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	2024 (1 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	1,354,379
Data storage capacity	(Gbytes)	/	/	25,000	1,000

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 only problem encountered, till now, is that the GPU cost of running HySEA for landslide-tsunamis has been underestimated and need for more GPUs has been raised during these first 6 months of the project. Due to the limited availability of GPU resources proposed by ECMWF, we are currently exploring the use of other GPU sources to be able to create the surrogate model(s) in Mayotte.

Summary of plans for the continuation of the project

During the rest of the project, the numerical simulations needed to build the surrogates/emulators will be run with Landslide-HySEA for the tsunamis in Mayotte, with the CoriSC model for Derechos over Corsica and with the ATAL (and TIGAR) model for acoustically-driven planetary meteotsunami events.

List of publications/reports from the project with complete references

No publication is yet associated with the project.

Summary of results

During the first 6 months of the project the main tasks have been to start defining the different modelling framework and/or surrogate models for the 3 applications (landslide tsunamis in Mayotte, Derechos in Corsica and acoustically-driven planetary meteotsunami waves).

1. Landslide tsunamis in Mayotte

As the characteristics of the landslides highly depend on the geographical location and the depth where they are triggered, three different surrogates were already designed to properly capture and distinguish the different kind of landslides that could occur around the Piton area in Mayotte (Figures 1). For the moment we generated two nearshore surrogates, defined along the 150m depth isoline, on the top and bottom of Piton and one offshore surrogate from the 300m depth isoline (Figure 2). For all the surrogates in Mayotte, the landslides are approximated as an ellipsoid of additional material put on top of the bathymetry and a maximum of 5 stochastic variables are used: latitude where the landslide occurs along the chosen isoline (LAT), eastwards distance from the chosen isoline (X in m), semi-major axis of the ellipse defining the landslide (R in m), maximum height of the landslide (H in m) and angle of friction (A in °). Given the sparsity of available measurements and the rarity of strong landslide events, it is assumed that all stochastic variables follow a uniform distribution as follows.

Top nearshore Piton surrogate: $LAT(\omega) \sim U([8589000, 8594000])$, $R(\omega) \sim U([100\text{m}, 600\text{m}])$, $\text{H}(\omega) \sim U([10m, 100\text{m}])$, $A(\omega) \sim U([4°, 12°])$

Bottom nearshore Piton surrogate: $LAT(\omega) \sim U([8584000{,}8589000])$, $R(\omega) \sim U([100\text{m}, 600\text{m}])$, $\text{H}(\omega) \sim U([10m, 100\text{m}])$, $A(\omega) \sim U([4°, 12°])$

Offshore Piton surrogate: $LAT(\omega) \sim U([8588000, 8590000])$, $X(\omega) \sim U([0m, 300m])$, $R(\omega) \sim U([100m, 600m])$, $H(\omega) \sim U([50m, 250m])$, $A(\omega) \sim U([4°, 12°])$

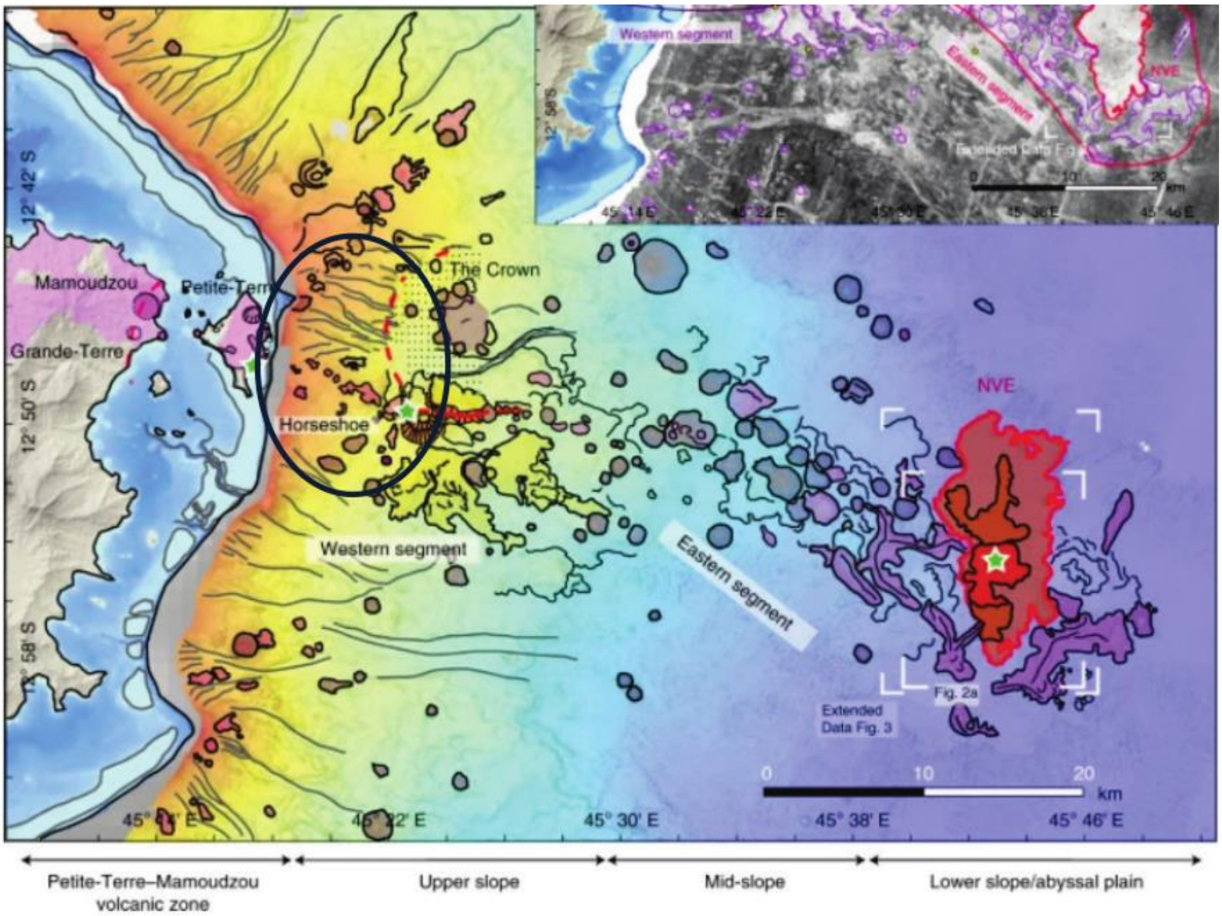

Figure 1. Schematic representation of the areas where landslides could occur in Mayotte with the Piton area circled in black.

The mathematical framework of the surrogate is based on polynomial expansions that decompose into deterministic coefficients and random orthogonal bases. The coefficients, which are the projection of the tsunami surges (maximum elevation at the coast) onto each polynomial basis, are derived from a quadrature based approximation using numerical simulations.
These simulations are undertaken with the HySEA (Hyperbolic Systems and Efficient Algorithms) hydrodynamic model, originally developed for Earthquake and landslide tsunamis. 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).
A nonintrusive quadrature-based Pseudo-Spectral Approximation (PSA) of the tsunami surge generalized Polynomial Chaos Expansions (gPCE) is used. The number and values of the landslide conditions forcing the deterministic HySEA model only depend on the total order p of the polynomial decomposition and the chosen quadrature rule. Hereafter, the tsunami surrogate model is built using the delayed Gauss–Patterson nested rules which keep a high precision but dramatically reduce the number of stochastic variable values needed for the quadrature compared to other rules.

For the moment, the PSA method for landslide tsunamis in Mayotte is tested for a total order p=5 of the polynomial decomposition for the 3 surrogates (Figure 2).

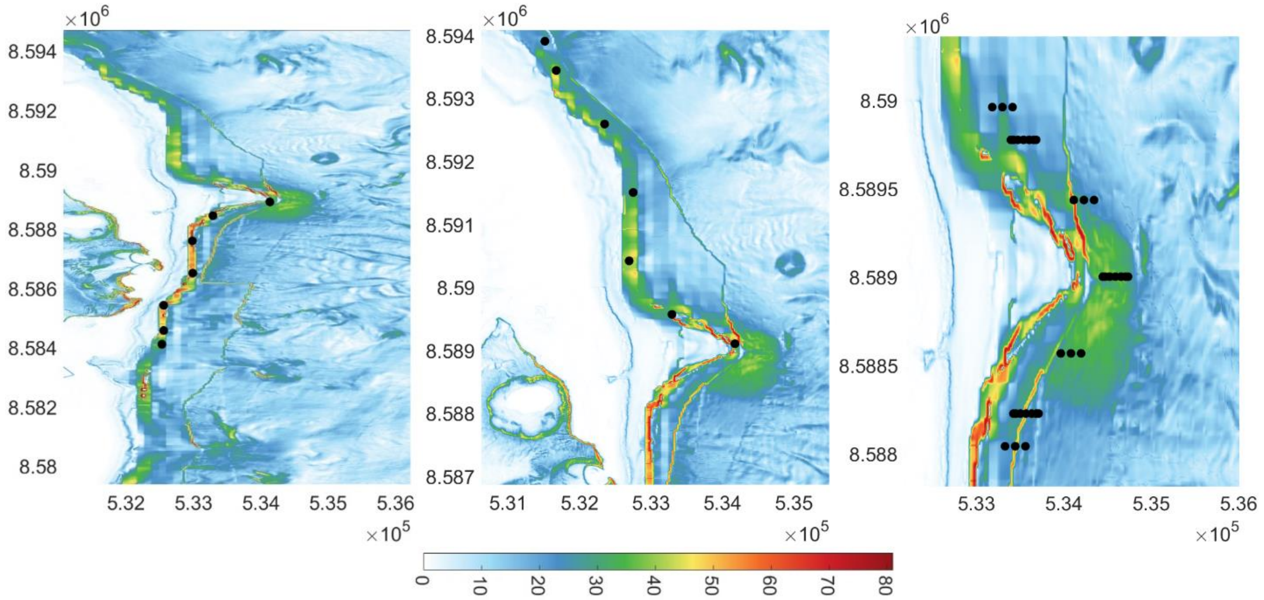

Figure 2. Locations of both the nearshore (two left panels) and offshore (right panel) landslides for a polynomial order 5 (black dots) on top of the slope (coloured maps).

2. Derechos in Corsica

The 18th of August 2022 severe storms occurred in a swath from Menorca (Balearic Islands, Spain) through Corsica (France), northern Italy, Slovenia, Austria, and southern Czechia. In total, 12 people died and 106 people were injured by wind and hail (up to 11 cm of diameter). All fatalities, and most of the injuries, were caused by a long-lived convective system, also known as a Derecho, that produced extremely severe wind gusts (up to 62.2 m/s recorded in Corsica) and rapidly moving showers. In Corsica, this Derecho resulted in the death of 5 people.

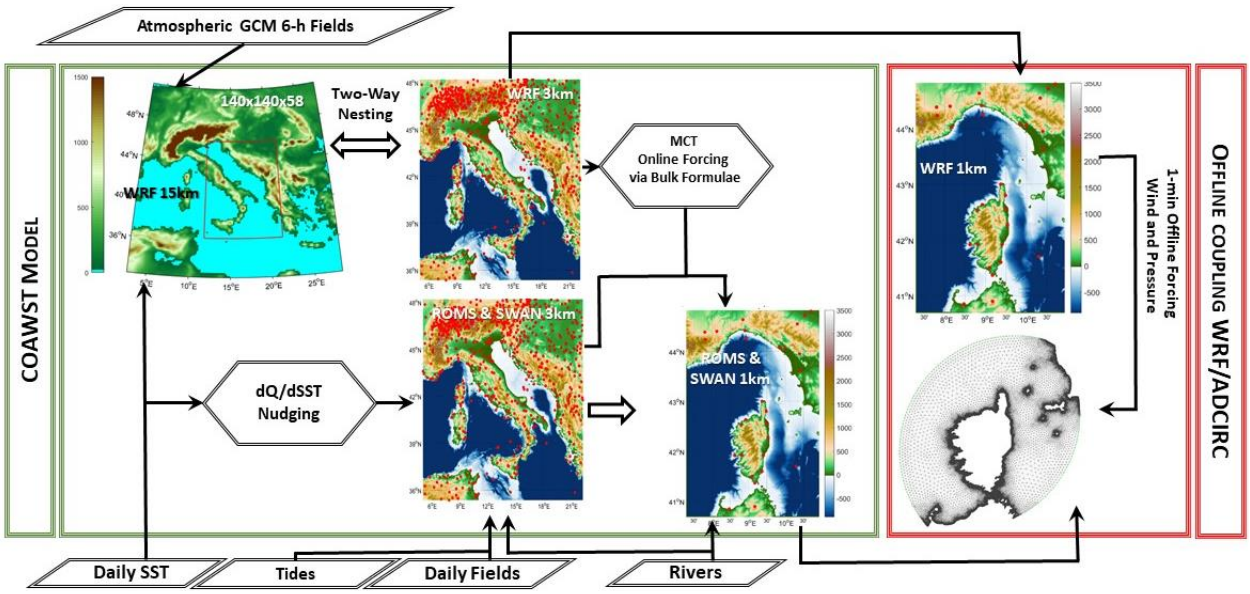

Figure 3. Setup of the Corsica island Sea and Coast (CoriSC) modelling suite.

The Corsica island Sea and Coast (CoriSC) modelling suite replicates the configuration used in the Adriatic Sea (AdriSC model) to forecast meteotsunami events (Figure 3). At this stage of the project the CoriSC model is not optimized in terms of the speed and cost of the runs but has been improved in comparison with the AdriSC model by using the EFAS model for the river inputs instead of monthly climatologies and by imposing the sea surface temperature from the ROMS 1-km to the WRF 1-km grid.

For the moment the main work done during the project was to evaluate the skills of the CoriSC model to reproduce the observations over Corsica during the 18th of August 2022 Derecho. For this, 4 simulations were performed starting from the 14th, 15th, 16th and 17th of August and forced by either reanalysis products (ERA5) or ECMWF forecasts (Figures 4 and 5).

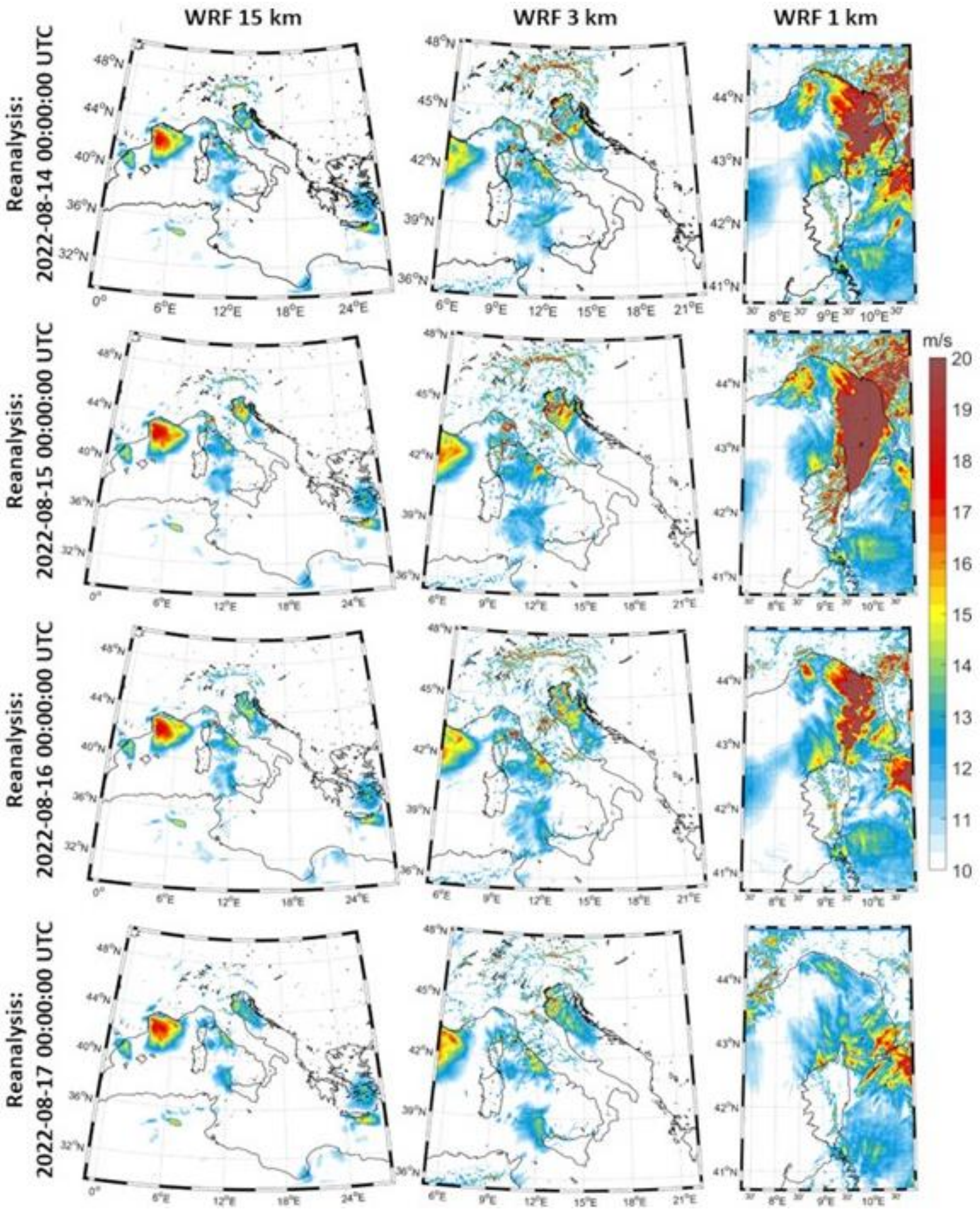

Figure 4. Setup of the Corsica island Sea and Coast (CoriSC) modelling suite.

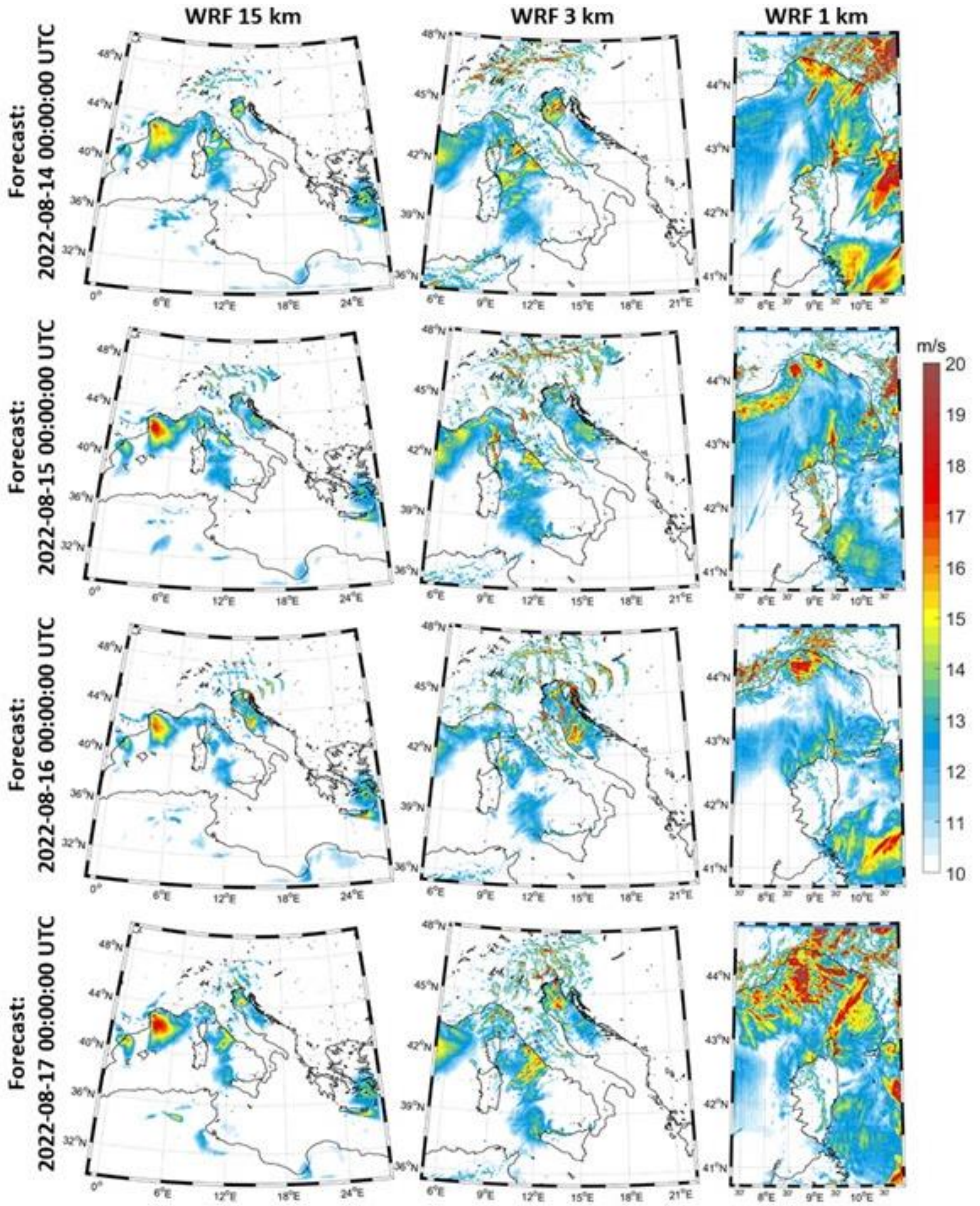

Figure 5. Setup of the Corsica island Sea and Coast (CoriSC) modelling suite.

The evaluation of the 8 simulations is made only during the 18th of August 2022 when the maximum of the rain and wind speed is extracted and compared to the observations over Corsica (Figure 6). Overall the CoriSC model struggles to reproduce the event except for the reanalysis simulation starting the 15th of August 2022. In this simulation the CoriSC model maximum wind speed reach more than 25 m/s in the northern Corsica which is in good agreement with the observations. However, the extreme rain modelled by CoriSC is also located in northern Corsica while the observations show that it occurred along the western coast. It should also be noted that the modelled extreme winds and rainfall occur 3 hours earlier than the observed ones. All these difficulties to reproduce the event demonstrate the necessity to use ensemble modelling which can be achieve at lower cost by building a surrogate model accounting for the uncertainty of the deterministic solution.

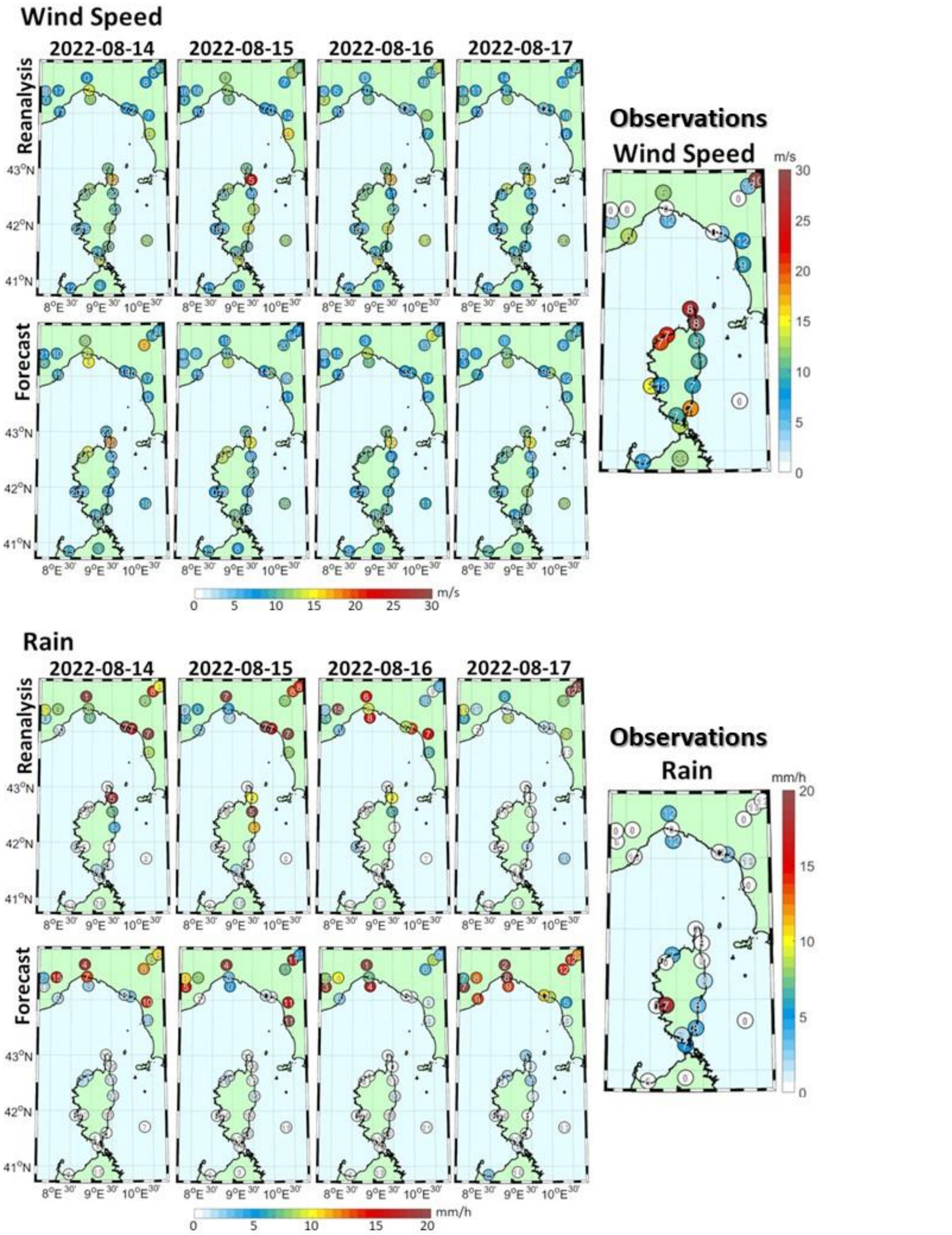

Figure 6. Evaluation of the CoriSC model against the maximum wind speed and rain measurements observed over Corsica. The numbers inside the circles represent the hour when the maximum occur.

The way to build surrogate models for Derecho in Corsica has still yet to be investigated but several avenues are explored in order to optimize the numerical cost of building such an emulator.

3. Acoustically-driven planetary meteotsunami waves

Research on acoustic meteotsunamis only gained traction after the Hunga Tonga–Hunga Ha'apai (HTHH) volcano eruption on January 15, 2022, when such meteotsunami waves were globally recorded by hundreds of tide gauges. Consequently, as little is still known about the full explosion-atmosphere-ocean dynamics during these events, in this project we decided to first start building a simple surrogate model for the HTHH volcano – $(x_0 = -175.38\,°\mathrm{E}, y_0 = -20.54\,°\mathrm{N})$ coordinates in the in the (x, y) referential – based on 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 $c(\omega) \sim U([200\,m/s, 350\,m/s])$. An example of the propagation of such atmospheric disturbance is presented in Figure 7.

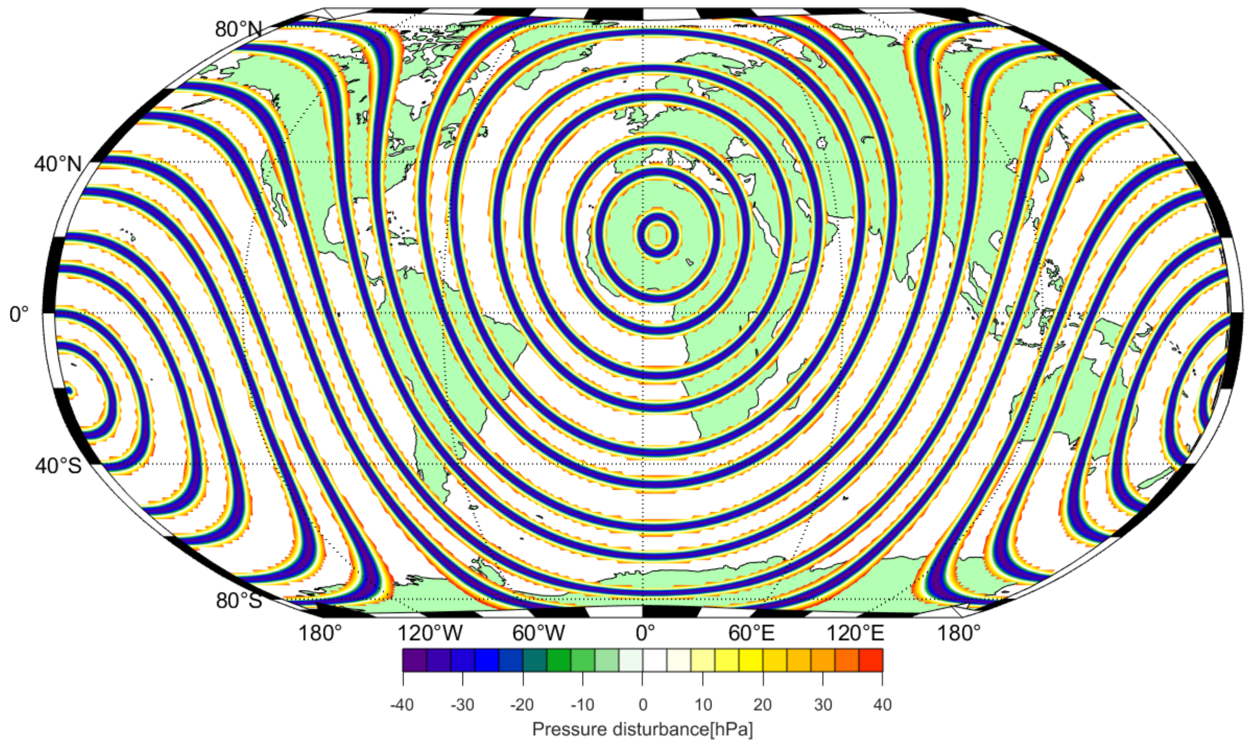

Figure 7. Synthetic pressure disturbance generated by a HTHH explosion with $P_A = 40\,hPa$, $\lambda = 600\,km$ and $c = 315\ \ m/s$ and plotted every 2 hours.

Identically to the landslide-tsunamis in Mayotte, a nonintrusive quadrature-based Pseudo-Spectral Approximation (PSA; Gerstner and Griebel, 1998; Constantine et al., 2012) with an order p=5 is used to build this simplified surrogate of planetary meteotsunami waves due to HTHH potential explosions. For this simplified surrogate only the Atmospheric Tsunamis Associated with Lamb waves (ATAL) model is used.