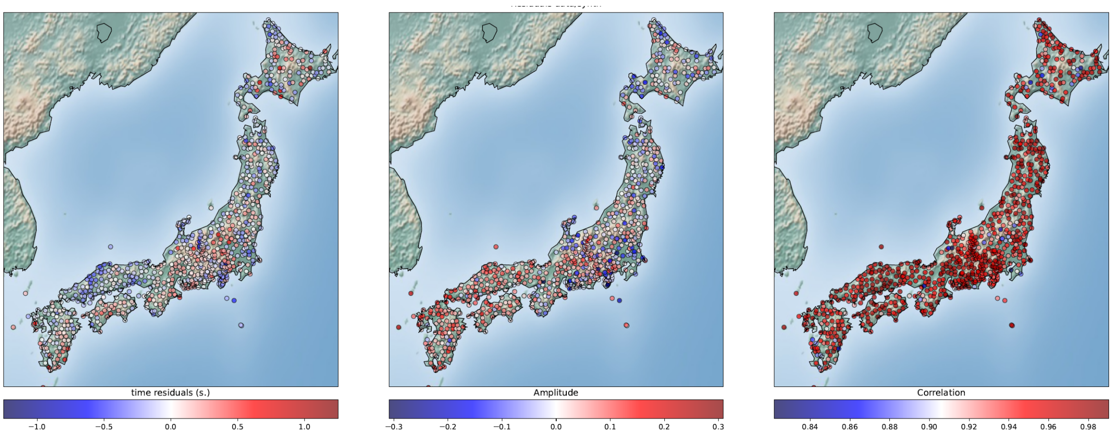
and in the final model, with a better agreement between synthetics and data:



Also a map displaying residuals, amplitude misfit and correlation in initial model :



And for the final model :



Thus the metrics are improved by the inversion and the resulting model explains the data much better than the original model.

Model visualization

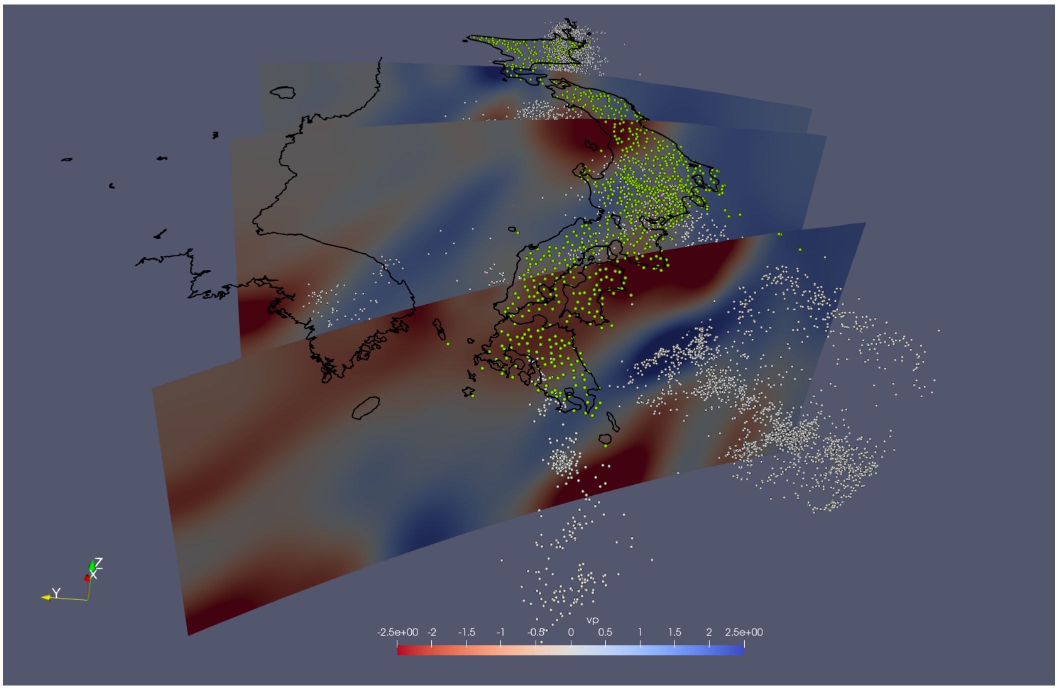
The results of each iteration are stored in the VISU directory. For example the file VISU/model__iter_20026.xdmf means that the model corresponds to iteration 26 of frequency band 2. To view the model, simply open this file with the open source code paraview (https://www.paraview.org/).

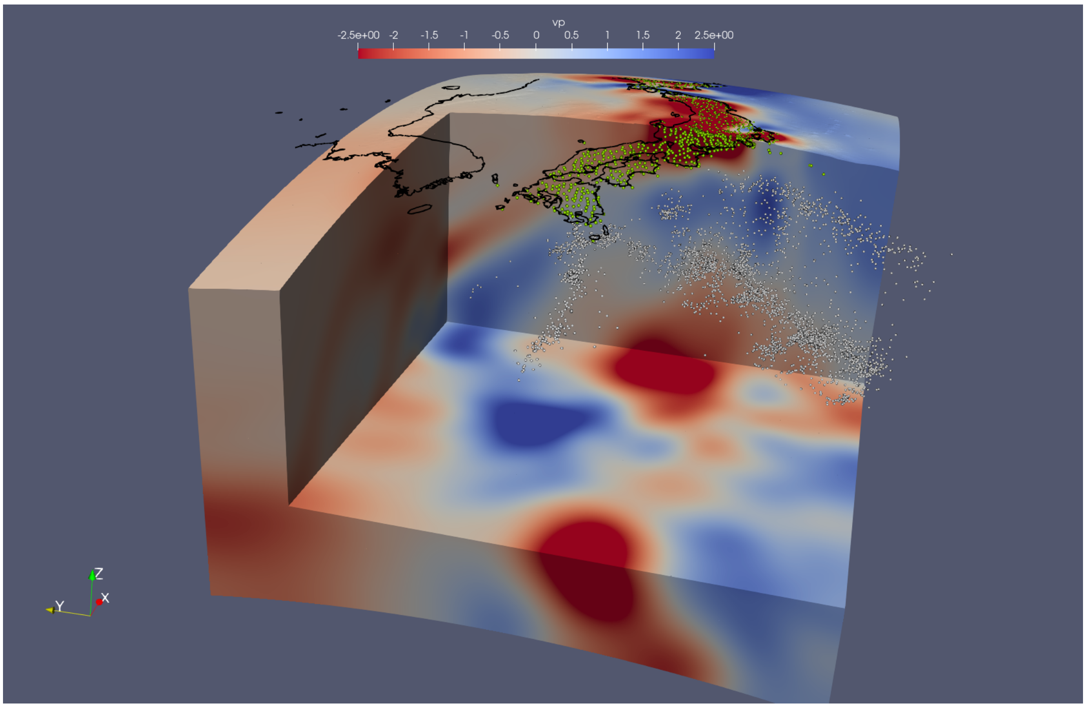
Visualization in specfem_fwi mesh

To visualize the perturbations of the model with respect to the reference model, it is necessary to use a utility that will calculate this difference (in %) and store it in the file VISU/pert_model.xdmf. For example the command,

```
> pyspecfem_compute_pert_xdmf –cur model__iter_20026.xdmf
```

computes the perturbations of the iteration 26 model in the frequency band 2 and the visualization with paraview can be done by opening the file pert_model.xdmf. The paraview code allows to make 3D visualization, make slices in the model as on the following figures and also allows to plot the stations (green dots) and also the seismicity (white dots) as well as the coastlines (in black).





Other output formats

We have other options for output file formats, when the hdf5_enabled flag in the setup file is set to 1, all output merges to hdf5 which allows for better portability. In addition we can also use a regular grid for model output rather than the spectral element grid. In this case we need to set the ouput_fd_model flag in the inversion_fwi_par file to .true.

Appendix

A.1 setup file

An example of setup file is given below. Line with # is a comment. The file contains keywords and values separated by a “:”

```
#################################################################
#
#                 Domain for specfem
#             meshing Chunk with meshfem                  #
#################################################################

# use coupling specfem and axisem
axisem_coupling : 1
mesh_chunk : 1
use_new_script_axisem : 1
model_type : ak135
ngnod : 27

# ---------------------------------------------------
#  Domain in spherical coordiantes
#  since axisem use geocentric coordiantes
#  we assume that the following values are
# given in geocentric coordinates.

# center of domain lat, lon  (decimal degrees)
LONGITUDE_CENTER_OF_CHUNK : 136.
LATITUDE_CENTER_OF_CHUNK :  37.5

# extention of chunk before rotation (decimal degrees)
LONGITUDE_EXTENTION : 24
LATITUDE_EXTENTION : 15.5

# vertical size (km)
VERTICAL_EXTENTION : 900

# rotation of domain with respect to z axis in the center  # of domain given with
respect to the north
# (0. means no rotation)
AZIMUTH_OF_CHUNK : -50

# depth for buried box (km)
# set 0. for domain that reach the free surface of the
# earth
DEPTH_OF_CHUNK : 0.


#-------------------------------------------------------
# topography file path  (comment to unuse it)
```

```
# pyspecfem will automatically download ETOPO1 or
# SRTM15_PLUS file from NOAA server if not found at the
# specified path.
ETOPO1 : home/topo/etopo1_bed_g_f4.flt

#-------------------------------------------------------
# discretization of domain in km
SIZE_EL_LONGITUDE_IN_KM : 20.0
SIZE_EL_LATITUDE_IN_KM  : 20.0

# depth
SIZE_EL_DEPTH_IN_KM : 20.0

# doubling
#number_of_doubling : 1
#depth_of_doubling : -47.5

# tripling
number_of_tripling : 1
depth_of_tripling : -55.

##############################################################
#
#                  AXISEM CONFIG                            #
##############################################################

# background model used in axisem : ak135, iasp91, prem
EARTH_MODEL : ak135

# Dominant period [s]
DOMINANT_PERIOD : 8

# MPI decomposition for AXISEM
NTHETA_SLICES :  20
NRADIAL_SLICES : 2

SOURCE_FUNCTION : errorf

##############################################################
#
#                SOURCES - RECEIVERS
#
##############################################################

SIMULATION_TYPE : moment
```

```
input_directory : fwi_input_files

# path to station file (format :
# station_name  network_name latitue(deg) longitude(deg)
# elevation(m) buried(m)
station_type : geo_file

############################################################
#
#                SPECFEM CONFIG                      #
############################################################

# MPI decomposition for SPECFEM
number_of_mpi_domain : 4

# simultanteous runs
number_of_simultaneous_runs : 1

# simulation time sampling
time_length : 500
time_step : 0.025
time_step_data : 0.05

############################################################
#
#                COUPLING CONFIG                     #
############################################################


mpi_process_coupling : 80

############################################################
#
#                  PATHS                                                      #
############################################################

# PATH TO SPECFEM DIRECTORY
specfem_directory : /path/to/specfem_fwi
AXISEM_COUPLING_DIR : /path/to/specfem_fwi/src/extern_packages/coupling/

############################################################
#
#                  SCRIPTS                      #
############################################################
script_type : slurm
use_inverse_problem : 1
```

use_gpu : 1
use_input_fwi : 1
hdf5_enabled : 0
number_of_io_node : 0

A2. inversion_fwi_par file

The inversion_fwi_par file contains all the parameter that control the inversion, like stopping criteria, smoothing, damping, output controls. Frequency band to use in inversion. Below, this is an example of such file with line begins with "#" are comments and the other lines are composed by a key and value separated by ":"

```
############# stopping criteria ###########################
Niter : 25           # maximum FWI iteration allowed
relat_grad : 1e-5    # relative decrease on gradient
relat_cost : 1e-5    # relative decrease on cost function

##########   cost function  #################################
prior_data_std : 0.001    # estimated error on data
normalize_data :  .true.  # normalize data by maximum of
                    # the traces
max_relative_pert : 0.9   # maximum relative perturbation
                    # allowed for line search

######## parameter to invert ###############################
# choice : ISO=(rho, vp, vs), VTI=(rho, vp, vs, ep, gm, de)
param_family : ISO
nb_inver : 3
param_to_inv : rho vp vs

# damping of each parameter in [0. 1.] (1 => allows 100%
# variation)
normalize_inv_parameter : 0.1 0.1 0.1
# smoohting length on model (meters)
laplac_filter : 150000. 150000. 150000.

#############  Band frequency ##############################
use_frequency_band_pass_filter : 2
fl : 0.04  0.04  # min frequency Hz
fh : 0.1   0.12  # max frequensy Hz

########## PRECOND  ######################################
# gradient=0 outisde domain : xmin, xmax, ymin, ymax, zmin,
# zmax defined below
taper : -1334339. 1334339. -861760.7 861760.7 -1135664.  0.

########### I/O options ###################################
```

input_fd_model : .true. **# read input model on regular grid**
output_model : .true. **# write ouput model on the spectral**
 # element mesh

#**########### save some fields at each iteration ##########**
dump_model_at_each_iteration : .true.
dump_gradient_at_each_iteration : .true.
dump_descent_direction_at_each_iteration : .true.

PD 12. High-resolution volcanic ash dispersal forecast

PD 12	High-resolution volcanic ash dispersal forecast
Leader	BSC
Participants	BSC, INGV, IMO
Workflow	GFS/WRF-ARW/FALL3D
Numerical Engine	FALL3D
TRL achieved	8-9
Access Mode	By request

Description of the pilot use-case

Volcanoes encompass a range of hazardous phenomena that precede, accompany, and follow volcanic eruptions. Fragmented magma and gases released during explosive eruptions rise up to a neutral buoyancy level where volcanic aerosols and ash can be transported thousands of kilometres by upper-level winds. Specifically, volcanic ash clouds can be a threat to air navigation and flight safety, whereas the subsequent ash fallout can affect buildings (e.g. causing structural damage due to excessive ash loading), communication networks, airports, power plants, and water and energy distribution networks. Many parts of Europe are potentially threatened by this volcanic hazard. Specifically, impact areas include regions close to the volcano (e.g. tephra fallout at Naples with 3 million people at risk from Vesuvius and Campi Flegrei) or even regions over continental scales (e.g. impacts from ash clouds on civil aviation like the massive aviation shutdown occurred during the 2010 Eyjafallajökull eruption in Iceland).

Management of volcanic risk and related strategies for reducing its impacts on aerial navigation can benefit from accurate forecasts of volcanic dispersal produced by volcanic ash transport and dispersion (VATD) models. For example, operational institutions like the Volcanic Ash Advisory Centers (VAACs) rely on VATD models to deliver volcanic ash forecasts to aviation stakeholders, civil protection agencies, and governmental bodies. VATD models aim at simulating the main processes involved in the life cycle of atmospheric ash and gas species released during volcanic eruptions: emission, atmospheric transport, and ground deposition.

Forecasting what will occur in the next hours when a volcano is erupting or to quantify potential impacts from future events are relevant issues to aviation stakeholders and to civil protection agencies and governmental bodies. Volcanic ash cloud forecasts are performed shortly before or during an eruption in order to predict expected fallout rates in the next hours/days and/or to prevent aircraft encounters with volcanic clouds. These forecasts constitute the main decision tool for flight cancellations and aeroplane re-routings avoiding contaminated airspace areas. However, an important gap exists between current operational products and the actual requirements from the aviation industry and related stakeholders. Specifically, this is a list of crucial aspects to be addressed:

1. Temporal and spatial scales of present official forecasts are too coarse
2. Official operational products are still based on semi-quantitative or qualitative forecasts using, for instance, a binary classification strategy (i.e. ash/no ash criterion)

3. Ash dispersal forecasts are prone to epistemic/aleatory uncertainties which are unknown or poorly constrained.
4. VAACs products focus on far-range ash cloud dispersal, with no consideration on proximal fallout potentially affecting airports and landing/take-off operations
5. The emergence of near-real-time geostationary satellite measurements with high spatial and temporal resolutions provides the opportunity to improve the accuracy of operational forecasts. In particular, data assimilation (DA) is one of the most effective ways to reduce forecast errors through the incorporation of observation data into numerical models. However, no DA technique is currently used by operational centres to produce forecast products

PD objectives

HPC is pivotal to address the forementioned critical aspects by: i) making operational ash cloud forecasts compatible with the spatio-temporal constraints of aircraft operations (emergency management scenarios and related urgent computing) and, ii) allowing ensemble-based modelling approaches that reduce and quantify uncertainties and reduce related economic losses.

The idea of PD12 is to consolidate our monitoring and forecasting capacity for volcanic clouds and turn it into an operational service able to produce reliable and realistic forecasts. Specifically, the main objectives of PD12 are:

1. To increase the grid resolution of current operational ash dispersal model setups by one order of magnitude. For the European domain (including Iceland), this implies a 4 km grid resolution in a domain of roughly 4.000 km x 4.000 km
2. Develop a new version of the FALL3D ash dispersal model. The code has been redesigned and rewritten from scratch in modern Fortran Include ensemble modelling capabilities in order to produce probabilistic forecasts
3. To implement an ensemble-based data assimilation system based on the coupling between the FALL3D dispersal model and the PDAF library in order to assimilate high-resolution geostationary satellite retrievals
4. To design a prototypical workflow for implementation of the PD as a potential future service on tier-0 and on tier-1/tier-2 machines (PD version at reduced model resolution), the latter could include e.g. the IMO and INGV clusters.

Workflow structure

A diagram of the modelling system workflow is shown in Figure 12.1. This system is prepared to be run multiple times per day during a volcanic eruption in a semi-automatic way. For instance, if eruptive conditions change, it could be necessary to update model parameters to properly represent the emission source term (e.g. eruption column height).

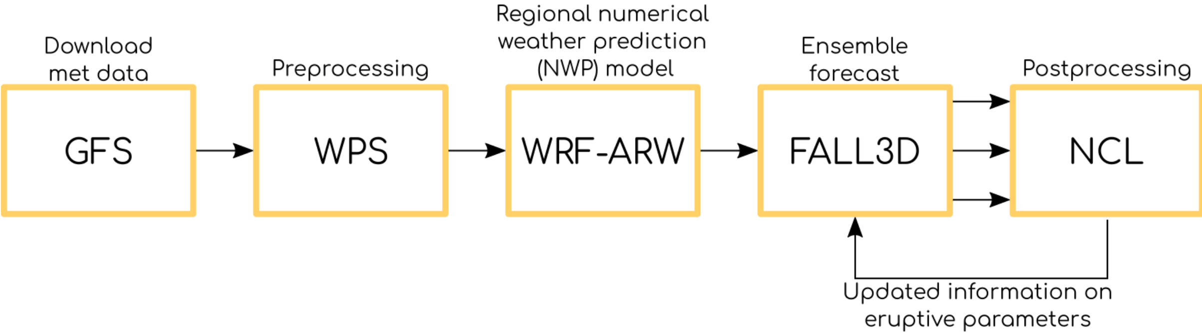


Figure 12.1 Diagram of the modelling system workflow

The components of the workflow are:

1. Downloading meteorological data: In the first step, the most recent global meteorological data must be downloaded from the NCEP (National Centers for Environmental Prediction, US) data server. In particular, we use the NCEP operational Global Forecast System (GFS) forecast. Grids are on a 0.25 by 0.25 global latitude longitude grid and include forecast time steps at a 1 hourly interval from 0 to 240. Model forecast runs occur at 00, 06, 12, and 18 UTC daily.
2. Regional numerical weather prediction model (optional): This step is optional. It could be required for small domains when downscaled meteorological fields with high spatial resolution are required. In this step, a regional Numerical Weather Prediction (NWP) model should be run. Our workflow is based on the WRF (Weather Research and Forecasting) model. Firstly, the global meteorological data is interpolated by the preprocessing module (WPS) and converted into the regional domain. This information can be read by the WRF-ARW model, which produces new meteorological forecasts with downscaled fields.
3. Dispersal model: The FALL3D dispersal model is run next. FALL3D is an open-source offline Eulerian model for atmospheric passive transport and deposition based on the so-called advection–diffusion–sedimentation (ADS) equation. In addition to the meteo driver, the dispersal model requires volcanological inputs (e.g. eruption column height, eruption start time, grain size distribution for tephra, vertical mass distribution).
4. Post-processing: The final products are generated in the last step. A set of scripts written in NCL (NCAR Command Language) are available to produce deterministic and probabilistic products. Examples of some deterministic products are:
   - Ensemble-mean tephra/SO2 column mass load
   - Ensemble-mean tephra/SO2 concentration at relevant flight levels
   - Ensemble-mean tephra deposit thickness
   - Cloud-top height
   - Examples of some probabilistic products are:
   - Column mass exceeding probability
   - Concentration exceeding probability at specific flight levels
   - Fallout deposit exceeding probability

It should be noted that if changes in the eruptive conditions occur and eruptive source parameters have to be updated, only steps (3) and (4) of the workflow should be repeated (see Figure 1).

Required input data

As stated by the following table, two sources of input data are required: the global meteorological data and volcanological information required to define the emission source term parameters.

Input Data	**Format**	**Specific requirements**	**Accessibility**
GFS global forecasts	GRIB-2	Hourly forecasts with spatial resolution of 0.25º	NCEP data server (public repository)
Volcanological inputs	VAA message (simple text file)	Eruption start time, Eruption column height, Grain size distribution	Messages from Volcanic Ash Advisory Centers (VAAC's)

		(e.g. mean, standard deviation, modes)	

Table 12.1. Input data required for High Resolution Ash Dispersal Forecast simulation workflow.

Output data

Ensemble forecasts of volcanic ash and gases are generated by the FALL3D model and stored in a single final output in NetCDF format. Information of all ensemble members is collected in this output file and statistics are computed from the ensemble forecast.

Output Data	Format	Number of files	File size	Level
FALL3D output	NetCDF	1	~100Mb	Level 2

Modality of accessing the codes/workflows

- ☐ Workflow scripts: The workflow is based on BASH/NCL/Python scripts and is accessible to expert users under collaborative actions.
- ☐ GFS global forecasts: NCEP Data Product GFS can be accessed via the NCEP data server at: https://www.nco.ncep.noaa.gov/pmb/products/gfs/
- ☐ WRF-ARW model: The WRF system is available via free download through https://www2.mmm.ucar.edu/wrf/users/
- ☐ FALL3D dispersal model: FALL3D-8.0 and its test suite are available under version 3 of the GNU General Public License (GPL) at https://gitlab.com/fall3d-distribution/v8

  Virtual Access to the numerical engine is also available via the Zenodo ChEESE Community Repository at:
  www.zenodo.org at https://zenodo.org/record/6343786

  Code/Model documentation for FALL3D is available at:
  https://gitlab.com/fall3d-distribution/v8/-/wikis/home

  Further information is available in the following publications:
  https://doi.org/10.5194/gmd-13-1431-2020
  https://doi.org/10.5194/gmd-14-409-2021

- ☐ Post-processing: NCL (NCAR Command Language) is an open source interpreted language, designed specifically for scientific data processing and visualisation. All information on how to download and install NCL on a UNIX-based operating system (Linux, MacOSX and Cygwin/X) can be found on the NCL installation web page: http://www.ncl.ucar.edu/Download/

Scaling of computational resources

Target architecture	Node specs	Target size	Total number of nodes per use-case	Nodes per run	Number of runs	Duration of a single run	Data storage needs	Total CPU hours
Nord III cluster (BSC) Processors: Intel SandyBridge -EP E52670 cores at 2.6 GH	All IBM dx360 M4 node contain: *2x E52670 SandyBridge-EP 2.6GHz cache 20MB 8-core *500GB 7200 rpm SATA II local HDD *Default RAM conguration: 32 GB/node	medium	16 nodes with 256 cores	16	1 (Only FALL3D)	15'	FALL3D 15Gb	74 h (FALL3D)
		large	84 nodes with cores: 512 (WRF-ARW) + 1344 (FALL3D)	84	2 (WRF/ARW and FALL3D)	45' (WRF-ARW) + 40' (FALL3D)	GFS: 6.3G + WPS: 45G + WRF-ARW: 59Gb + FALL3D: 18G	800 h (WRF-ARW) + 733 h (FALL3D)