SPECIAL PROJECT PROGRESS REPORT

All the following mandatory information needs to be provided. The length should *reflect the complexity and duration* of the project.

Reporting year	2025 (2nd year)
Project Title:	Exploring the potential of uncertainty quantification and machine learning techniques to forecast rare extreme events
Computer Project Account:	SPCRDENA
Principal Investigator(s):	Cléa Denamiel
Affiliation:	Ruđer Bošković Institute
Name of ECMWF scientist(s) collaborating to the project (if applicable)	Frédéric Dias (University College Dublin - UCD, Ireland), Serge Guillas (University College London - UCL, UK), Xun Huan (University of Michigan, USA), Anne Mangeney (University Paris Cité, France)
Start date of the project:	01-01-2024
Expected end date:	31-12-2026

Computer resources allocated/used for the current year and the previous one

(if applicable)

Please answer for all project resources

		Previous year		**Current year**	
		Allocated	Used	Allocated	Used
High Performance Computing Facility	(units)	20,000,000	38,871,347	20,000,000	29,797,178
Data storage capacity	(Gbytes)	25,000	10,000	50,000	20,000

Summary of project objectives

The forecast of rare events (e.g., Derechos, meteotsunamis, tsunamis) with traditional ensemble methods (if even possible) is extremely costly in terms of numerical resources and is reaching the limit of what state-of-the-art numerical models can simulate in forecast mode. Consequently, the question of whether or not the cost of (not) forecasting these events is acceptable (i.e., human casualties vs. modelling capabilities and efforts) can be raised. In this project, we propose to explore the potential of using uncertainty quantification (UQ) and machine learning (ML) techniques (e.g., surrogate models/emulators) for the forecast of rare extreme events and to leverage the costs and benefits of such an approach. In particular, different surrogates/emulators will be combined and compared for 3 selected rare extreme events: Derechos in Corsica, acoustically-driven planetary meteotsunami waves and tsunamis caused by landslides in Mayotte.

Summary of problems encountered

The numerical cost of the generation of the surrogate models for acoustically-driven planetary meteotsunami waves has been underestimated in the original proposal and additional resources have been requested in both 2024 & 2025.

Summary of plans for the continuation of the project

The Mayotte landslide-tsunami subtask is now in the last phase of being achieved and an article is in preparation. The acoustically-driven planetary meteotsunami subtask is well-advanced but three more volcanoes will be added. For now, not much has still been done concerning the Corsica Derecho subtask and this will be the main task that will be performed during the rest of the project.

List of publications/reports from the project with complete references

Denamiel, C., Esposti Ongaro, T., and Huan, X.: Silent Threat: Predicting Acoustic Meteotsunami Global Hazards, EGU General Assembly 2025, Vienna, Austria, 27 Apr–2 May 2025, EGU25-7820, https://doi.org/10.5194/egusphere-egu25-7820, 2025.

González del Pino, A., Macías Sánchez, J., Castro Díaz, M., and **Denamiel, C.**: Meteo-HySEA: A GPU accelerated code for simulating atmospherically-driven tsunamis on real bathymetries. Evaluating the performance of the newly implemented nested grids system., EGU General Assembly 2025, Vienna, Austria, 27 Apr–2 May 2025, EGU25-9150, https://doi.org/10.5194/egusphere-egu25-9150, 2025.

Behrens, J., Babeyko, A., Baptista, M. A., **Denamiel, C.**, González Vida, J. M., Hancilar, U., Jalayer, F., Lorito, S., Løvholt, F., Macias, J., Murphy, S., Özer Sözdinler, C., Ragu Ramalingam, N., Romano, F., Rudloff, A., Selva, J., Volpe, M., and Kanoglu, U.: Announcing the Global Tsunami Model Association, EGU General Assembly 2025, Vienna, Austria, 27 Apr–2 May 2025, EGU25-5074, https://doi.org/10.5194/egusphere-egu25-5074, 2025.

Gibbons, S. J., Bader, M., **Denamiel, C. L.**, Díaz, M. J. C., Gabriel, A.-A., González del Pino, A., Lorito, S., Macías Sánchez, J., Romano, F., Storrøsten, E. B., Ulrich, T., Wille, M., and Løvholt, F.: A workflow for Complex Multi-Source Tsunami Modelling, EGU General Assembly 2025, Vienna, Austria, 27 Apr–2 May 2025, EGU25-11615, https://doi.org/10.5194/egusphere-egu25-11615, 2025.

Summary of results

During this 2nd year period, the main tasks performed have been to build the surrogate models for both Mayotte landslide tsunamis and acoustically-driven planetary meteotsunamis.

1. Landslide tsunamis in Mayotte

Mayotte is a volcanically active territory located in the northern Mozambique Channel of the Indian Ocean. This archipelago, composed of the main islands of Grande Terre and Petite Terre, is characterized by a shallow submarine shelf that encloses a large lagoon (Fig. 1). Since the onset of an intense seismo-volcanic crisis in 2018 (Mercury et al., 2022), geophysical investigations have revealed that the flanks of the newly formed submarine edifice east of Mayotte—the so-called "Piton" area—are prone to gravitational collapse (Feuillet et al., 2021; Roger, 2019; Poulain et al., 2022; Marboeuf et al., 2025). Such collapses could trigger submarine landslide-generated tsunamis with devastating consequences for the coastal population. Notably, landslide-induced tsunamis differ from their seismic counterparts by their extremely short arrival times, potentially larger initial wave heights near the source, and the absence of clear precursors detectable by traditional Early Warning Systems (EWS; Roger et al., 2024; Lemoine et al., 2020b). This combination of rapid onset and high impact underscores the need for efficient hazard assessment tools capable of supporting local preparedness and emergency planning. In this project, a surrogate modelling framework specifically developed for landslide-generated tsunami scenarios within Probabilistic Tsunami Hazard Assessment (PTHA) workflows: the Landslide-Tsurrogate v1.0 model. The framework is designed to integrate sparse input sampling to approximate complex numerical tsunami models with minimal computational overhead and is used in the Mayotte region, where emerging submarine landslide hazards demand rapid, probabilistically robust tsunami hazard assessments.

Figure 1. Mayotte lagoon geomorphology including Grande Terre, Petite Terre and Piton locations.

The construction of surrogate models in Mayotte with Landslide-Tsurrogate v1.0 follows seven building blocks or steps, as illustrated in Figure 2.

The first step is the user specifications which primarily involves selecting the stochastic variables that characterize submarine landslides. For the Mayotte, 3 stochastic variables are used: location, volume (V) and angle of friction (α) of the landslides. To minimize the number of stochastic variables, landslide locations are defined using only the latitude (y) values along isolines of the steepest slope in

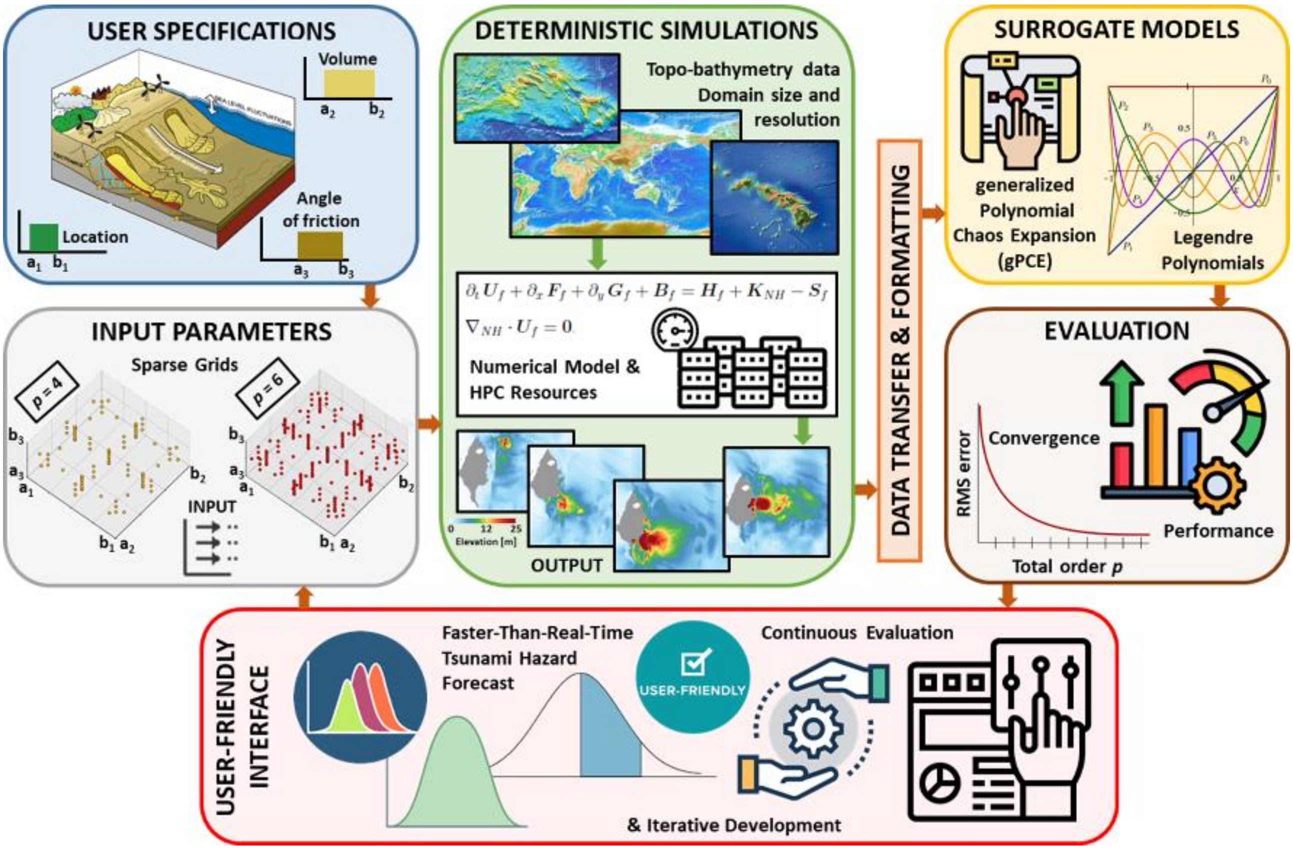


Figure 2. Overview of the seven Landslide-Tsurrogate v1.0 building blocks enabling faster-than-real-time Probabilistic Tsunami Hazard Assessments (PTHA) based on surrogate models. These building blocks are: (1) User Specifications, defining the stochastic variables; (2) Input Parameters, providing the inputs for deterministic simulations; (3) Deterministic Simulations, using complex numerical models; (4) Data Transfer and Formatting; (5) Surrogate Models, creating with the deterministic simulations; (6) Evaluation, assessing surrogate model convergence and performance; and (7) User-Friendly Interface, generating the PTHA outputs based on user inputs.

the Piton area, where landslides are considered most likely to occur (Fig. 3). The longitude coordinate is obtained by interpolating along the selected isolines of highest slope. Surrogate models have been constructed along three distinct isolines of steepest slope: Piton, Piton South, and Mayotte South (Fig. 3). Hereafter we will only present the results for the Piton area. Given the sparsity of the available measurements, the prior distribution of the stochastic variables could not be assessed accurately. It is thus assumed that all the stochastic variables follow uniform distributions (i.e., within the range of the uniform distributions, the generation of any pressure disturbance is equally probable) given by:

$$\begin{cases} y(\omega) \sim U\left([8588254.0, 8589736.0]\right) \text{ m} \\ V(\omega) \sim U\left([1, 200]\right) \text{ Mm}^3 \\ \alpha(\omega) \sim U\left([4.0, 12.0]\right) \, ^\circ \end{cases}$$

The second step is the generation of the input parameters which is critical in constructing the surrogate models within the Landslide-Tsurrogate v1.0 framework. It involves generating the input parameter sets for the deterministic simulations. The input parameters correspond to the 3 stochastic variables: location, volume and angle of friction. For Mayotte, a sparse delayed Gauss–Patterson quadrature is constructed up to a total order 6. The quadrature nodes—representing collocation points in the stochastic parameter space—are mapped from the canonical space to the physical space to ensure that deterministic simulations are concentrated in regions of high probability density, thereby improving the efficiency and accuracy of the input space sampling. Each combination of parameters defines a deterministic landslide scenario that is then simulated using a numerical tsunami model. For the Piton area, the final number of deterministic simulations is 207.

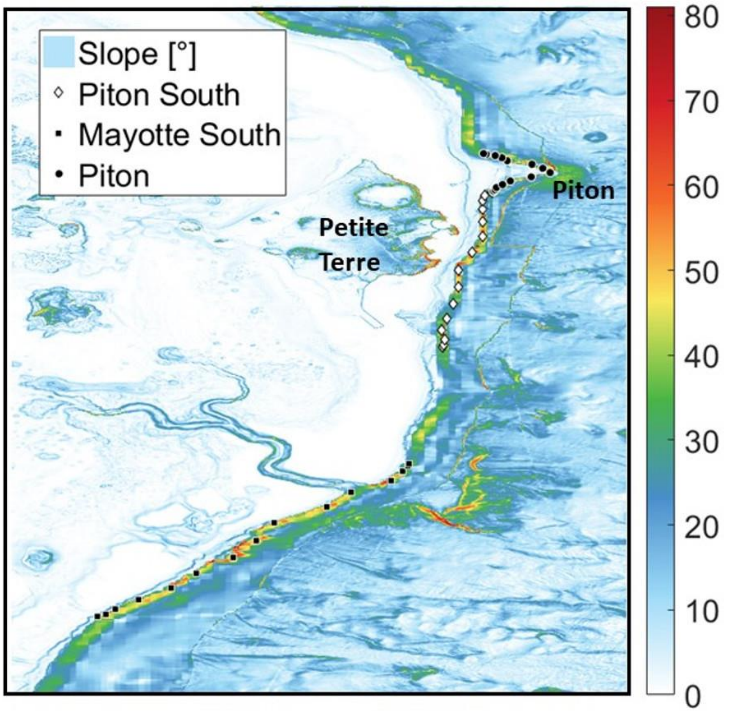


Figure 3. Landslide locations

This deterministic simulation step primarily involves selecting and running a numerical model to simulate the landslide-generated tsunamis based on the input parameters generated in the previous step. For Mayotte, the HySEA (Hyperbolic Systems and Efficient Algorithms) hydrodynamic model, originally developed for Earthquake and landslide tsunamis is used. The HySEA software consists of a family of geophysical codes based on either single-layer, two-layer stratified systems, or multilayer shallow-water models. The CUDA-based HySEA codes are developed by the EDANYA group from UMA (the University of Málaga).

The Data Transfer and Formatting step deals with the large volume of data produced by the deterministic runs. In the Piton area, the data is formatted in a file containing the maximum elevation, maximum speed, and time of arrival of the 207 tsunami waves extracted at the selected 863 locations where the surrogate models are to be constructed (Fig. 4).

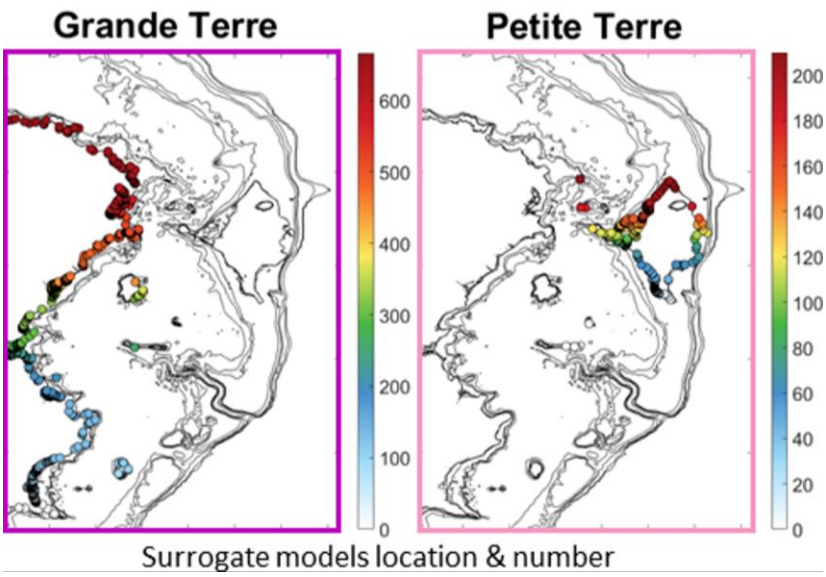


Figure 4. Surrogate model locations (and number) along the Grande Terre and Petite Terre coastlines for the Mayotte test case.

The Surrogate Model step is the main building block of Landslide-Tsurrogate v1.0. Here, the deterministic coefficients of the surrogate models are computed based on the selected quadrature rule and the faster-than-real-time probabilistic tsunami hazard assessments are provided at the chosen 863 locations.

The evaluation step is a key component of Landslide-Tsurrogate v1.0, aimed at assessing the convergence (necessary to define the truncation order of the surrogate models), accuracy, and sensitivity (needed to exclude certain stochastic variables) of the surrogate models. A key strength of the framework used in Landslide-Tsurrogate v1.0 lies in its ability to assess convergence through hierarchical polynomial orders. By comparing the surrogate model outputs at a given total order p with those at the next order p+1, especially when using the Gauss–Patterson nested quadrature rules, the stability and accuracy of the expansion can be quantitatively

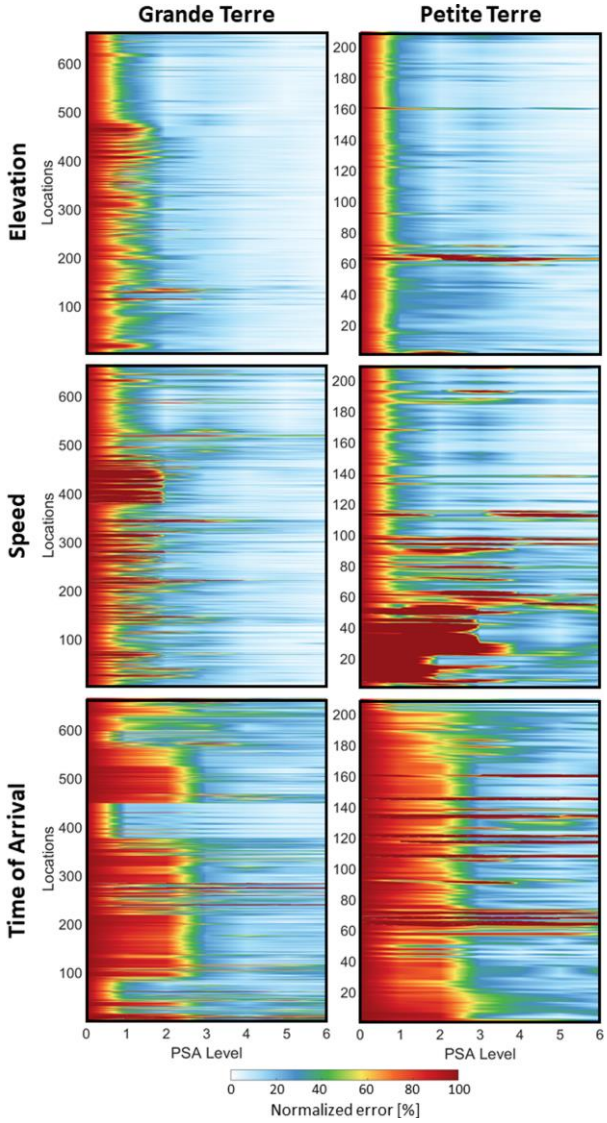
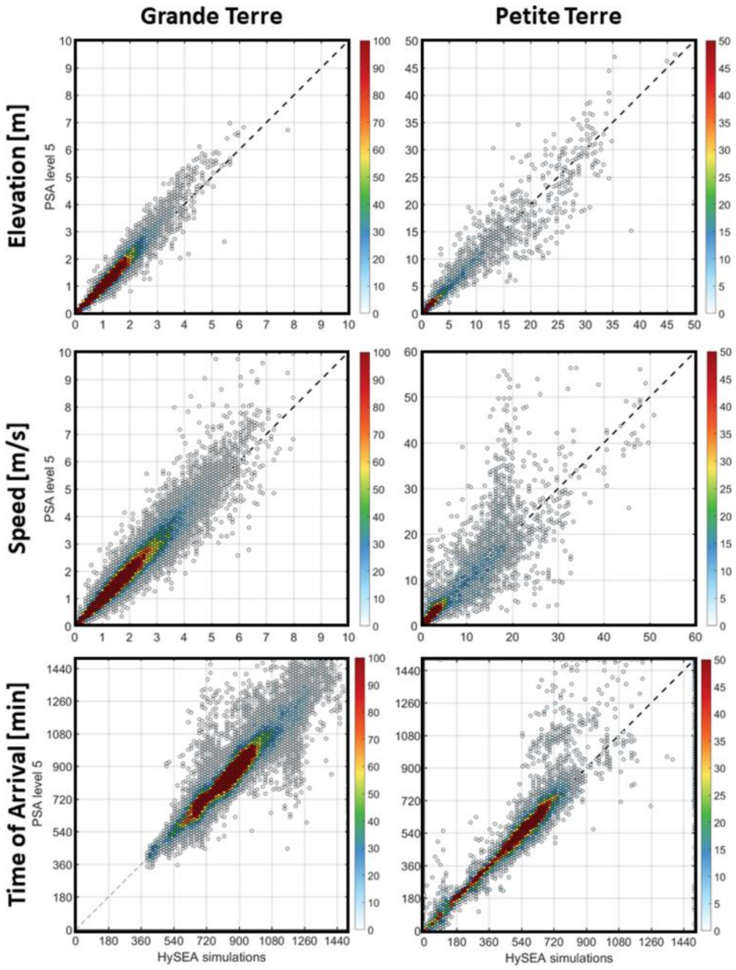


Figure 5. Convergence and accuracy of the surrogate models along the coastlines of Petite Terre and Grande Terre.



evaluated. Practically, for the Piton area, the convergence (Fig. 5, left panels) show that the error is minimized at a total order 5. This means that the surrogate models can be built at the total order 5 and the remaining simulations, needed to reach the total order 6, can be used to evaluate the surrogate model results with scatter plots (Fig. 5 right panels). Following these results, the accuracy of the surrogate models is judged satisfactory.

The last step is to develop a user-friendly interface to facilitate the practical use of surrogate models by a wide range of end users—including scientists, engineers, and decision-makers. The interface for the Mayotte landslide-tsunami surrogate models is under development and will consist in two different GUIs: a Matlab interface and a Jupyter Widget-based web-application.

2. Acoustically-driven planetary meteotsunami waves

This part of the project deals with representing the acoustic-gravity wave-driven ocean dynamics resulting from explosive volcanic eruptions. The final aim is to integrate this new global hazard within existing tsunami early warning systems. However, we face two major limitations. However, two major limitations are encounterd. First, the vast majority of the scientific knowledge about acoustically-driven meteotsunami events was derived from observations and modelling experiments performed after the Hunga Tonga-Hunga Ha'apai (HTHH) volcanic eruption. Therefore, the explosive-atmosphere-ocean dynamics might not be as well understood as it seems from the extensive literature already available. Second, due to the destructive power of the blasts at the epicenter and the potential generation of tsunamis from other sources, acoustic meteotsunamis will often be a secondary or even tertiary hazard in the aftermath of catastrophic volcanic explosions or asteroid impacts. Consequently, the cost of assessing acoustic meteotsunami hazards in early warning systems should be minimal. To address these limitations, surrogate models based on Uncertainty Quantification (UQ) methods (Ghanem et al., 2017; Le Maître and Knio, 2010) are built. As seen in Figure 6, these surrogate models propagate the uncertainties of the volcanic explosion to the worldwide meteotsunami surge results and are designed to optimize the balance between lead-time and robustness/stability of the real-time hazard assessments.

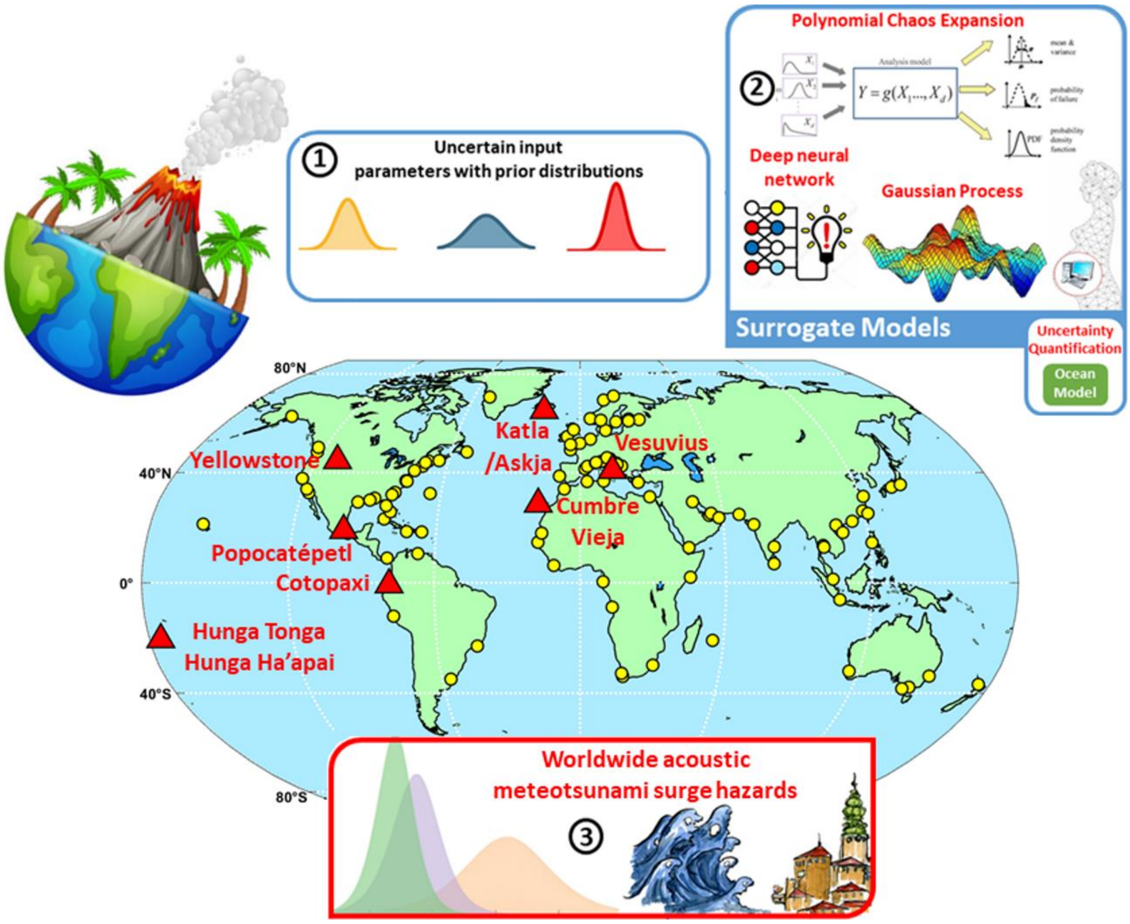


Figure 6. Strategic framework. Step 1. Defining the uncertain input variables of the atmospheric disturbances generated by volcanic eruptions and their associated prior distributions. Step 2. Implementing surrogate models based on uncertainty quantification and high-fidelity ocean models. Step 3. Assessing acoustic-gravity wave-driven ocean dynamics and acoustic meteotsunami surge global hazards. Artwork from Frits Ahlefeldt: https://fritsahlefeldt.com.

In this part of the project, the surrogate models are implemented for 7 different volcanoes and for each of the most populated and/or endangered coastal cities in the world. As volcanic eruptions occur at the geological scale, these models are built through the numerical reproduction of all potential events with thousands of high-fidelity simulations forced by a basic synthetic atmospheric disturbance defined as follows:

$P(x,y,t,\omega) = P_A(\omega) cos\left(\frac{2\pi}{\lambda(\omega)}(r - c(\omega)t) + \pi\right)$ if $0 \leq (r - c(\omega)t) + \frac{\lambda(\omega)}{2} \leq \lambda(\omega)$

$P(x,y,t,\omega) = 0$ otherwise

Here, the 3 stochastic variables depending on (ω) and defining the pressure disturbance are: $P_A(\omega)$ the amplitude, $\lambda(\omega)$ the wavelength and $c(\omega)$ the propagation speed. The range of variation of these 3 stochastic input parameters has been defined based on pressure measurements and numerical simulations as follows: $P_A(\omega) \sim U([4\,hPa, 4200\,hPa])$, $\lambda(\omega) \sim U([300\,km, 900\,km])$ and $c(\omega) \sim U([200\,m/s, 350\,m/s])$.

In total 641 simulations are performed per volcano with 177 simulations representing a Volcanic Explosivity Index (VEI) of 5 & 6 corresponding to a volume ejecta of 5 to 100 km3 and a pressure amplitude of 10 to 200 hPa.

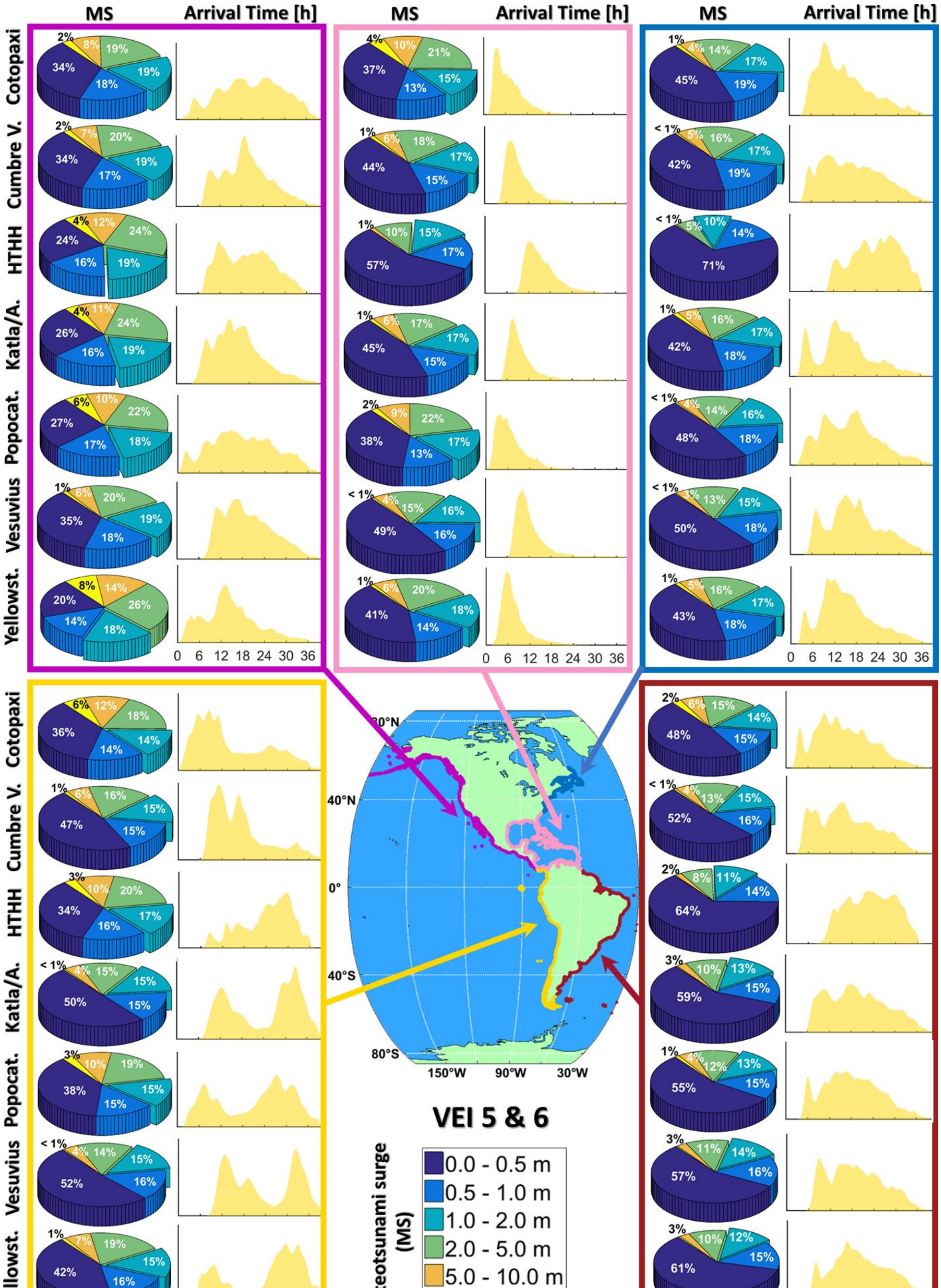


Figure 7. Acoustic meteotsunami probabilistic hazard assessments for VEI 5 & 6. Distribution of the meteotsunami surges (MS; as pie charts) and their associate time of arrival (as density plot) along the coasts of north-west (purple), north-east (blue), south-west (yellow), south-east (brown) and central-east (pink) America.

Before building the surrogate models, the results of the simulations for VEI 5 & 6 are analysed along the worldwide coastline. An example along the American coastlines is shown in Figure 7.

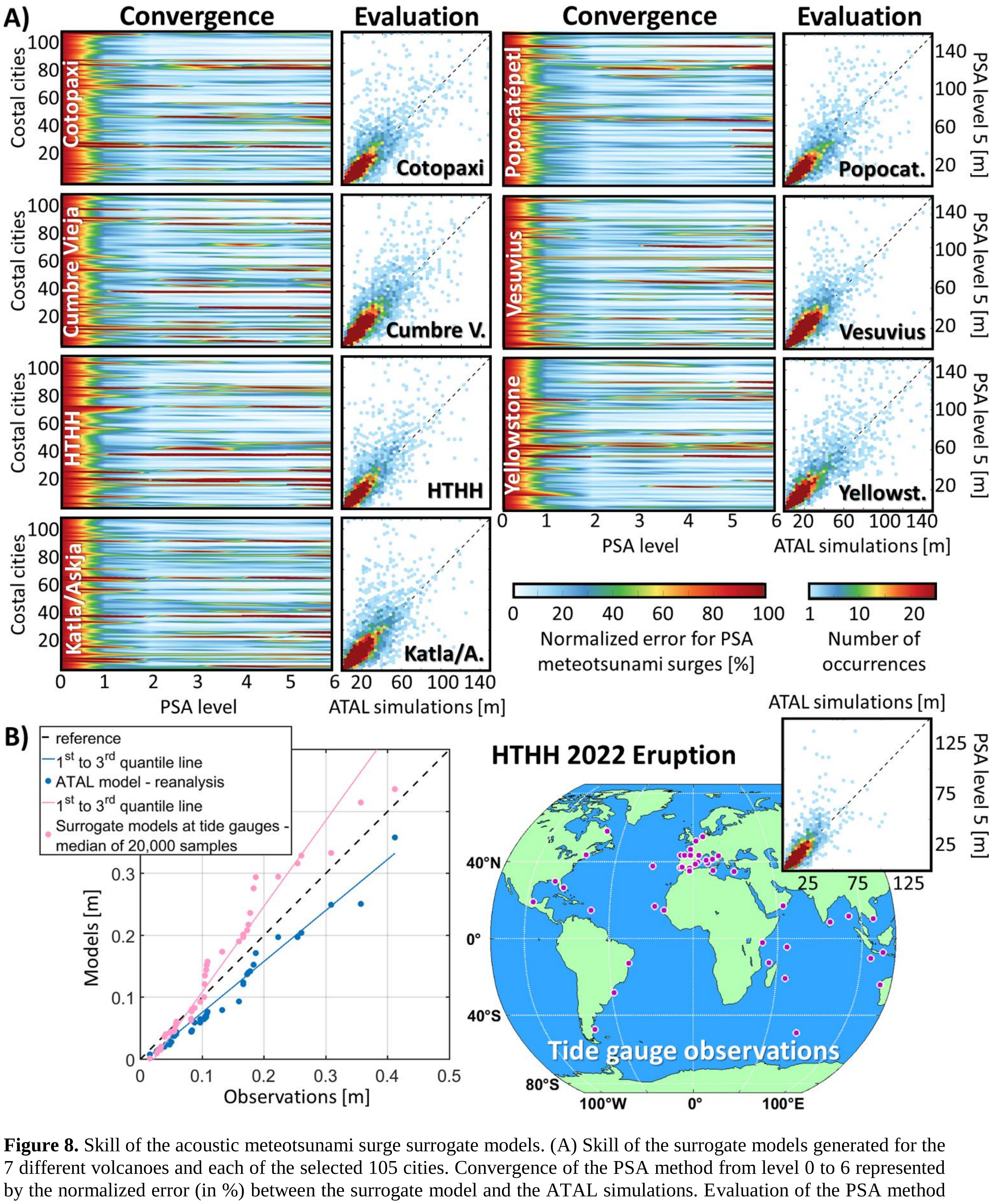


Figure 8. Skill of the acoustic meteotsunami surge surrogate models. (A) Skill of the surrogate models generated for the 7 different volcanoes and each of the selected 105 cities. Convergence of the PSA method from level 0 to 6 represented by the normalized error (in %) between the surrogate model and the ATAL simulations. Evaluation of the PSA method level 5 against the 256 independent simulations generated to reach the level 6 of the PSA method presented as scatter plots of the results obtained for the entire set of cities. (B) Skill of the HTHH surrogate models during 2022 volcano eruption. Evaluation of the PSA method level 5 against observations as well as ATAL 256 independent and reanalysis simulations at 43 tide gauge locations.

Then the surrogate models are constructed following the same steps defined for the Mayotte landslide-tsunami task and the evaluation is presented for 105 cities around the world in Figure 8. It should also be noted that for the HTHH volcano surrogate models have been built at tide gauges' locations only affected by the meteotsunami generated by the 2022 eruption in order to compare the results of the surrogate models with the observations. Overall, the skills of the surrogate models in assessing the acoustically-driven meteotsunami hazards have been demonstrated.

In conclusion, the project is for the moment progressing at the expected pace with several positive outcomes ready to be published in the coming year.