Landslide-Tsurrogate v1.0: A computationally efficient framework for probabilistic tsunami hazard assessment applied to Mayotte (France)

Cléa Denamiel[1,2,*], Alexis Marboeuf[3,*], Anne Mangeney[3], Anne Le Friant[3], Marc Peruzzetto[4], Antoine Lucas[3], Manuel J. Castro Díaz[5], and Enrique Fernández-Nieto[6]

[1]Ruđer Bošković Institute, Division for Marine and Environmental Research, Bijenička cesta 54, 10000 Zagreb, Croatia
[2]Institute for Adriatic Crops and Karst Reclamation, Put Duilova 11, 21000 Split, Croatia
[3]University Paris Cité, Institut de Physique du Globe de Paris, 1, rue Jussieu, 75238 Paris cedex 05, France
[4]BRGM, F-75012 Paris , France
[5]Departamento Análisis Matemático, Estadística e Investigación Operativa, y Matemática Aplicada, Universidad de Málaga, Campus Teatinos S/N, Málaga 29080, Spain
[6]Departamento de Matemàtica Aplicada I, Universidad de Sevilla, E.T.S. Arquitectura. Avda Reina Mercedes, 41012 Sevilla, Spain
[*]Equal contributions

Correspondence: Cléa Denamiel (cdenami@irb.hr) and Alexis Marboeuf (marboeuf@ipgp.fr)

Abstract. Landslide-Tsurrogate v1.0 is an open-source Python and MATLAB tool that helps scientists quickly estimate the tsunami hazards generated by submarine landslides. Instead of running thousands of heavy deterministic numerical simulations, the software builds surrogate models that reproduce the main results with a fraction of the computational cost. The method relies on a mathematical approach called generalized polynomial chaos expansion, which efficiently explores how uncertain landslide parameters affect tsunami generation. Users can perform sensitivity analyses, identify the most influential parameters, and quantify the variability of possible outcomes. The tool includes a Jupyter Notebook User Manual and interactive MATLAB and Jupyter Notebook interfaces, making it easy to understand the methodology, set up the surrogate simulations and visualize the results. The Landslide-Tsurrogate v1.0 model's performance is demonstrated through a real-world test case involving five zones in Mayotte (France). For this application, the surrogate models achieve convergence with only 135 deterministic simulations per zone and produce probabilistic results in less than 2 seconds within the user-friendly interfaces used on a basic laptop, demonstrating the computational efficiency of the approach. Beyond this example, the framework can be applied to any coastal region prone to submarine landslides. By combining physical modeling, statistical analysis, and user-friendly design, Landslide-Tsurrogate v1.0 enables faster and more transparent probabilistic tsunami hazard assessments.

1 Introduction

In the last decade, surrogate model approaches—providing fast-running approximations of complex dynamical models—have been gaining significant traction in the geoscience community, particularly within the Destination Earth framework which promotes the creation of Digital Twins. For example, Formaggia et al. (2013) built a surrogate model of basin-scale geochemical compaction, Wang et al. (2016) studied acoustic uncertainty predictions, Sraj et al. (2014) estimated the wind drag parameter forcing an ocean model, Giraldi et al. (2017) documented the propagation of earthquake ocean floor displacement uncertainty to tsunami wave parameters and Bulthuis et al. (2019) used a surrogate model to quantify the uncertainty of the multi-centennial response of the Antarctic ice sheet to climate change. Compared to traditional sampling techniques such as Monte Carlo simulations, surrogate models are markedly more efficient for uncertainty quantification. Once constructed for a particular application or geographic domain, their use requires very limited further computational resources (Knio and Le Maître, 2006; Najm et al., 2009). Furthermore, surrogate models are inherently flexible: they can be constructed based on sparse or incomplete prior knowledge and tailored to situations where observational and computational resources are limited. For instance, uniform prior distributions can be adopted in the absence of comprehensive observational data, and sparse grids can be used to reduce the number of model evaluations required during training.

Given these advantages, surrogate models are particularly well-suited for Probabilistic Tsunami Forecast (PTF) in regions where observational data are scarce and computational capacities are limited. Unlike traditional high-fidelity simulations, which can be computationally intensive, surrogate models offer a more efficient framework to approximate tsunami dynamics across a wide range of potential source scenarios. These models enable rapid propagation of uncertainties associated with various tsunami generation mechanisms—be it seismic, volcanic, landslide, or atmospheric—to estimate the probability of coastal hazard metrics exceeding critical thresholds pertinent to infrastructure planning, emergency response, and long-term risk mitigation.

Recent studies have demonstrated the efficacy of surrogate modeling in diverse probabilistic tsunami modeling contexts. For instance, Denamiel et al. (2019, 2020, 2021) developed a surrogate model approach to quantify meteotsunami hazards in the Adriatic Sea, addressing challenges posed by limited observational data and the coarse resolution of atmospheric models. In the Cascadia Subduction Zone, Guillas et al. (2018) introduced a functional emulator to approximate high-resolution tsunami simulations, facilitating efficient uncertainty quantification in tsunami hazard assessments for the Pacific Northwest. Additionally, Salmanidou et al. (2021) constructed statistical emulators, enabling high-resolution probabilistic tsunami predictions in northern Cascadia. The integration of machine learning techniques into surrogate modeling has further expanded its applicability and efficiency. For example, Brecht et al. (2025) utilized physics-informed neural networks to model tsunami inundation, offering a meshless approach that simplifies the modeling process while maintaining accuracy. Such advancements highlight the growing potential of surrogate models to serve as foundational tools for tsunami hazard assessment, especially in contexts where data limitations and computational constraints hinder the application of fully deterministic modeling strategies. Collectively, these examples underscore the versatility and effectiveness of surrogate models in probabilistic tsunami modeling,

providing a computationally viable means to assess tsunami hazards across various scenarios and regions. This versatility is especially valuable in areas where tsunamis can cause damage to infrastructure and pose serious risks to the population.

A compelling example of such a setting is landslide-induced tsunamis which differ from their seismic counterparts by (i) their extremely short arrival times since the source is often located near the coast, (ii) their potentially larger initial wave heights near the source, and (iii) the absence of clear precursors detectable by traditional Early Warning Systems (EWS; Roger et al., 2024; Lemoine et al., 2020). 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 study, a surrogate modeling framework specifically designed to support probabilistic tsunami forecasting (PTF) workflows for landslide-generated tsunamis is presented: the Landslide-Tsurrogate v1.0 model. The framework combines sparse grid sampling (Smolyak, 1963; Gerstner and Griebel, 1998; Constantine et al., 2012) with generalized Polynomial Chaos Expansions (gPCE; Xiu and Karniadakis, 2002; Soize and Ghanem, 2004) in order to approximate the response of computationally expensive tsunami models while requiring only a limited number of deterministic simulations and simultaneously providing direct access to sensitivity indices and uncertainty quantification metrics. Consequently, gPCE-based surrogate modeling offers an efficient alternative to machine learning approaches that often require extensive training data and may provide limited interpretability regarding the influence of uncertain input parameters.

Importantly, this surrogate modeling strategy is not tied to a specific tsunami generation mechanism. The gPCE approach approximates the response of the numerical tsunami model as a function of a set of uncertain input parameters, and is therefore applicable to any tsunami source that can be parameterized with a finite set of inputs. The main adaptation required for different tsunami sources lies in the definition of these input variables and their probability distributions. The framework can therefore be readily adapted to other tsunami sources provided that the dominant uncertain parameters controlling tsunami source generation can be identified. For example, in the case of submarine landslides considered in this study, the stochastic variables describe the physical and geometrical properties of the landslide (e.g., location, volume, etc.). In the case of sub-aerial landslides entering the ocean, if the model does not describe the sub-aerial flow, the parameters could also include the landslide mass entering the water, the entry velocity, the impact geometry, or the impact location. Similarly, for seismic tsunamis, the stochastic variables may represent earthquake source parameters such as fault slip, rupture length, or hypocentral location. Finally, for the forecast of atmospheric tsunamis (or meteotsunamis), the methodology has already been applied by defining the parameters of pressure disturbances (e.g., amplitude, direction, speed, period, etc.) driving these events in the Adriatic Sea (Denamiel et al., 2019, 2020).

Here, this broadly applicable approach is applied in the Mayotte region, where since the beginning of the seismo-volcano crisis in 2018, one particular concern was the question of a tsunami triggered by submarine landslides linked to intense seismic activity. Mayotte is a volcanic 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 May 2018, the island of Mayotte has registered intense seismic activities related to the birth of a large new submarine volcano 50 km offshore Petite Terre (Cesca et al., 2020; Lemoine et al., 2020; Feuillet et al., 2021). The intense seismo-volcanic crisis that has affected Mayotte since 2018, the location of earthquakes near the steep slopes

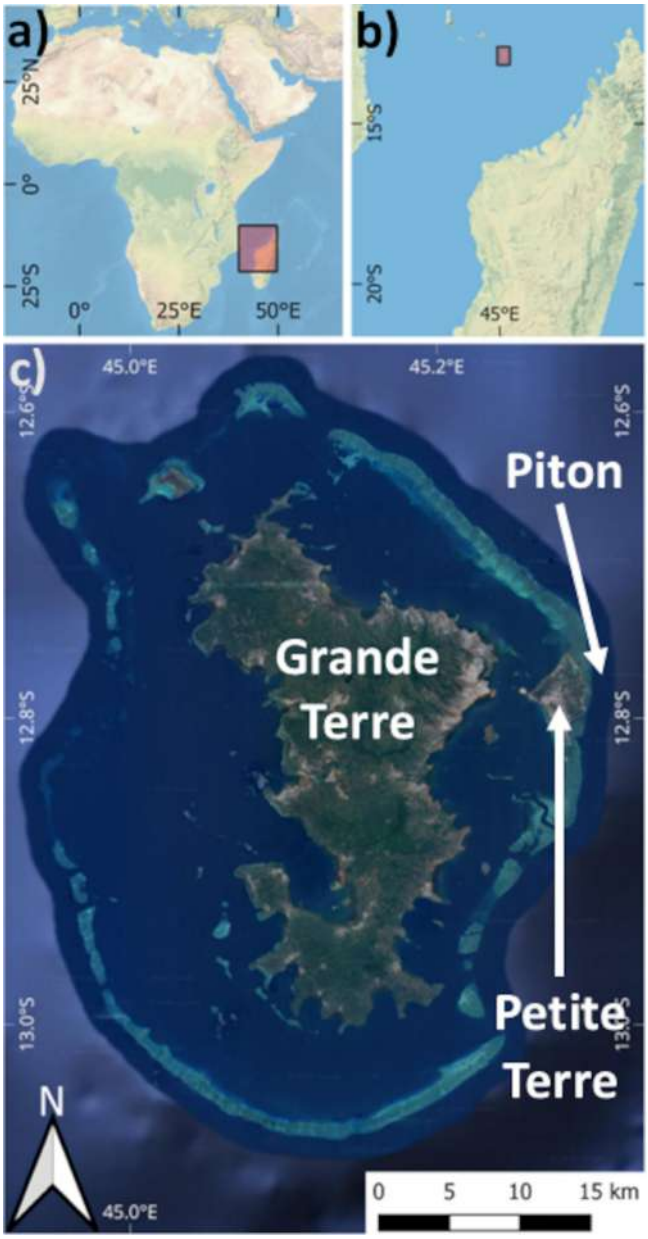


Figure 1. Location of the study area. (a) and (b) General maps (Map source: ESRI World Physical Map). The rectangle in (a) gives the extent of the map in (b), and the rectangle in (b) the extent of map in (c). (c) Mayotte lagoon geomorphology including Grande Terre, Petite Terre and Piton locations. (Map source: Google Satellite).

surrounding the lagoon around the islands, and the construction of a new volcanic structure (Lemoine et al., 2020; Feuillet et al., 2021; Mercury et al., 2022) may trigger submarine instabilities offshore East of Mayotte. Such collapses could trigger
85 submarine landslide-generated tsunamis with serious consequences for the coastal population as shown by recent simulations

(Roger, 2019; Lemoine et al., 2020; Poulain et al., 2022; Marbœuf et al., 2025). This application in Mayotte serves both to validate the surrogate model approach and to highlight its operational potential for hazard-prone regions with limited resources.

The article is structured as follows. To ensure transparency and reproducibility, Section 2 presents the full mathematical framework of Landslide-Tsurrogate v1.0, which forms the scientific and computational basis of the model. Section 3 and Appendix A outline the implementation of the framework, from the selection and representation of input uncertainties to their integration within a complete tsunami hazard assessment workflow. These sections aim to provide sufficient detail for readers to produce surrogate models in different contexts. Section 4 and Appendix B then demonstrate the use of Landslide-Tsurrogate v1.0 through an application to submarine landslide-generated tsunamis in Mayotte, serving as a real-world validation of the approach. Finally, Section 5 presents the limits of the gPCE-based surrogate modeling approach and Section 6 summarizes the main conclusions and discusses potential future developments.

2 Surrogate Modeling Approach

The primary objective of this study is to develop a model capable of delivering faster-than-real-time probabilistic forecasts of tsunami hazards triggered by submarine landslides. In this context, the uncertainty associated with submarine landslide parameters needs to be fully explored through the evaluation of large ensembles of landslide–tsunami simulations. Running such ensembles with complex deterministic numerical models can be computationally prohibitive, which motivates the development of surrogate modeling approaches capable of rapidly approximating the system response across the parameter space. To address this challenge, Landslide-Tsurrogate v1.0 employs a computationally efficient surrogate modeling strategy based on generalized Polynomial Chaos Expansions (gPCE; Xiu and Karniadakis, 2002; Soize and Ghanem, 2004). Hereafter, to illustrate the mathematical concepts, Figure 2 provides a schematic overview of the main components of the surrogate modeling approach.

2.1 Theoretical Framework

A key advantage of the gPCE method lies in its suitability for uncertainty propagation, as statistical moments and sensitivity indices can be directly derived from the gPCE coefficients. The choice of gPCE as the mathematical framework for Landslide-Tsurrogate v1.0 is grounded in the assumption by Ernst et al. (2012) that the gPCE representations of scalar tsunami hazard quantities—such as maximum tsunami elevation, maximum tsunami speed, and time of arrival, hereafter collectively referred to as $\xi_{\max}$—converge. Specifically, the mean-square error between the gPCE approximation $\xi_{\mathrm{gPCE},p}$ and the hazard quantity $\xi_{\max}$ tends toward zero as the expansion order p increases. This assumption is justified provided that $\xi_{\max}$ has finite variance; a condition that is generally reasonable from a physical standpoint. Figure 2 illustrates this assumption for a theoretical non-physical modeled quantity of interest $\xi_{\max}$ (no unit) and an expansion order $p \in 0, \ldots, 6$.

Additionally, one of the main difficulties in assessing submarine landslide-tsunami hazards lies in generating plausible landslide scenarios. This challenge stems from scarce data on critical factors such as the volume, shape, location, and frequency of tsunamigenic landslides, the availability of high-resolution digital elevation models, and incomplete knowledge of landslide

mechanics and material properties (Roger et al., 2024; Løvholt et al., 2020). Consequently, stochastic variables required to generate submarine landslide scenarios may vary in number r between regions and have various prior distributions (i.e., normal, uniform, etc.). However, due to the lack of observations, the stochastic variables are often assumed to follow uniform distributions—i.e., within the range of the uniform distributions, the generation of any submarine landslide is equally probable. For a set of elementary events ($\omega \in \Omega$) and $i \in 1, 2, \ldots, r$, the stochastic variable prior distributions are thus given by:

$$R_i : \begin{cases} \Omega \to \mathbb{R} \\ \omega \to R_i(\omega) \end{cases} \quad , \quad R_i(\omega) \sim \mathcal{U}([a_i, b_i]) \tag{1}$$

Figure 2a shows an example of Landslide-Tsurrogate v1.0 employing $r = 3$ stochastic input variables, each defined as uniformly distributed random variables. In this illustrative case, no specific physical meaning or units are assigned to these variables.

Importantly, the stochastic variables must be mutually independent, and the bounds of each uniform distribution must be selected to encompass all conditions known to trigger landslide-generated tsunamis within a given region. The use of uniform distributions and the assumption of independence are both grounded in the maximum entropy principle. According to this principle, when limited information is available, the distribution with the highest entropy is preferred, as it is based on the fewest assumptions (Jaynes, 1957). Nevertheless, adopting this approach implies that many of the randomly generated submarine landslide scenarios may, in fact, lack the capacity to produce significant tsunamis in the area under investigation.

2.2 Generalized Polynomial Chaos Expansions

2.2.1 General Framework

The gPCE mathematical framework is based on polynomial expansions that decompose into deterministic coefficients ξ_α and multivariate orthogonal polynomial bases $\Psi_\alpha(R_1, R_2, \ldots, R_r)$ depending on the multi-index α. The multi-index $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_r]$ provides the polynomial degree α_i associated with the stochastic variable $R_i, \quad i \in 1, 2, \ldots, r$. Figure 2b illustrates the derivation of the multi-index α for $r = 3$ and $p \in 0, 2, 4, 6$.

The optimal expansion—minimizing the error for a maximum total order p such as $|\alpha| = \sum_{i=1}^{r} \alpha_i \leq p$—is then given by the following truncated gPCE:

$$\xi_{\text{gPCE},p}(R_1, R_2, \ldots, R_r) = \sum_{|\alpha| \leq p} \xi_\alpha \Psi_\alpha(R_1, R_2, \ldots, R_r) \tag{2}$$

In the context of Landslide-Tsurrogate v1.0, the multivariate orthogonal polynomial basis is expressed as the product of the Legendre polynomials $\text{Le}_{\alpha_i}, \quad i \in 1, 2, \ldots, r$ which are orthogonal with respect to uniform distributions—i.e., $\mathbb{E}[\Psi_\alpha, \Psi_\beta] = \delta_{\alpha,\beta} \|\Psi_\alpha\|^2$ with $\delta_{\alpha,\beta}$ the Kronecker symbol—and are given by:

$$\Psi_\alpha(R_1, R_2, \ldots, R_r) = \prod_{i=1}^{r} \text{Le}_{\alpha_i} \left[\frac{(2R_i - a_i - b_i)}{(b_i - a_i)} \right] \tag{3}$$

Figures 2c,d, illustrate the values of the Legendre polynomials Le_{α_i} for $\alpha_i \in 0, 1, \ldots, 6$ and the orthogonal polynomial bases $\Psi_\alpha(R_1, R_2, R_3)$ for $p \in 0, 2, 4, 6$.

The deterministic coefficients ξ_α are the projection of the landslide-generated tsunami hazards represented by ξ_{max} onto each polynomial basis of the gPCE:

$$\xi_\alpha = \frac{1}{\|\Psi_\alpha\|^2}\mathbb{E}[\xi_{\max}\Psi_\alpha(R_1, R_2, \ldots, R_r)] \tag{4}$$

However, in the absence of an analytical formulation of tsunami hazards, numerical integration methods are employed to approximate the expectation $\mathbb{E}[\xi_{\max}\Psi_\alpha(R_1, R_2, \ldots, R_r)]$. One common method for computing multidimensional integrals is to express them as a series of nested one-dimensional integrals using Fubini's theorem (also known as the tensor product rule). However, this technique leads to an exponential increase in the number of required function evaluations—and thus, in the presented context, the deterministic landslide-tsunami simulations—as the number of dimensions grows, a phenomenon known as the curse of dimensionality. To address this issue, three main alternative methods have been developed to mitigate the computational burden: the Monte Carlo, Sparse Grids and Bayesian Quadrature methods.

2.2.2 Pseudo-Spectral Approximation (PSA)

Within Landslide-Tsurrogate v1.0, the deterministic coefficients ξ_α are computed using a Smolyak-based sparse-grid approach (Smolyak, 1963): the non-intrusive Pseudo-Spectral Approximation (PSA) method (Gerstner and Griebel, 1998; Constantine et al., 2012). One of the key advantages of PSA is that the error decreases rapidly as the number of needed deterministic simulations increases (Fornberg and Sloan, 1994). In this context, the truncated generalized Polynomial Chaos Expansion (gPCE) can be reformulated to yield the main equation of the Landslide-Tsurrogate v1.0 model:

$$\xi_{\mathrm{PSA},p}(R_1, R_2, \ldots, R_r) = \sum_{max(0,p-r+1)\leq|\alpha|\leq p} (-1)^{p-|\alpha|} \binom{r-1}{p-|\alpha|} \sum_\beta C_{|\alpha|,\beta}\Psi_\beta(R_1, R_2, \ldots, R_r) \tag{5}$$

For each α from $max(0, p-r+1) \leq |\alpha| \leq p$, the multi-index $\boldsymbol{\beta} = [\boldsymbol{\beta_1}, \boldsymbol{\beta_2}, \ldots, \boldsymbol{\beta_r}] \in \boldsymbol{A}$ stores all the polynomial combinations needed to calculate the tensor product $\sum_\beta C_{|\alpha|,\beta}\Psi_\alpha(R_1, R_2, \ldots, R_r)$ with:

$$C_{|\alpha|,\beta} = \frac{Q_\alpha^{(1:r)}(\xi_{max}\Psi_\beta)}{\|\Psi_\beta\|^2} \tag{6}$$

and the multivariate quadratic rule:

$$Q_\alpha^{(1:r)}(\xi_{max}\Psi_\beta) = \sum_\alpha w_\alpha^{(1:r)}\xi_{max}(s_\alpha^{(1:r)})\Psi_\beta(s_\alpha^{(1:r)}) \tag{7}$$

The complete set of abscissas $s_\alpha^{(1:r)}$—i.e., the combination of parameters sampling the defined ranges of the stochastic variables—and their associated weights $w_\alpha^{(1:r)}$ required for computing the multivariate quadrature rule are then obtained through

the tensor products:

$$\begin{cases} s_{\alpha}^{(1:r)} = s_{\alpha_1}^{(1)} \otimes s_{\alpha_2}^{(2)} \otimes \cdots \otimes s_{\alpha_r}^{(r)} \\ \\ w_{\alpha}^{(1:r)} = w_{\alpha_1}^{(1)} \otimes w_{\alpha_2}^{(2)} \otimes \cdots \otimes w_{\alpha_r}^{(r)} \end{cases} \tag{8}$$

where the univariate sampling of each stochastic variable R_i is defined by the S_i abscissas $s_{\alpha_i}^{(i)} = \left\{ s_k^{(i)} \right\}_{k=1}^{S_i}$ and the corresponding weights $w_{\alpha_i}^{(i)} = \left\{ w_k^{(i)} \right\}_{k=1}^{S_i}$, for $i = 1, 2, \ldots, r$. In practice, the unique set of abscissas extracted from $s_{\alpha}^{(1:r)}$ forms the ensemble of input parameters that will be fed to the deterministic landslide-tsunami model. For the illustrative case, Figure 2e shows examples of different set of unique abscissas $s_{\alpha}^{(1:3)}$ selected from the distributions of the $r = 3$ stochastic variables depending on the total order $p \in 0, 2, 4, 6$.

2.3 Quadrature Rule

The selection of an appropriate quadrature rule for sampling the stochastic variables is critical, as it determines both the accuracy of the surrogate models and the total number of deterministic landslide-tsunami simulations to be performed. Among the most commonly used quadrature rules for uniform distributions are the Gauss–Legendre (Abramowitz and Stegun, 1965), Clenshaw–Curtis (Clenshaw and Curtis, 1960), and Gauss–Patterson (Patterson, 1968) rules.

The Gauss–Legendre quadrature is the standard n-point rule for uniform distributions, offering exact integration for any polynomial of degree up to $2n - 1$. However, a key limitation of the Gauss–Legendre rule is that its abscissas are not nested—i.e., the points used in the n-point rule are entirely different from those in the $(n - 1)$-point rule, preventing reuse of previous evaluations.

Table 1. Number of abscissa S depending on the level of different nested quadrature rules.

	Level l						
Quadrature rule	0	1	2	3	4	5	6
Clenshaw–Curtis	1	3	5	9	17	33	65
Gauss–Patterson	1	3	7	15	31	63	127
Delayed Gauss–Patterson	1	3	3	7	7	7	15

In contrast, nested quadrature rules allow the abscissas of a lower-order rule to be included in higher-order ones, which significantly enhances computational efficiency. The Clenshaw–Curtis and Gauss–Patterson rules are the most frequently used nested quadrature rules for uniform distributions. The main difference between these two lies in their precision and growth rate: the Gauss–Patterson rule typically provides nearly double the accuracy and point growth compared to the Clenshaw–Curtis rule. In nested rules, the concept of level is introduced to denote the quadrature rule where the abscissas are nested. At level l, the number of abscissas required is $2^l + 1$ for Clenshaw–Curtis and $2^{l+1} - 1$ for Gauss–Patterson (Table 1). To make Gauss–Patterson

rules more efficient—by limiting the rapid growth in the number of abscissas—the introduction of successive quadrature levels can be delayed. This delayed Gauss–Patterson rule (Holtz, 2011) uses a reduced set of abscissas (Table 1) while preserving high integration accuracy, typically at least $2l-1$. Consequently, this approach significantly reduces the number of required quadrature points without sacrificing the quality of the surrogate model.

In Landslide-Tsurrogate v1.0, only the Gauss–Patterson (GP) and delayed Gauss–Patterson (DGP) quadrature rules are implemented, as they offer greater precision than the Clenshaw–Curtis rule. Despite this theoretical advantage, both Gauss-Patterson and Clenshaw–Curtis rules have demonstrated comparable performance in practice (Trefethen, 2008). However, following the findings of Holtz (2011) and Denamiel et al. (2019), the DGP rule not only reduces the number of deterministic simulations required but might also enhances the accuracy of the surrogate models. Therefore, the use of the DGP rule is strongly recommended for sampling the definition intervals of the stochastic variables in Landslide-Tsurrogate v1.0. Figure 2f presents the number of deterministic simulations N needed to create gPCE-PSA-based surrogate models with the DGP quadrature rule for the modeled quantity of interest ξ_{max} depending on the total order $p \in 0,\dots,6$.

Finally the GP and DGP rules solve univariate quadrature problem for stochastic variables following the standard distribution $Z \sim \mathcal{U}([-1,1])$. The relation between standard distributions and the distributions used in Landslide-Tsurrogate v1.0 can be expressed as follows, for a set of elementary events ($\omega \in \Omega$) and $i \in 1,2,\dots,r$:

$$\begin{cases} Z_i : \begin{cases} \Omega \rightarrow \mathbb{R} \\ \omega \rightarrow Z_i(\omega) \end{cases} , \quad Z_i \sim \mathcal{U}([-1,1]) \\ \\ R_i = \left(\frac{a_i+b_i}{2}\right) + \left(\frac{b_i-a_i}{2}\right) Z_i, \quad R_i \sim \mathcal{U}([a_i,b_i]) \end{cases} \tag{9}$$

Similarly the relationship between the abscissa and weights calculated for the standard distributions and the ones used in Landslide-Tsurrogate v1.0 can be written, for $i \in 1,2,\dots,r$:

$$\begin{cases} \begin{cases} s_{\alpha_i}^{(i)} = \left\{ s_k^{(i)} \right\}_{k=1}^{S_i} \\ \\ w_{\alpha_i}^{(i)} = \left\{ w_k^{(i)} \right\}_{k=1}^{S_i} \end{cases} , \quad Z_i \sim \mathcal{U}([-1,1]) \\ \\ \begin{cases} \left(\frac{a_i+b_i}{2}\right) + \left(\frac{b_i-a_i}{2}\right) s_{\alpha_i}^{(i)} \\ \\ \left(\frac{b_i-a_i}{2}\right) w_{\alpha_i}^{(i)} \end{cases} , \quad R_i \sim \mathcal{U}([a_i,b_i]) \end{cases} \tag{10}$$

Landslide-Tsurrogate v1.0 is built upon the full system of equations [Eq. (1) to Eq. (10)] presented in the previous subsections, while the numerical simulations required to compute the quadrature rule can be generated using any suitable landslide-tsunami model. For a given region and a given set of stochastic variables, the surrogate models are built at specific locations

$m \in 1, 2, \ldots, M$ generally situated along the coastline. Figure 2g illustrate the behavior of a surrogate model in function of the total order $p \in 0, 2, 4, 6$. However, the convergence of surrogate models is slower or even impossible when different sets of abscissas (i.e., input parameters of the landslide-tsunami model) are linked to the same value of ξ_{max} (e.g., $0\mathrm{m}$ for the maximum elevation in the inundation zone when there is no inundation). Consequently, in this study, the locations where the surrogate models are built are deep enough to remain submerged—i.e., outside of the inundation zone—ensuring that results of maximum elevation, maximum speed and time of arrival are continuous—i.e., defined with different values for all the deterministic landslide-tsunami simulations.

2.4 Convergence and Accuracy

Once generated with Landslide-Tsurrogate v1.0, two different evaluation steps are needed to demonstrate the surrogate model skill in estimating the tsunami hazards—maximum elevation, maximum speeds and time of arrival. These steps are the convergence and accuracy of the surrogate models based on the comparison between the tsunami hazards predicted with the N-member ensembles of deterministic simulations $\{\xi_{\text{max}}^{m,n}\}_{n=1}^{N}$ and surrogate model results $\left\{\xi_{\text{PSA},p}^{m,n}\right\}_{n=1}^{N}$ at the chosen locations $m \in 1, 2, \ldots, M$ and for the total order $p \in 1, 2, \ldots, P$.

Firstly, the convergence of the surrogate models is analyzed with the normalized root-mean-square errors at each total order $p \in 1, 2, \ldots, P$ and for each chosen location $m \in 1, 2, \ldots, M$:

$$e_p^m = \frac{\sum_{n=1}^{N} \left(\xi_{\text{max}}^{m,n} - \xi_{\text{PSA},p}^{m,n}\right)^2}{\sum_{n=1}^{N} \left(\xi_{\text{max}}^{m,n} - \xi_{\text{PSA},0}^{m,n}\right)^2} \tag{11}$$

The convergence criteria is fulfilled when the errors [Eq. (11)] follow $\lim_{p \to P_t} e_p^m = 0$. In this context, the total order of the gPCE selected to build the surrogate models is P_t. Figure 2g represents this convergence of a gPCE-PSA-based surrogate model $\xi_{PSA,p}$ with $p \in 0, 2, 4, 6$ towards the illustrative quantity of interest ξ_{max} when p increases.

Secondly, evaluating the accuracy of the surrogate models theoretically requires an ensemble of deterministic simulations that (1) are sampled using a Monte Carlo method and (2) are large enough—typically on the order of 10^4 runs—to produce statistically robust results. In practice, however, this approach is not viable, since the primary motivation for constructing surrogate models is to reduce the computational burden associated with running a large number of deterministic simulations. Additionally, at higher total polynomial orders, reusing the same simulations used to build the surrogate risks overfitting. Within the Landslide-Tsurrogate v1.0 framework, which relies on nested sparse grids, a more practical strategy involves constructing surrogate models at a total order higher $(P_t + 1)$ than the one selected for convergence (P_t). The additional deterministic simulations ΔN generated at this higher order being independent—i.e., not used to build the surrogate models—and effectively sampling the full range of the stochastic parameter space can therefore serve as a robust basis for accuracy assessment. This is shown for the illustrative case in Figures 2e,f where, for example, the additional set of abscissas and, hence, the additional deterministic simulations, $\Delta N = +48 + 72 = +120$ needed to go from $p = 4$ (in yellow) to $p = 6$ (in red) clearly covers the entire space of definition of the stochastic variables R_1, R_2, R_3.

2.5 Sensitivity Analysis

A key challenge is to determine which stochastic variable R_i for $i \in 1, 2, \ldots, r$ most significantly influence the variability of landslide-generated tsunami hazards at the selected locations of interest $m \in 1, 2, \ldots, M$. In other words, the goal is to identify the parameters to which the hazard estimates produced by the surrogate models are most sensitive.

This sensitivity analysis is carried out using variance-based decomposition methods (ANOVA), in particular Sobol' sensitivity indices (Sobol′, 2001; Saltelli, 2002), which can be efficiently derived from the coefficients of the gPCE representation (Crestaux et al., 2009). Following the notations introduced in the previous subsections, the Sobol' indices are derived from the gPCE coefficients for which $1 \leq |\alpha| \leq P_t$ and $\boldsymbol{\beta} \in \boldsymbol{A}, \boldsymbol{\beta} \neq 0$. These gPCE coefficients can be expressed as:

$$\widehat{\xi_{\beta}^{m}} = (-1)^{P_t - |\alpha|} \begin{pmatrix} r-1 \\ P_t - |\alpha| \end{pmatrix} C_{|\alpha|,\beta}^{m} \tag{12}$$

At each chosen location $m \in 1, 2, \ldots, M$, the variability of the tsunami hazards D^m and the total Sobol' indices for each stochastic variable $S_{R_i}^{T,m}$ with $R_i \quad i \in 1, 2, \ldots, r$ are given by:

$$\begin{cases} S_{R_i}^{T,m} = \frac{D_i^{T,m}}{D^m} \\ \\ D^m = \mathrm{Var}\left[\xi_{\mathrm{PSA},Pt}^{m}\right] = \sum\limits_{\substack{\boldsymbol{\beta} \in \boldsymbol{A} \\ \boldsymbol{\beta} \neq 0}} \left(\widehat{\xi_{\beta}^{m}}\right)^2 \\ \\ D_i^{T,m} = \sum\limits_{\boldsymbol{\beta} \in \boldsymbol{A}_i^T} \left(\widehat{\xi_{\beta}^{m}}\right)^2, \boldsymbol{A}_i^T \in \{\boldsymbol{\beta} \in \boldsymbol{A} : \beta_i > 0\} \end{cases} \tag{13}$$

The total Sobol' indices [Eq. (13)] quantify the relative contribution (ranging from 0 to 1) of each stochastic variable to the variability of landslide-generated tsunami hazard predictions at the selected locations. Since total indices account for both individual effects and all possible interactions with other variables, their sum may exceed 1—reflecting that interactions are counted multiple times. In practice, if a total sensitivity is negligible (e.g., below 0.01), the corresponding stochastic variable can be fixed to a deterministic value without significantly impacting the resulting hazard distributions.

3 Landslide-Tsurrogate v1.0

Landslide-Tsurrogate v1.0 is a computational tool designed to support the development of gPCE-based surrogate models for regions affected by landslide-generated tsunamis. It includes a comprehensive set of functions (i.e., MATLAB and Python codes), a step-by-step User Manual provided as a Jupyter Notebook to guide users through the surrogate modeling workflow, and example user-friendly interfaces implemented in both Jupyter Notebook and MATLAB, all illustrated with the Mayotte test case (https://zenodo.org/records/17519408; Denamiel, 2025).

The construction of surrogate models with Landslide-Tsurrogate v1.0 follows seven building blocks or steps, as illustrated in Figure 3. This section discusses in detail the strengths and limitations of the methodological choices implemented in Landslide-Tsurrogate v1.0. For the steps that involve user-defined inputs (marked with *), the various options available to users are also outlined to highlight the versatility of the tool.

1. ***User Specifications****

   This step involves defining the stochastic variables used to represent the uncertain characteristics of potential submarine landslides. These variables describe the physical and geometric properties of the landslide source and can be grouped into three main categories: (i) soil parameters, such as cohesion, friction angle, and density; (ii) landslide geometry, including the thickness, length, width, and orientation of the shapes used to represent the landslide volume initially released; and (iii) initial conditions, such as the mean sea level, which may vary due to tidal fluctuations and long-term sea-level rise. In principle, the initial released mass and failure characteristics of a submarine landslide could be determined using mechanical slope stability models that incorporate measured soil properties (e.g., density, cohesion, internal friction angle), as well as hydrogeological conditions, structural discontinuities, and fault networks. However, such detailed information is rarely available for offshore environments. Consequently, many landslide properties remain poorly constrained and represent a major source of epistemic uncertainty in landslide-tsunami hazard assessment.

   Within the Landslide-Tsurrogate v1.0 framework, these uncertainties are explicitly represented through stochastic variables. Potential landslide locations, for example, may be identified using geomorphological indicators such as steep seabed slopes, sediment accumulations, previously mapped instability features or areas experiencing high seismicity. Landslide volumes, geometric parameters, and rheological properties can then be sampled within plausible ranges derived from regional geological knowledge, previous studies, or expert judgment. Similarly, mean sea level can be treated as an additional stochastic variable to capture the nonlinear interactions between tsunami propagation, tidal variability, and long-term sea-level rise.

   By representing the uncertain landslide characteristics using a limited set of stochastic variables (e.g., location, volume, friction angle, and sea level), the dimensionality of the problem can be kept manageable while still capturing the dominant sources of uncertainty affecting landslide-generated tsunamis. These variables define the stochastic parameter space used to generate the deterministic landslide-tsunami simulations required to construct the surrogate models. The stochastic space is subsequently sampled using sparse quadrature rules, allowing the surrogate models to efficiently approximate the response of the deterministic simulations across the range of possible landslide scenarios.

   Finally, it is important to note that the definition of stochastic variables and their ranges is inherently site dependent. Some uncertainties reflect natural variability in environmental conditions (aleatory uncertainty), while others arise from limited knowledge of subsurface properties and landslide triggering mechanisms (epistemic uncertainty). The proposed framework does not prescribe a unique parameterization but instead provides a flexible structure for propagating these uncertainties through the deterministic simulations.

2. ***Input Parameters***

This step 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 stochastic variables used to describe the physical and geometric properties of the landslides. Once these stochastic variables and their probability distributions have been defined in the previous step, a sparse Gauss–Patterson quadrature is constructed up to a user-defined total order P. The quadrature nodes—representing collocation points in the stochastic parameter space—are mapped from the canonical space to the physical space. Each combination of parameters defines a deterministic landslide scenario that is then simulated using a numerical landslide-tsunami model.

The total number of deterministic simulations N is determined by three main factors: (i) the number of stochastic variables r, (ii) the selected total polynomial order P used to construct the surrogate model, and (iii) the quadrature strategy adopted (Gauss–Patterson or delayed Gauss–Patterson). In practice, increasing either the number of stochastic variables or the total order increases the number of collocation points required to accurately represent the stochastic space. The quadrature rule determines how efficiently this space is sampled.

For example, using six stochastic variables $(r = 6)$ and a selected total order $P = 5$, the number of required simulations is $N = 10,625$ with the Gauss–Patterson (GP) sparse grid and $N = 1,889$ using the delayed Gauss–Patterson (DGP) approach. The DGP method therefore provides a substantial reduction in the number of deterministic simulations while maintaining comparable surrogate accuracy.

The choice of P represents a trade-off between surrogate model convergence/accuracy and computational cost. Higher orders improve the ability of the surrogate model to represent nonlinear interactions between parameters but require more deterministic simulations. Because each deterministic simulation involves running a full numerical tsunami model, this stage determines the computational cost of the framework.

For this reason, surrogate models can be constructed incrementally by truncating the polynomial expansion to lower orders (e.g., $P = 4$) and only generating additional deterministic simulations to reach the polynomial order needed for the convergence of the surrogate models (e.g., $P_t = 6$).

3. ***Deterministic Simulations****

This 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. The surrogate models developed in Landslide-Tsurrogate v1.0 are trained on deterministic tsunami simulations, which therefore constitute a fundamental component of the overall framework. The choice of the numerical model used to generate these simulations should be guided by the physical complexity of the system considered. Depending on the characteristics of the landslide source, the bathymetry, and the relevant hydrodynamic processes, different levels of model complexity may be required. In some cases, depth-averaged shallow-water models may provide adequate accuracy for hazard assessment, while in other settings more sophisticated

approaches, such as dispersive or fully three-dimensional models, may be necessary to properly represent the generation mechanism and the resulting wave field. The surrogate model does not impose constraints on the deterministic solver and can in principle emulate the outputs of any numerical model, provided that the training dataset sufficiently samples the parameter space and that the underlying simulations capture the relevant physical processes. Consequently, the reliability of the probabilistic hazard estimates ultimately depends on the adequacy of the deterministic model used to generate the training simulations.

Among the state-of-the-art numerical models available within the submarine landslide community, several fully non-linear and dispersive solvers have been developed with varying levels of complexity, computational cost, and intended applications. NHWAVE (Ma et al., 2012) is a three-dimensional, non-hydrostatic model particularly suited for simulating submerged landslides and the resulting tsunami generation. It provides high accuracy in capturing the wave generation and early propagation phases, but its 3D nature makes it computationally expensive, especially for large domains or probabilistic assessments. Geowave (Watts et al., 2003) is a coupled model that integrates a Boussinesq wave solver with a solid slide motion solver, allowing for flexible representation of landslide kinematics but relying on simplified assumptions for slide rheology. BING (Løvholt et al., 2005) represents the landslide as a viscous fluid and is often coupled with long-wave tsunami solvers such as COMCOT (Liu et al., 1998) or COULWAVE (Lynett et al., 2002; Lynett and Liu, 2004), striking a balance between physical realism and computational cost. Finally, Landslide-HySEA (Macías et al., 2021a, b) offers a unique advantage through its multi-layer non-hydrostatic shallow water formulation, which is especially important for capturing the vertical structure and dynamics of submarine mass movements. This multi-layer approach enables the realistic representation of both submerged and subaerial landslide processes and their interaction with the water column, leading to improved accuracy in wave generation. The HySEA suite is also GPU-accelerated, which significantly reduces computational time despite its complexity, making it particularly attractive for regional hazard assessments that require large ensembles of simulations.

Each of these models has trade-offs: NHWAVE and Landslide-HySEA offer higher fidelity but at greater computational cost, while models like Geowave and BING are more efficient but rely on simplifying assumptions. In addition, the setup of the simulations including domain size, horizontal resolution, number of layers (e.g., MHWAVE and Landslide-HySEA), Manning friction coefficient, CFL number, dry land threshold value, etc., will greatly impact the accuracy of deterministic simulations and, hence, of the surrogate model-based PTF. The balance between numerical cost and accuracy is thus constraint by the availability of computational resources which, practically, plays a crucial role on the choice of the numerical model and simulation setup.

4. ***Data Transfer and Formatting****

The deterministic simulations required to build accurate surrogate models can generate a large volume of data. As such, transferring and formatting this dataset becomes a critical step. However, data management largely depends on the computing environment used (e.g., HPC vs. local server). Landslide-Tsurrogate v1.0 only requires the user to provide

a specifically formatted file that must contain the maximum elevation, maximum speed, and time of arrival of the N tsunami waves extracted at the selected M locations, where the surrogate models are to be constructed.

5. ***Surrogate Models***

This is the main building block of Landslide-Tsurrogate v1.0. Here, the deterministic coefficients of the gPCE decomposition are computed based on the selected quadrature rule and the faster-than-real-time probabilistic tsunami hazard assessments are provided at the chosen M locations. A couple of important aspects of the methodological choices used to create the surrogate models should be highlighted.

Firstly, in Landslide-Tsurrogate v1.0, the logarithm of the tsunami hazard quantities is used to build the surrogate models instead of their raw values. This offers several advantages, particularly when dealing with quantities that span multiple orders of magnitude or exhibit skewed distributions. Log-transforming these quantities tends to normalize their distribution, reduce heteroscedasticity (i.e., variability that changes with the magnitude), and suppress the influence of extreme outliers. This results in smoother response surfaces that are more amenable to polynomial approximations such as those used in generalized Polynomial Chaos Expansions (gPCE). Additionally, the log-transformation improves the numerical stability and convergence of the surrogate models, especially in regions where the hazard quantity may approach zero. However, care must be taken when applying the log-transform, as it requires all input values to be strictly positive. This condition is generally met by selecting offshore locations outside the inundation zone. Additionally, the log-transform can compress extreme values—slightly reducing the model's sensitivity to rare but intense events. Overall, applying the logarithmic transformation—used by default in Landslide-Tsurrogate v1.0—remains a valuable and robust strategy for enhancing surrogate model performance in tsunami hazard assessments.

Secondly, a notable characteristic of the generalized Polynomial Chaos Expansion (gPCE) approach is that the surrogate models are constructed independently at each location of interest. That is, the gPCE is built pointwise, meaning that the statistical relationships between input parameters and tsunami hazard quantities—such as maximum elevation, speed, or arrival time—are captured separately for each site without considering spatial correlations. This is in contrast to other methods such as Gaussian Process Regression (GPR; e.g., Rasmussen and Williams, 2005) or Deep Neural Networks (DNNs; e.g., Luo et al., 2025), which can exploit shared structure or dependencies across space and potentially model multiple outputs jointly. While the pointwise nature of gPCE ensures computational simplicity and interpretability (Xiu and Karniadakis, 2002; Sudret, 2008), it may limit the method's ability to account for spatial coherence in the tsunami signal, particularly in complex coastal environments where hazard responses are spatially correlated.

In the context of landslide-generated tsunami hazard assessment, the gPCE choice made in Landslide-Tsurrogate v1.0 is particularly well-suited. The underlying physics of tsunami generation and propagation are smooth and governed by well-understood equations, making them ideal candidates for polynomial-based surrogates. Moreover, the relatively low number of stochastic input parameters—carefully defined by the user—keeps the dimensionality within the range where gPCE remains both accurate and efficient. The method's ability to provide an interpretable analytical structure, its natural convergence monitoring through sparse quadrature rules, and its rapid evaluation speed make it a robust and practical

choice. While alternative surrogate approaches may offer advantages in high-dimensional or highly nonlinear contexts, gPCE provides a highly effective balance of performance, transparency, and computational efficiency for the specific demands of probabilistic tsunami hazard assessment.

6. ***Evaluation***

The evaluation step is a key component of Landslide-Tsurrogate v1.0, aimed at assessing the convergence (necessary to define the gPCE truncation order), accuracy, and sensitivity (needed to exclude certain stochastic variables) of the surrogate models.

A key strength of the gPCE framework used in Landslide-Tsurrogate v1.0 lies in its ability to assess convergence at total order P_t 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, users can quantitatively evaluate the stability and accuracy of the expansion and define the order P_t for which the convergence is reached. This nested structure allows for computational efficiency since the nodes from the lower-order quadrature are reused in higher-order computations, minimizing redundant simulations. Moreover, this approach provides a natural mechanism for adaptive refinement, as discrepancies between orders highlight regions where the surrogate may be under performing. However, this method also has limitations. The convergence between successive orders may be slow or misleading in cases where the model response is highly non-linear or discontinuous in the stochastic space, particularly in the presence of rare but extreme events. Additionally, while nested grids improve efficiency, the cost of running additional deterministic simulations for higher orders can become prohibitive as the dimensionality increases. Nonetheless, convergence checks remain a practical and valuable diagnostic tool for ensuring surrogate model robustness within the gPCE framework.

In Landslide-Tsurrogate v1.0, the practical approach used to assess the accuracy of gPCE surrogate models is to validate the surrogate built at total order P_t using the deterministic simulations corresponding to the quadrature nodes at order P_t+1. One key strength of this method is its efficiency: it avoids the additional computational burden of calculating all coefficients required for a full order P_t+1 surrogate, while still providing a meaningful diagnostic of the surrogate's generalization capability. This approach acts as a cross-validation scheme, testing how well the surrogate interpolates or extrapolates to unseen, but structured, parameter combinations. Moreover, because the P_t+1 quadrature includes more nodes and often better captures variability in the stochastic space, it can reveal weaknesses in the surrogate built at order P_t, especially in regions where the model response is highly sensitive. However, this methodology has its limitations. The validation points are not randomly distributed, but follow the deterministic structure of the quadrature grid, which might bias the assessment — particularly if certain regions of the input space are poorly covered by the higher-order nodes. Also, since the evaluation is based only on a subset of the full parameter space, it may overlook local deficiencies or provide an overly optimistic view of surrogate performance in regions where higher-order interactions dominate. Furthermore, this method does not capture errors related to the truncation of the polynomial chaos expansion itself,

which a full surrogate at order $P_t + 1$ might better expose. Still, when used judiciously, this validation strategy offers a valuable compromise between computational cost and model fidelity.

Finally, using total Sobol' sensitivity indices to discard stochastic variables is a practical strategy for reducing the complexity of surrogate models in Landslide-Tsurrogate v1.0. The main strength of this approach lies in its ability to quantify the overall influence of each input variable—including interactions—on the output variance, thus offering a solid basis for identifying parameters with negligible impact. Discarding such variables can significantly lower the computational cost of the surrogate construction without sacrificing accuracy. However, a key limitation is that Sobol' indices are global measures (i.e., they reflect the properties of the surrogate models across all the scenarios generated to cover the interval of definition of the stochastic variables), and their estimation can be sensitive to sampling strategy and noise. Additionally, a variable with low total influence on the entire domain might still be critical in localized regions or under specific conditions. Hence, while total Sobol' indices provide valuable guidance, decisions to eliminate variables should be made with caution and, ideally, confirmed by localized or scenario-specific analysis.

7. ***User-friendly Interface****

To facilitate the practical use of surrogate models by a wide range of end users—including scientists, engineers, and decision-makers—developing user-friendly interfaces is essential. Several approaches can be adopted depending on the target audience and deployment environment. A Python-based GUI offers flexibility, integration with scientific libraries (e.g., NumPy, Matplotlib), and cross-platform compatibility. It is well-suited for research and prototyping, but may require users to install dependencies and be comfortable with a scripting environment. A MATLAB-based GUI benefits from a cohesive environment and strong plotting capabilities, making it particularly attractive in academic settings where MATLAB is already in use. However, it depends on licensed software, limiting accessibility and long-term sustainability. In contrast, web-based services using frameworks like Flask (Python), Streamlit, or full-stack applications (e.g., with React or Django) can reach a broader audience and require no local installation—users only need a browser. These solutions are ideal for sharing models with stakeholders or policy users, but can be more complex to develop and maintain, especially if real-time performance or secure data handling is needed. Ultimately, the choice of interface depends on balancing accessibility, development resources, and the technical proficiency of the end users.

In practice, Landslide-Tsurrogate v1.0 is implemented following the technical description and workflow presented in Appendix A based on the seven steps (or building blocks) described above.

4 Mayotte Test Case

The Mayotte test case demonstrates the application of the Landslide-Tsurrogate v1.0 model by following the sequential steps presented as building blocks in Figure 3 and as a workflow pipeline in Figure A2.

4.1 Steps 1 and 2: User Specifications and Input Parameters

Mayotte's submarine geomorphology is strongly shaped by its volcanic origin. Steep submarine slopes offshore transition abruptly from shallow shelf (< 9°) to flank slopes reaching 25–60° locally, segmented by deep canyons and gullies (Lemoine et al., 2020; Feuillet et al., 2021). The eastern submarine slope of Mayotte would be probably prone to slope failure because it is composed of accumulation of carbonate and volcanic sediments on steep slopes. These mass-wasting events could be triggered by a combination of factors: steep slopes, weak layers of eroded volcanic ash or clay, hydrostatic pressure from groundwater, and seismic shaking linked to ongoing seismic swarms and active volcanic degassing near submarine edifices (Lemoine et al., 2020; Feuillet et al., 2021). Sultan et al. (2023) analyzed physical properties and mineralogy and performed dynamic triaxial tests on 25 sediment cores offshore of the eastern side of Mayotte to understand the hazards related to earthquake-induced submarine liquefaction. They show that the main parameter controlling the liquefaction potential offshore of Mayotte is the presence of low-density layers with high calcite content accumulating along the slope during lowstands.

Additionally, in the context of the on-going seismo-volcanic activity, Lemoine et al. (2020), Poulain et al. (2022) and Marbœuf et al. (2025) already implemented numerical simulations of potential submarine landslides and associated tsunamis in Mayotte. First, they showed that the most impacting submarine landslide scenarios were located on a large portion of the transition between the lagoon and offshore zones, presenting the steepest slopes and could have a volume that varied greatly depending on these locations. Second, they also highlighted the uncertainties around the friction angles in the landslide rheology and their high impact on the granular flow and deposit. Finally, of all the free parameters involved in their model, they demonstrated that these friction angles have the strongest impact on the wave field, compared, for example, to the Manning coefficient and the interlayer friction coefficient. Consequently, based on these previous studies, in this first attempt to apply the Landslide-Tsurrogate v1.0 framework in Mayotte, the number of stochastic variables is limited to three: location, volume and friction angle.

In practice, within the Landslide-Tsurrogate framework, Mayotte submarine landslides are pragmatically located along the isolines of steepest slope (extracted from high-resolution bathymetry) close to the seismic swarm related to the seismo-volcanic activity. They serve as coherent proxies for collapse-prone zones in the absence of additional comprehensive subsurface data. As illustrated in Figure 4, five distinct zones have been selected in the Mayotte region based on the isolines of steepest slope: Piton, Piton Offshore, Piton South, Mayotte South, and Mayotte Offshore. For the volume of the submarine landslides, volumes between 1–200 $10^6\,\mathrm{m}^3$ have been selected based on (i) the works of Lemoine et al. (2020) and Poulain et al. (2022), (ii) the bathymetric data and (iii) modeling studies of landslides in shallow marine environments (Løvholt et al., 2015). The largest volumes $V = [1, 200]\ 10^6\,\mathrm{m}^3$ are used for the Piton location while $V = [1, 100]\ 10^6\,\mathrm{m}^3$ are chosen for the deeper zones of Piton Offshore and Mayotte Offshore. Smaller landslide volumes of 1–50 $10^6\,\mathrm{m}^3$ are selected for the Piton South and Mayotte South zones that are associated with lower scarps volumes. The rheology and associated friction angles that should be used in models of submarine landslides are still a debate. However, previous works calibrated the friction coefficient that could be used in the Mayotte's case (Poulain et al., 2022; Marbœuf et al., 2025). Following these studies, a conservative range of friction angles [4,12]° is used to simulate potential submarine landslides in Mayotte. Such small friction angles are also necessary to

reproduce the dynamics and deposits of submarine landslides involving large volumes (Poulain et al., 2023). The origin of this small dissipation in large landslides is however a challenging question (Poulain et al., 2022). Furthermore, the selected small values are aimed to describe the worst-case scenarios as often considered in tsunami hazard assessments, since low friction angles lead to higher velocities and runout distances (Løvholt et al., 2015).

In summary, for each of the five selected zones, the Landslide-Tsurrogate v1.0 framework is implemented with three stochastic variables: the along-isoline distance from the northernmost point (i.e., the point with the maximum latitude; D), the volume of the submarine landslide (V), and the friction angle (δ). These variables are defined using uniform probability distributions based on the range of physically plausible values derived from geological and geomorphological constraints, and expert knowledge:

- Piton (12.76°S, 45.31°E):

$$\begin{cases} D \sim \mathcal{U}([0, 919]\,\mathrm{m}) \\ V \sim \mathcal{U}([1, 200]\,10^6\,\mathrm{m}^3) \\ \delta \sim \mathcal{U}([4, 12]\,^\circ) \end{cases} \tag{14}$$

- Piton Offshore (12.77°S, 45.31°E):

$$\begin{cases} D \sim \mathcal{U}([0, 1190]\,\mathrm{m}) \\ V \sim \mathcal{U}([1, 100]\,10^6\,\mathrm{m}^3) \\ \delta \sim \mathcal{U}([4, 12]\,^\circ) \end{cases} \tag{15}$$

- Piton South (12.81°S, 45.30°E):

$$\begin{cases} D \sim \mathcal{U}([0, 5045]\,\mathrm{m}) \\ V \sim \mathcal{U}([1, 50]\,10^6\,\mathrm{m}^3) \\ \delta \sim \mathcal{U}([4, 12]\,^\circ) \end{cases} \tag{16}$$

- Mayotte South (12.90°S, 45.26°E):

$$\begin{cases} D \sim \mathcal{U}([0, 6071]\,\mathrm{m}) \\ V \sim \mathcal{U}([1, 50]\,10^6\,\mathrm{m}^3) \\ \delta \sim \mathcal{U}([4, 12]\,^\circ) \end{cases} \tag{17}$$

- Mayotte Offshore (12.89°S, 45.30°E):

$$\begin{cases} D \sim \mathcal{U}([0, 3568]\,\mathrm{m}) \\ V \sim \mathcal{U}([1, 100]\,10^6\,\mathrm{m}^3) \\ \delta \sim \mathcal{U}([4, 12]\,^\circ) \end{cases} \tag{18}$$

For the maximum total order corresponding to $P = 6$ and the delayed Gauss-Patterson (DGP) sparse grid sampling selected for the Mayotte test case, the number of combinations of location, volume and friction angle that define the input parameters of the deterministic simulations is $N = 207$ per zone. This number results directly from the DGP sparse quadrature construction for $r = 3$ stochastic variables and a polynomial expansion of total order $P = 6$ [Eq. (14) to Eq. (18)]. In practice, the surrogate models are constructed using a training dataset consisting of $N = 135$ deterministic simulations per zone corresponding to $P_t = 5$, while the remaining 72 simulations are used as an independent testing dataset to assess the predictive performance and robustness of the surrogate model. Consequently, a total of 1035 deterministic simulations are required to construct and validate the surrogate models across the five defined zones. The choice of $P_t = 5$ represents a compromise between surrogate accuracy and computational cost and depends on the convergence of the surrogate models that will be further discussed in Section 4.3.2. Lower orders may not adequately capture nonlinear interactions between landslide parameters, while higher orders would significantly increase the number of required deterministic simulations. For tsunami applications with a limited number of stochastic variables (typically $r \leq 5$), polynomial orders $P_t \in 4 \dots 6$ provide a suitable balance between accuracy and computational efficiency (e.g., Denamiel et al., 2019, 2020).

More generally, the number of deterministic simulations required by the Landslide-Tsurrogate framework depends primarily on the number of stochastic variables and the selected polynomial order. In practical applications, it is recommended to (i) limit the number of stochastic variables to the most influential physical parameters and (ii) select a polynomial order that captures the dominant nonlinear behavior without excessively increasing the number of deterministic simulations. The computational cost associated with these deterministic simulations and the parallelization strategy adopted for the Mayotte test case are discussed in Section 4.2.1.

4.2 Step 3: Deterministic Simulations

4.2.1 Deterministic Model Setup

The topo-bathymetric dataset used in this study is derived from the compilation presented by Lemoine et al. (2020), which combines multiple data sources with spatial resolutions ranging from 1 m to 100 m. These datasets were merged to construct an initial Digital Elevation Model (DEM) with a horizontal resolution of 10 m. For the deterministic tsunami simulations, this DEM was resampled to a 30 m resolution by removing rows and columns from the raster grid. Keeping in mind the total number of 1035 simulations to be performed, this resolution was selected as a compromise between spatial accuracy and computational efficiency while maintaining a realistic representation of the coastal and nearshore bathymetry. The computational domain covers a 21×25.5 km area defined by the local coordinate bounds $[x_{min}, x_{max}] \times [y_{min}, y_{max}] = [520075, 541075] \times [8570890, 8596390]$ in the EPSG:4471 (RGM04 / UTM zone 38S) coordinate system. The computational domain and the 30 m Mayotte's bathymetric data are displayed on Figure 4 left panel. Middle and right panels of Figure 4 show all the submarine landslide locations and the associated isolines of steepest slopes (dashed white lines) for the five considered zones: Piton (red dots on the middle panel), Piton Offshore (white dots on the middle panel), Piton South (white triangles on

the middle panel), Mayotte South (black dots on the right panel) and Mayotte Offshore (white dots on the right panel). The coastline from the SHOM Histolitt dataset is represented in all panels by a bold black line.

All the simulations are performed with Multilayer HySEA, the layer-averaged version of the HySEA model family (Macías et al., 2021a, b) described in Appendix B. Multilayer HySEA couples the landslide dynamics and the generated water waves through a friction term between landslide and water layers. Multilayer HySEA has been shown to be well suited to simulate landslide generated tsunamis in Mayotte. Indeed, a multilayer configuration is required to properly describe the effect of the strong topography variations related to the presence of a coral reef while keeping a reasonable computational cost compared to full 3D models. This enables performing the numerous simulations required for the surrogate model with the parameters set according to Marbœuf et al. (2025). The main features of all the simulations are given by:

- Final time $t_f = 1800$ s with CFL = 0.55
- Non-hydrostatic pressures with 4 vertical layers as recommended in Marbœuf et al. (2025) for Mayotte
- Open boundary conditions with a sponge layer of 4 cells absorbing the waves
- $\mu(I)$ rheology with a grain size diameter $d = 1$ m, a solid volume fraction $\Phi = 0.8$ and the first friction angle δ as a stochastic variable: see [Eqs. B3 and B4]
- Density ratio $r = \rho_f/\rho_s = 0.5$ where $\rho_f = 1000\ \mathrm{kg/m^3}$ and $\rho_s = 2500\,\Phi = 2000\ \mathrm{kg/m^3}$
- Variable Manning coefficient ranging from 0.025 to $0.07\ \mathrm{s.m^{-1/3}}$ offshore (Lemoine et al., 2020; Marbœuf et al., 2025)
- A mass / water friction of $m_f = 0.004\ \mathrm{m.s^{-1}}$

The second friction angle is calculated thanks to $\delta_2 - \delta = 10^\circ$ (Pouliquen and Forterre, 2002; Poulain et al., 2023; Marbœuf et al., 2025).

Finally, to quantify the computational performance of the Landslide-Tsurrogate v1.0 framework for the Mayotte test case, the cost of the deterministic simulations required to construct the surrogate models using Multilayer HySEA is addressed. Each deterministic simulation required approximately 50 minutes of runtime when executed on 1 NVIDIA GPU A100 card (40GB) of the S-CAPAD/DANTE platform at IPGP, France. Because the simulations associated with the quadrature nodes are independent, the training dataset was generated using a parallel workflow in which multiple deterministic simulations were executed simultaneously. A total of 1035 simulations were required to construct and validate the surrogate models across the five landslide zones considered in this study. When distributed across 6 GPU cards, the total wall-clock time required to generate the full dataset was approximately 5.6 days, whereas the equivalent sequential runtime would be approximately 36 days. This parallelization strategy substantially reduces the computational burden associated with constructing the surrogate models and makes the framework tractable for large parameter studies or probabilistic hazard analyses involving multiple potential landslide sources.

4.2.2 Landslide volumes

In Multilayer HySEA, the initial moving mass is defined from two input grids: (1) the topo-bathymetry of Mayotte, and (2) the same topo-bathymetry modified by adding the initial failure volume. For each of the 1035 simulations, these two files are required.

To generate the failure surface automatically, we developed an algorithm that modifies the 10 m DEM of Mayotte based on two stochastic parameters: the along-isoline distance D, which determines the coordinates (x_c, y_c) of the failure center, and the target released volume V. The shape of the failure surface is controlled by three geometric parameters (l_1, l_2, r):

- l_1 and l_2 are the semi-axes of the ellipsoidal footprint along and across the local slope,
- rl_1 controls the maximum depth (thickness) of the failure.

These parameters define a paraboloid representing the sliding mass. Figure 5 shows different failure surfaces and sliding masses generated by the algorithm. Note that the shapes of the failure surfaces differ at each location and depend on the local bathymetry.

To match a prescribed volume V, the algorithm iteratively adjusts (l_1, l_2, r) by minimizing

$$f = \frac{|V - V_c|}{V},$$

where V_c is the computed volume from the generated surface. The minimization uses the *Nelder-Mead* method (*scipy.optimize.minimize*), with convergence tolerances of 0.001 on f and 0.01 on the parameters. The construction of the failure surface (illustrated in Figure 6 for the simulation 0 of the Mayotte South zone) follows these steps:

1. A square box $\mathcal{B}$, centered at (x_c, y_c), defines the local computation domain (Fig. 6a).
2. A plane $\mathcal{P}$ is fitted to the bathymetry inside $\mathcal{B}$ (Fig. 6c).
3. A paraboloid thickness $h_e(x, y)$, centered at (x_c, y_c), is created:

   $$h_e(x, y) = r\, l_1 \left[1 - \left(\frac{x'}{l_1}\right)^2 - \left(\frac{y'}{l_2}\right)^2\right],$$

   where (x', y') are translated and rotated coordinates (angle θ) aligned with the local slope direction (Figs. 6a, 6b and 6d).
4. The surface $\mathcal{S}$ (height z_s) is defined by the points $\mathbf{x}_s = \mathbf{x}_p - h_e\, \mathbf{n}$ where $\mathbf{x}_p$ are the points of the plane $\mathcal{P}$ and $\mathbf{n}$ its normal (Fig. 6e).
5. The initial bathymetry (height z_b) is locally lowered following the surface $\mathcal{S}$ to form the failure surface (height z_d), i.e. $z_d = \min(z_b, z_s)$ (Fig. 6f).
6. The removed volume is computed as

   $$V_c = \Delta x\, \Delta y \sum (z_b - z_d),$$

   where $\Delta x = \Delta y = 10$ m is the grid size.

The parameter bounds are set to $l_1, l_2 \in [40, 800]$ m and $r \in [0.05, 0.2]$, ensuring that the generated surface is smooth at 10 m resolution and not excessively deep (to avoid trapping material).

For computational efficiency, the resulting 10 m DEMs are later downsampled to 30 m resolution, with a volume discrepancy below 1.7%.

4.3 Steps 4 and 5: Surrogate Models and Evaluation

This section aims to illustrate how the Landslide-Tsurrogate v1.0 model operates and, hence, only the results of the surrogate models implemented along the shores of Petite Terre are presented hereafter. The corresponding results for Grande Terre are not discussed in the article but are provided via the user-friendly interfaces.

4.3.1 Locations

To provide PTF along all vulnerable Petite Terre coastal areas, the surrogate models are built pointwise (i.e., with no spatial correlation between the sites) at targeted depths of $1.0 - 1.5$ m outside of the inundation zone. These target locations are distributed along the Petite Terre coastline in areas where inundation maps are required for risk reduction, emergency preparedness, and evacuation planning. Practically, the spatial discretization used in the deterministic simulations leads to a total of $M = 211$ coastal target points (Fig. 7). To simplify the analysis, the surrogate models are numbered and divided into four areas: east coast (1–51), south-west coast (52–113), north-west coast (114–151 and 190–211), and lagoon (152–189). Practically, the surrogate models are built using the outputs of the 1035 deterministic tsunami simulations initiated from the five previously defined submarine landslide-prone zones (Piton, Piton Offshore, Piton South, Mayotte South, and Mayotte Offshore). For each landslide scenario, tsunami propagation is simulated, and relevant hazard metrics such as maximum tsunami elevation, maximum tsunami speed and time of arrival are extracted at the coastal target points.

This process enabled the construction of efficient surrogate models of tsunami hazards that capture the complex, non-linear relationships between submarine landslide source characteristics and coastal impacts along the Petite Terre coastline.

4.3.2 Evaluation

The convergence, accuracy and sensitivity of the 211 surrogate models built along the Petite Terre coastline are presented in Figures 8–10.

1. Convergence

   The convergence analysis of the 211 surrogate models built for Petite Terre (Fig. 8) provides important insights into the reliability of the generalized Polynomial Chaos Expansion (gPCE) methodology when applied to landslide-generated tsunami hazards. Across the three key tsunami variables—maximum elevation, maximum speed, and time of arrival—the results highlight both the strengths and the limitations of the surrogate modeling approach in capturing the complex dynamics of landslide-induced tsunamis in the Piton, Piton Offshore, Piton South, Mayotte South, and Mayotte Offshore zones.

For maximum elevation, the surrogate models display robust convergence across nearly all zones and all surrogate model locations. Normalized errors $\left\{e_p^m\right\}_{p=0}^{6}, \quad m \in 1, 2, \ldots, 211$ [Eq. (11)] decrease rapidly with increasing gPCE order p, with most locations m achieving normalized errors below 5%. This demonstrates the strong capacity of the surrogate models to approximate the spatial distribution of maximum wave heights, even in dynamically complex environments. The convergence of maximum speed is less uniform and somewhat slower. While some zones—particularly Piton and Piton Offshore—achieve acceptable error levels at higher gPCE orders, localized bands of moderate errors persist, particularly for surrogate models number 25 and 27 which are located in a protected bay. This heterogeneity may be attributed to small-scale variations in bathymetry and shoreline geometry that strongly influence flow velocities (e.g., via refraction and reflection). This uneven convergence indicates that while surrogate models provide a good first approximation of flow speed, site-specific refinements may be required for accurate hazard characterization in sensitive locations.

The time of arrival emerges as the most difficult variable to approximate using the surrogate models. With the exception of most of the east-coast locations (1-51) and the northern most locations of the north-west coast (189-211) for the three Piton zones and the Mayotte South zone (only east-coast), errors generally decrease more slowly compared to elevation and speed, with many surrogate model locations retaining high normalized errors (>60$\%$) even at the highest gPCE orders tested. This underscores the sensitivity of arrival time to small uncertainties in landslide initial conditions and wave propagation processes. Additionally, it should be noted that, in HySEA, the time of arrival is set to -1 when the wave height remains below $0.01\mathrm{m}$. For the construction of the surrogate models, realistic values were required; therefore, the -1 values were replaced with 30 minutes, corresponding to the full length of the simulation. While this substitution ensured consistency in the dataset, it may have affected the generation of surrogate models in locations minimally impacted by the tsunami waves (e.g., the west-coast and the lagoon). This suggests that while surrogate models can offer useful probabilistic estimates of arrival times, their predictive skill remains limited in certain contexts, necessitating caution if used for early-warning purposes.

Overall, although the convergence patterns are not uniformly perfect across all zones, all surrogate model locations and all variables, the results clearly demonstrate that the gPCE method achieves near-optimal performance by total order $p = 5$. The marginal improvement observed at total order $p = 6$ does not significantly enhance convergence, indicating that increasing the polynomial order beyond $p = 5$ provides diminishing returns. This finding underscores that a total order of $P_t = 5$ strikes a practical balance between computational efficiency and model accuracy, ensuring reliable surrogate model representations of tsunami hazards without incurring unnecessary complexity.

2. Accuracy

The scatter plots in Figure 9 assess the predictive skill or accuracy of the surrogate models built using the gPCE truncated at total order $P_t = 5$. Importantly, the surrogate results are compared against the 72 additional deterministic simulations needed to reach the total order $p = 6$. These deterministic simulations can be considered as independent as they are not used in building the surrogate models. This approach provides a robust validation of the surrogates by testing their ability to generalize beyond the training dataset.

For maximum elevation, the surrogate models demonstrate excellent predictive accuracy across all five zones. The data points cluster tightly around the 1:1 reference line, with minimal dispersion, confirming that elevation—a relatively smooth and integrated measure of tsunami intensity—is reliably reproduced by the surrogates even in unseen scenarios. The predictive performance for maximum speed is slightly less consistent, with more noticeable scatter around the reference line, particularly in Piton South and Mayotte South. This behavior is expected, as speed is a more sensitive and localized variable influenced by fine-scale bathymetric features and flow dynamics. Despite this, the surrogate models capture the overall trends well, with the densest regions of the scatter (Fig. 9, occurrences between 20 and 50, in red) still following the 1:1 relationship. For time of arrival, the agreement remains strong overall, though some divergence appears for Piton Offshore, Piton South and Mayotte Offshore. As explained previously, this variability may partly reflect the methodological choice to replace non-physical HySEA values (originally set to -1 for wave heights below $0.01\mathrm{m}$) with the full simulation length of 30 minutes. Such replacements could affect the accuracy at sites with negligible tsunami impact, yet the bulk of the results remain closely aligned with the deterministic benchmarks.

In summary, the comparison demonstrates that surrogate models built at total order $P_t = 5$ not only converge internally but also generalize effectively to independent simulations, providing reliable predictions of tsunami elevation, speed, and arrival time along the coastal areas of Petite Terre.

3. Sensitivity

Finally, Figure 10 presents the total sensitivity indices [Eq. (13)] of the surrogate models for maximum elevation, maximum speed, and time of arrival of landslide-generated tsunamis across the 211 Petite Terre locations. The decomposition highlights the relative importance of the three uncertain input parameters—landslide location, volume, and friction angle—within each of the five study zones.

Across all zones and most locations, landslide volume emerges as the dominant contributor to the output variance of the maximum elevation and flow speed, with its sensitivity index typically approaching or exceeding 0.75. This result is consistent with the established understanding that the magnitude of the failing mass strongly controls the intensity of landslide-generated tsunamis. Rather than representing a new physical insight, this outcome provides an important consistency check, confirming that the surrogate modeling framework correctly reproduces the expected physical sensitivities of the system.

In contrast, the landslide location shows a more variable influence depending on the region and the considered metric, reflecting the role of local bathymetric conditions and coastal geometry in modulating tsunami propagation and coastal impacts. In the Piton South and Mayotte South zones, location accounts for a substantial fraction of the sensitivity, especially for maximum speed and time of arrival, reflecting the strong influence of coastal geometry and bathymetric gradients on tsunami propagation and impact timing. The other zones exhibit a reduced location sensitivity, where volume largely governs the results except for the time of arrival.

Finally, The friction angle consistently contributes the least to the variance across all zones and metrics, with values generally below 0.1. Its role appears minor compared to volume and location, though it may still introduce subtle effects

on flow dynamics in specific sites. Interestingly, particularly for the speed and time of arrival, total sensitivities exceed 1, indicating non-linear interaction effects among the input parameters and demonstrating that location and volume combine to drive complex arrival-time variability.

Overall, the analysis demonstrates that while landslide volume is the primary control on tsunami hazard characteristics, the importance of location cannot be neglected, particularly in areas where coastal complexity modulates local hazard intensity. The friction angle, though secondary, still contributes non-negligibly in sensitive locations.

This evaluation (Figs. 8–10) highlights a consistent relationship between sensitivity structure, convergence, and accuracy of the surrogate models. For maximum elevation and speed, in zones where landslide volume dominates the response (e.g., Piton, Piton Offshore, Mayotte Offshore), the reduced effective dimensionality (i.e., the fact that the surrogate model results mostly depend of the landslide volume) leads to faster convergence (Fig. 8) and higher predictive accuracy when tested against independent simulations (Fig. 9). Conversely, in zones where the influence of location is stronger—particularly for speed and time of arrival in Piton South and Mayotte South—the surrogate models exhibit greater scatter around the 1:1 reference line (Fig. 9) and require higher orders to stabilize (Fig. 8), reflecting the increased complexity of the tsunami response. It is noteworthy that the Piton South and Mayotte South zones encompass the largest geographical extents within the study area and are characterized by pronounced slope heterogeneity. Such geomorphological variability can increase the complexity of the surrogate model parameter space, thereby hindering model convergence. Furthermore, the substantial distances separating potential landslide sources within these zones imply that tsunami arrival times will be more sensitive to the location of the initiating event. Thus, the observed sensitivity patterns directly explain why some variables, locations, and zones achieve better convergence and accuracy than others, reinforcing the physical interpretation of the surrogate model framework.

4.3.3 Step 7: Mayotte Landslide Probabilistic Tsunami Forecast (PTF)

An important aspect of the Mayotte submarine landslide test case is the emphasis on user-friendly interfaces providing faster-than-real-time forecasts (Fig. A3) that allow both researchers and decision-makers to interact with probabilistic tsunami hazard data efficiently.

However, the physical constraints of the Mayotte setting highlight the short delay between the landslide and the impact on the coast (a few minutes in some cases). Consequently, running deterministic simulations in real time, even with rapid computational tools, may be challenged by the speed of tsunami onset. Once trained using the deterministic simulations, the Landslide-Tsurrogate v1.0 framework enables rapid exploration of the stochastic parameter space describing potential submarine landslides. In this study, the probabilistic hazard associated with $1,000$ stochastic realizations of landslide parameters at 211 coastal locations along Petite Terre can be evaluated in less than 2 seconds on any laptop while the same number of deterministic simulations would run in about 6 days across 6 GPU cards. This capability makes it possible to characterize the statistical variability of tsunami impacts across a large number of plausible landslide scenarios, which would be computationally impractical using only full deterministic simulations.

The resulting probabilistic simulations (Fig. 11) indicate that tsunamis generated in the Piton zone can reach the airport and nearby settlements in less than three minutes for a significant subset of the explored scenarios. This result highlights the extremely short warning times associated with nearby submarine landslide sources and illustrates how the surrogate framework enables rapid evaluation of arrival-time distributions across a wide range of possible landslide parameters. Such short warning times strongly limit the feasibility of traditional evacuation strategies and emphasize the need for complementary adaptation measures. These may include vertical evacuation shelters, reinforced structures, pre-identified safe zones, clear evacuation protocols, and automated alert systems—strategies that have been studied in Mayotte (Leone et al., 2023) and implemented in other tsunami-prone regions such as Japan and Indonesia.

By integrating fast, accessible computational tools with proactive adaptation measures, Landslide-Tsurrogate not only advances scientific understanding but also provides actionable solutions for emergency preparedness. This combination ensures that hazard assessments translate directly into timely and informed decisions, ultimately enhancing community resilience in the face of submarine landslide tsunamis in Mayotte.

5 Limitations

While the Landslide-Tsurrogate v1.0 framework provides an efficient approach for propagating uncertainties in landslide-generated tsunami forecasts, several limitations should be acknowledged. These limitations arise from the assumptions made in the representation of landslide sources, the deterministic numerical simulations used to construct the surrogate models, and the surrogate modeling approach itself.

5.1 Surrogate modeling approach

The Landslide-Tsurrogate v1.0 framework relies on gPCE to approximate the response of the deterministic landslide–tsunami simulations across the stochastic parameter space. While gPCE provides an efficient and mathematically rigorous surrogate modeling approach, it also introduces specific limitations that must be considered.

First, the construction of a polynomial chaos surrogate requires that the quantity of interest be defined for all deterministic simulations corresponding to the quadrature nodes. In practice, this means that all model evaluations must return finite values. However, in tsunami simulations certain output variables—particularly those related to coastal inundation—may exhibit undefined or discontinuous behavior. For example, at locations where the tsunami does not reach the coastline for some parameter combinations, quantities such as maximum elevation, time of arrival or maximum speed may be zero or undefined, while they become positive for other combinations. These wet–dry transitions introduce discontinuities in the response surface that are difficult to represent accurately with polynomial expansions.

Second, polynomial surrogate models generally perform best when the system response varies smoothly with respect to the stochastic parameters. In the context of landslide-generated tsunamis, strongly nonlinear processes such as shoreline interaction, or abrupt changes in landslide mobility can introduce localized nonlinearities or threshold effects. These behaviors

may reduce the convergence rate of the polynomial expansion and lead to reduced surrogate accuracy unless sufficiently high polynomial orders are used.

Finally, inundation-related quantities of interest, such as maximum run-up or flooded area, are particularly challenging to represent within the gPCE framework because they depend on complex wetting–drying dynamics and may exhibit non-smooth spatial patterns. For this reason, the present implementation primarily focuses on tsunami wave characteristics at locations outside of the inundation zone, where the response is generally smoother and more amenable to polynomial approximation. Extending the framework to accurately represent inundation processes may require alternative surrogate modeling strategies or hybrid approaches combining gPCE with other statistical or machine learning methods.

5.2 Deterministic modeling strategy

The accuracy of the surrogate models ultimately depends on the ability of the underlying deterministic numerical model to represent the physical processes governing landslide-tsunami generation, propagation, and coastal interaction. The surrogate model does not introduce new physical information but rather approximates the response of the deterministic model across the stochastic parameter space. Consequently, any simplifications or inaccuracies in the deterministic model are directly inherited by the surrogate model.

In particular, the representation of submarine landslide dynamics may vary significantly depending on the assumptions made in the models (e.g., rigid-block motion, depth-averaged granular flow models, or fully coupled hydro-mechanical approaches). The suitability of a given model depends on the characteristics of the considered site, including bathymetric complexity, sediment properties, and the expected failure mechanisms. For some environments, simplified landslide parameterizations may provide reasonable approximations of tsunami generation, whereas other cases may require more advanced three-dimensional or multilayer models to capture complex flow behavior and landslide–water interactions. The present framework is therefore not tied to a specific deterministic model but relies on the user to select an appropriate model for the site under investigation.

Furthermore, in the Mayotte test case, submarine landslides are represented using simplified geometric parameterizations (e.g., paraboloid shapes for the initial mass released) and a limited number of stochastic variables describing their location, volume, and rheological properties. While this representation allows the dimensionality of the stochastic space to remain manageable, it necessarily simplifies the complex geometry and mechanical behavior of real submarine landslides.

Natural submarine slope failures may involve irregular geometries, heterogeneous sediment properties, progressive failure mechanisms, or multi-stage collapses. Such processes may not be fully captured by simplified parametric representations. As a result, the surrogate models should be interpreted as providing probabilistic estimates of tsunami forecasts for a range of idealized landslide scenarios rather than precise predictions of specific failure events.

5.3 Dimensionality of the stochastic parameter space

The efficiency of the surrogate modeling approach depends strongly on the number of stochastic variables used to describe the system. As the number of uncertain parameters increases, the number of deterministic simulations required to construct

accurate polynomial surrogate models grows rapidly. Although sparse quadrature methods substantially reduce the number of required simulations compared with full tensor grids, the approach may still become computationally demanding when the dimensionality of the stochastic space is large.

For this reason, the framework is best suited to applications where the dominant sources of uncertainty can be represented by a relatively small number of stochastic variables (up to 6). In practice, careful parameter selection and sensitivity analysis may be necessary to identify the most influential parameters and avoid unnecessary expansion of the stochastic space.

5.4 Representation of uncertainty

The Landslide-Tsurrogate framework represents uncertainty through prior probability distributions assigned to the stochastic variables describing landslide characteristics and environmental conditions. The definition of stochastic variables and their probability distributions are inherently site dependent. Bathymetric conditions, sediment characteristics, tectonic setting, and coastal morphology all influence landslide dynamics and tsunami propagation.

However, in many practical applications, the available geological and geotechnical information may be limited, and the choice of probability distributions may rely partly on expert judgment or regional analogues. Due to these limitations, the prior distributions of all stochastic variables are assumed uniform in the current version of the Landslide-Tsurrogate framework. As a result, the probabilistic outputs of the framework reflect both natural variability and epistemic uncertainty associated with limited knowledge of the system.

Future applications could benefit from improved constraints on landslide occurrence, sediment properties, and triggering mechanisms, which would allow more realistic parameter distributions to be defined. Consequently, future developments of the Landslide-Tsurrogate framework should include the possibility to build surrogate models with non-uniform prior distributions of the stochastic variables.

In summary, the Landslide-Tsurrogate framework should be viewed as a flexible methodological approach that can be adapted to different environments. Applying the framework to other sites requires careful consideration of the local geological setting, the selection of appropriate deterministic models, and the definition of stochastic variables that adequately represent the dominant uncertainties affecting landslide-generated tsunamis.

6 Conclusions

This study addresses the critical challenge of efficiently assessing tsunami hazards generated by submarine landslides, a phenomenon characterized by high physical complexity and considerable uncertainty. Traditional Probabilistic Tsunami Forecast (PTF) approaches are computationally expensive, limiting the ability to explore the full range of uncertain landslide parameters. To overcome this limitation, Landslide-Tsurrogate v1.0, a surrogate modeling framework designed to approximate tsunami maximum elevation, maximum speed and time of arrival as functions of submarine landslide characteristics, has been developed to provide fast and reliable predictions suitable for probabilistic analyses.

The key takeaways of the Landslide-Tsurrogate v1.0 model are:

1. Surrogate modeling efficiency for probabilistic assessment of landslide-generated tsunamis

   The Landslide-Tsurrogate framework combines high-fidelity numerical simulations with generalized Polynomial Chaos Expansion (gPCE) and uniform distributions of the submarine landslide parameters to construct surrogate models. The methodology incorporates a rigorous training and validation workflow, ensuring that the surrogate accurately captures the non-linear response of the tsunami system while drastically reducing computational costs compared to full deterministic numerical simulations. However, adopting this approach implies that many of the randomly generated submarine landslide scenarios may, in fact, lack the capacity to produce significant tsunamis in the area under investigation. This provides strong motivation to better constrain uncertainty distributions, that is, to more accurately define the prior distributions of the stochastic parameters using additional data. Such refinement can be achieved through Bayesian inference approaches (e.g., von Toussaint, 2011), or more generally via model calibration and data assimilation techniques. These aspects fall outside the scope of the present study and are left for future work.

2. Sensitivity Analysis

   Within the gPCE framework, performing sensitivity analysis is trivial and demonstrates that key landslide parameters—such as volume and geographical location—dominate tsunami maximum elevation, maximum speed and time of arrival, confirming both the physical relevance of the model and the ability of gPCE to reproduce complex dependencies.

3. Real-world application

   Application to the Mayotte submarine landslide case illustrates the practical utility of Landslide-Tsurrogate v1.0 for real-world PTF studies. Using landslide scenarios informed by geological and geophysical data, the surrogate models along Petite Terre efficiently generated probabilistic maximum elevation, maximum speed and time of arrival distributions and identified regions of elevated tsunami risk. Results highlight both the expected range of tsunami characteristics and the relative influence of individual landslide parameters on hazard outcomes, providing actionable insight for risk management and early warning considerations.

4. Scalability

   The Landslide-Tsurrogate v1.0 approach can be extended to multi-slide scenarios, higher-resolution bathymetry, and real-time hazard forecasting for enhanced coastal risk management.

Overall, Landslide-Tsurrogate v1.0 demonstrates that surrogate modeling is a powerful tool for bridging the gap between high-fidelity tsunami simulations and practical PTF applications. The framework allows extensive uncertainty quantification, sensitivity analysis, and scenario exploration while maintaining computational efficiency. Future work will focus on addressing missing values generated by the deterministic landslide–tsunami models (e.g., time of arrival set to -1 in HySEA) to enable the extension of the surrogate model implementation to the inundation zone and the production of accurate inundation maps essential for risk assessment. In addition, Landslide-Tsurrogate v1.0 will be expanded to incorporate alternative prior distributions of the stochastic variables beyond the uniform case, thereby enhancing its general applicability. By enabling rapid yet reliable

tsunami hazard assessment, Landslide-Tsurrogate v1.0 represents a valuable step toward improved coastal risk management in landslide-prone regions.

Code and data availability. The Landslide-Tsurrogate v1.0 model including the MATLAB and Python codes, the Jupyter Notebook User Manual, the MATLAB and Python GUIs for the Mayotte test case and the data associated with the Mayotte test case are published in Denamiel (2025), under Zenodo with a GNU General Public license v3.0 (https://zenodo.org/records/17519408).

Author contributions. C.D. developed the Landslide-Tsurrogate v1.0 MATLAB and Python codes, implemented the Jupyter Notebook User Manual and the MATLAB and Python GUIs. C.D. designed the study and contributed to the methodology and interpretation of results for the Mayotte test case. Alexis M. performed all the HySEA simulations and implemented the methodology to construct the failure surfaces (in collaboration with M.P. and A.L.) for the Mayotte test case. Anne M. designed the study, supervised the work, contributed to the methodology and interpretation of results for the Mayotte test case. A.L.F., M.P., A.L., E.F.-N. and M.C. contributed to the methodology and interpretation of results for the Mayotte test case. C.D. prepared the manuscript with contributions from all co-authors.

Competing interests. The authors declare that none of them have any competing interests.

Acknowledgements. The authors thank both anonymous reviewers for their careful reading of the manuscript and for their constructive and insightful comments. Numerical computations were partly performed on the S-CAPAD/DANTE platform, IPGP, France and thanks to the support of Geo-INQUIRE (HORIZON-INFRA-2021-SERV-01, number 101058518) through the SLAM project Transnational Access (TA2-531-3, HySEA - Earthquake and landslide generated tsunami simulations). Special thanks are given for the support of the European Centre for Medium-Range Weather Forecasts (ECMWF) staff and for the ECMWF computing and archive facilities used in this research under the special project: "Exploring the potential of uncertainty quantification and machine learning techniques to forecast rare extreme events". The authors acknowledge the support of the Agence Nationale de la Recherche under the France 2030 programme with the reference ANR-22-EXIR-0005, (ROM Project PEPR Risques, IRiMa). This work was supported by ChEESE-2P (EU-EuroHPC JU-101093038) and DT-GEO (HORIZON-INFRA-2021-TECH-01-01, number 101058129) projects. Manuel J. Castro Díaz and Enrique D. Fernández-Nieto acknowledge PID2022-137637NB-C21 and PID2022-137637NB-C22 funded by MCIN/AEI/10.13039/501100011033/ and ERDF "A way of making Europe".

Appendix A: Landslide-Tsurrogate v1.0: Technical Description

Landslide-Tsurrogate v1.0 is implemented in both Python and MATLAB to ensure accessibility and flexibility for a broader range of users within the scientific community. The same simple structure (Fig. A1) is used for both the Python and MATLAB codes. To run the model, three folders are required: `code`, `data`, and `results`.

The code folder contains the seven user-oriented programs that each produce a single output file. They should be executed sequentially following the workflow presented in Figure A2. While each program can run independently, it depends on input files generated by the preceding step. This modular structure is intentional, as certain steps—such as running deterministic simulations—require user intervention and can be time-consuming. Separating the programs in this way allows for greater flexibility and practicality in managing the workflow.

The data folder includes two MATLAB-format input files that the user must generate when running Landslide-Tsurrogate v1.0. The first, input_simus.mat, stores the outputs of the N deterministic simulations corresponding to the highest total order $P_t + 1$. It contains the maximum tsunami elevation zeta_max_surf[N,nx,ny], the maximum tsunami speed velo_max_surf[N,nx,ny], and the tsunami time of arrival time_max_surf[N,nx,ny], all defined over the simulation domain with spatial dimensions [nx,ny]. The second file, surrogate_model_locations.mat, stores the indices index_coast[M]—following MATLAB conventions in the example included in the distribution of the code—corresponding to the M locations where the surrogate models are to be constructed.

Finally, the MATLAB-format output files of the seven main programs are created in the results folder.

The sequential execution of the seven user-oriented programs and their associated outputs (illustrated in Fig. A2) corresponds to the first six core building blocks of the workflow while the implementation of the user-friendly interfaces is done outside of the main workflow and illustrated, for the Mayotte test case, in Figure A3.

1. Landslide_Tsurrogate_step_1 _user_input

   The user specifies the number r of stochastic variables, the definition intervals $[a_i, b_i] \quad i \in 1, 2, \ldots r$ for each stochastic variable, the maximum total order $P_t + 1$, and the quadrature rule (0 for GP, 1 for DGP). This program generates the file output_users.mat.

2. Landslide_Tsurrogate_step_2 _input_parameters

   This program creates the combinations of parameters that sample the definition intervals of the stochastic variables, according to the quadrature rule selected by the user and following the theoretical framework described in Section 2. It calls the main function surrogate_model_gauss_patterson_param, which constructs the final nested grids using the Heiss and Winschel algorithm (Heiss and Winschel, 2008).

   In practice, the multi-index set α is generated by the function get_seq, while the multivariate abscissa [Eq. (8)] are computed by the function gauss_patterson_nested_rule_nd, based on the univariate GP abscissa [Eq. (10)] provided by gauss_patterson_nested_rule_1d.

   The final parameter set is given by the unique combinations of these multivariate abscissa. These are saved in the file output_param.mat and used as inputs for the deterministic landslide-generated tsunami simulations.

3. Landslide_Tsurrogate_step_4 _format_simulations

   After the deterministic simulations are completed, the user must provide the tsunami hazard variables—maximum tsunami elevation, maximum tsunami speed, and time of arrival—in the specific format required by Landslide-Tsurrogate

v1.0. This program serves as a template for formatting those outputs and should be edited and adapted to match the structure and naming conventions of the user's simulation results. This step is essential to ensure compatibility with the subsequent surrogate modeling process and the results are saved in the file `output_model.mat`.

4. `Landslide_Tsurrogate_step_5 _coefficients`

   This program requires the inputs generated in the three previous steps to compute the deterministic coefficients of the gPCE expansion [Eqs. (6) and (7)]. It relies on the main function `surrogate_model_gauss_patterson_coeff`, which builds upon the same functions used in Step 2, with the addition of `norm2_legendre_polynomials` to compute the norms of the multivariate Legendre polynomials. This program produces the main output of Landslide-Tsurrogate v1.0 which is saved in the file `output_coeff.mat`.

5. `Landslide_Tsurrogate_step_6 _evaluation`

   This program also requires the inputs generated in all the previous steps. It is based on the main functions `surrogate_model_gau` and `surrogate_model_gauss_patterson _sensi`, and produces two essential output files used for the validation of the surrogate models: `output_evals.mat` and `output_sensi.mat`. Subsequently, the user can run the program `Landslide_Tsurrogate_step_6_bis _visualization` to explore the results and make informed decisions regarding the truncation of the gPCE expansion or the elimination of certain stochastic variables.

6. `Landslide_Tsurrogate_step_7_PTF`

   In this final step, the program uses inputs from Step 1 and Step 4 to compute the Probabilistic Tsunami Forecast (PTF) at the locations where surrogate models have been constructed. It relies on the function `surrogate_model_gauss_patterson_` which evaluates the surrogate models based on a user-defined input distribution of the stochastic variables. The resulting PTF data are stored in the file `output_PTF.mat`.

7. User-friendly interface

   The generation of user-friendly interfaces with MATLAB and Python is illustrated for the Mayotte test case (see Fig. A3). In MATLAB, the interface can be converted into a standalone executable, which eliminates the need for internet access or complex software installations. This ensures that users in remote or resource-limited environments can run simulations and visualize results immediately, making the tool practical for real-world decision-making scenarios. In Python, the user-friendly interface leverages Jupyter Notebooks and Voilà (https://voila.readthedocs.io/en/stable/index.html; last visited 31/10/2025) to offer the possibility to provide interactive web applications. Jupyter Notebooks are widely used in scientific computing to combine code, documentation, and visualization in a single environment. With Voilà, these notebooks can be converted into fully interactive web applications, allowing users to adjust input parameters, run simulations, and visualize results dynamically—all through a standard web browser. Importantly, both the MATLAB and

Python approaches require no programming knowledge from the end user: the complexity of the model is fully hidden behind intuitive controls, while outputs such as probability density functions, cumulative distributions, and hazard maps are immediately accessible. These design choices illustrate the philosophy of Landslide-Tsurrogate: creating operational products from advanced computational research, ensuring that Probabilistic Tsunami Forecasts (PTF) can be used by emergency planners, local authorities, and other decision-makers.

To summarize, Landslide-Tsurrogate v1.0 provides a modular and transparent workflow for constructing and evaluating surrogate models tailored to landslide-generated tsunami hazards. Each user-oriented program is designed to function independently while maintaining a logical sequence that ensures reproducibility and flexibility. From the initial specification of stochastic inputs to the final generation of PTF outputs, the tool incorporates rigorous mathematical foundations and practical considerations to support a wide range of scientific applications. By offering both Python and MATLAB implementations and supporting user customization, Landslide-Tsurrogate v1.0 aims to be accessible to diverse users while maintaining high standards of technical robustness. This flexibility is exemplified in the next section through its application to the Mayotte test case, which demonstrates the end-to-end capabilities of the model in a real-world context.

Appendix B: Multilayer HySEA Model

The deterministic simulations for the Mayotte case study were performed using the Multilayer HySEA (Hyperbolic Systems and Efficient Algorithms) model, which belongs to a broader class of long-wave numerical models designed to simulate tsunamis generated by submarine landslides (Macías et al., 2021a, b). Earlier one-layer configurations of HySEA have been extensively validated against laboratory benchmarks (e.g., Poulain et al., 2023) as well as real-world tsunami events at field scale (e.g., Linares-Sánchez et al., 2011; Gonzalez-Vida et al., 2019). The HySEA modeling framework is formulated within the shallow-water approximation and explicitly represents both the motion of the granular landslide and the hydrodynamics of the water column, including their dynamic interaction through a frictional coupling at the landslide–water interface (Macías et al., 2021a, b). In the one-layer configuration, the water motion is described using depth-averaged non-hydrostatic (dispersive) equations, which can be written as follows:

$$
\begin{cases}
\partial_t h + \partial_x (h u) + \partial_y (h v) = 0 \\
\partial_t (h u) + \partial_x (h u^2 + \dfrac{1}{2} g h^2 + h p) + \partial_y (h u v) \\
\qquad\qquad = (g h + p)\, \partial_x (H - h_s) - S_x \\
\partial_t (h v) + \partial_x (h u v) + \partial_y (h v^2 + \dfrac{1}{2} g h^2 + h p) \\
\qquad\qquad = (g h + p)\, \partial_y (H - h_s) - S_y \\
\partial_t (h w) + \partial_x (h w u) + \partial_y (h w v) = p \\
\nabla \cdot \mathbf{u} + \dfrac{w_\eta - w_0}{h} = 0.
\end{cases}
\tag{B1}
$$

The model variables include the water surface height h, the horizontal water velocity vector $\mathbf{u} = (u\ v)$, the vertical water velocity w, the non-hydrostatic pressure p. The gravity acceleration is denoted by g. The last equation comes from the incompressibility condition and involves the vertical velocity at the free surface w_η and at the bottom w_0. At the free surface, the non-hydrostatic pressure is prescribed to vanish. The term $\mathbf{S} = (S_x\ S_y)$ in the momentum equations include all the dissipative terms (see Marbœuf et al. (2025) for a detailed formulation of the model).

Accounting for more vertical layers than the one-layer model described by [Eq. (B1)] improves the representation of the velocity field (Fernández-Nieto et al., 2016; Garres-Díaz et al., 2020; Escalante et al., 2023; Marbœuf et al., 2025) and makes it possible to better capture dispersive effects (Macías et al., 2021a). We use here 4 layers as the best compromise between accuracy and computational cost to simulate landslide-induced tsunamis in Mayotte (Marbœuf et al., 2025).

Note that the model does not differentiate between offshore and inland domains, and the propagation of flow over land is handled without introducing additional models, apart from the standard wet–dry treatment applied at shallow water depth (Macías et al., 2021a; Castro et al., 2005). No other model—such as a breaking wave or a vertical viscosity term (Roeber et al., 2010; Scala et al., 2024)—is applied.

A Savage-Hutter model (Savage and Hutter, 1989) describes the landslide motion and complements the system [Eq. (B1)] as follows:

$$
\begin{cases}
\partial_t h_s + \partial_x (h_s u_s) + \partial_y (h_s v_s) = 0 \\
\partial_t (h_s u_s) \\
+ \partial_x (h_s u_s^2 + \frac{1}{2}(1-r) g h_s^2) + \partial_y (h_s u_s v_s) \\
\qquad = (1-r)\, g\, h_s\, \partial_x H - S_{s,x} \\
\partial_t (h_s v_s) \\
+ \partial_x (h_s u_s v_s) + \partial_y (h_s v_s^2 + \frac{1}{2}(1-r) g h_s^2) \\
\qquad = (1-r)\, g\, h_s\, \partial_y H - S_{s,y},
\end{cases}
\tag{B2}
$$

where the unknowns are the landslide thickness h_s and the horizontal velocity vector $\mathbf{u}_s = (u_s\ v_s)$. The density ratio r is defined as $r = \rho_f / \rho_s$ where ρ_f is the fluid density and ρ_s the grain phase density. The system [Eq. (B2)] differs from [Eq. (B1)] through two main features: (i) the reduced gravity term $(1-r)\, g$ taking into account buoyancy effects when the landslide is underwater, and (ii) the absence of dispersive effects (non-hydrostatic pressure and vertical velocity). The terms $\mathbf{S}_s = (S_{s,x}\ S_{s,y})$ are the friction forces where the friction coefficient $\mu(I)$ appears (Pouliquen and Forterre, 2002; Jop et al., 2006; Mangeney et al., 2007). The coefficient $\mu(I)$ is given by

$$
\mu(I) = \tan(\delta) + \frac{\tan(\delta_2) - \tan(\delta)}{1 + \frac{I_0}{I}}. \tag{B3}
$$

The inertial number I is chosen here as in Martin et al. (2023)

$$
I = \frac{\frac{5}{2} \|\boldsymbol{u}_s\|\, d}{h_s \sqrt{g\,(1-r)\, h_s\, \Phi}} \tag{B4}
$$

where the empirical value of the constant $I_0 = 0.279$ comes from Jop et al. (2006); Martin et al. (2023). The friction angles δ and δ_2, the grain diameter d and the solid volume fraction Φ are specified in subsection 4.2.1.

A complete description of the multilayer HySEA model and its implementation can be found in Castro et al. (2011); Escalante et al. (2018); Macías et al. (2021a, b) and Marbœuf et al. (2025).

References

Abramowitz, M. and Stegun, I. A.: Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables, Courier Corporation, ISBN 978-0-486-61272-0, 1965.

Brecht, R., Cardoso-Bihlo, E., and Bihlo, A.: Physics-Informed Neural Networks for Tsunami Inundation Modeling, Journal of Computational Physics, 536, 114 066, https://doi.org/10.1016/j.jcp.2025.114066, 2025.

Bulthuis, K., Arnst, M., Sun, S., and Pattyn, F.: Uncertainty Quantification of the Multi-Centennial Response of the Antarctic Ice Sheet to Climate Change, The Cryosphere, 13, 1349–1380, https://doi.org/10.5194/tc-13-1349-2019, 2019.

Castro, M. J., Ferreiro, A. M. F., García-Rodríguez, J. A., González-Vida, J. M., MacíAs, J., Parés, C., and Vázquez-Cendón, M. E.: The Numerical Treatment of Wet/Dry Fronts in Shallow Flows: Application to One-Layer and Two-Layer Systems, Mathematical and Computer Modelling, 42, 419–439, https://doi.org/10.1016/J.MCM.2004.01.016, 2005.

Castro, M. J., Ortega, S., de la Asunción, M., Mantas, J. M., and Gallardo, J. M.: GPU Computing for Shallow Water Flow Simulation Based on Finite Volume Schemes, Comptes Rendus Mécanique, 339, 165–184, https://doi.org/10.1016/J.CRME.2010.12.004, 2011.

Cesca, S., Letort, J., Razafindrakoto, H. N. T., Heimann, S., Rivalta, E., Isken, M. P., Nikkhoo, M., Passarelli, L., Petersen, G. M., Cotton, F., and Dahm, T.: Drainage of a deep magma reservoir near Mayotte inferred from seismicity and deformation., Nat. Geosci., 13, 87–93, https://doi.org/https://doi.org/10.1038/s41561-019-0505-5, 2020.

Clenshaw, C. W. and Curtis, A. R.: A Method for Numerical Integration on an Automatic Computer, Numerische Mathematik, 2, 197–205, https://doi.org/10.1007/BF01386223, 1960.

Constantine, P. G., Eldred, M. S., and Phipps, E. T.: Sparse Pseudospectral Approximation Method, Computer Methods in Applied Mechanics and Engineering, 229–232, 1–12, https://doi.org/10.1016/j.cma.2012.03.019, 2012.

Crestaux, T., Le Maı^tre, O., and Martinez, J.-M.: Polynomial Chaos Expansion for Sensitivity Analysis, Reliability Engineering & System Safety, 94, 1161–1172, https://doi.org/10.1016/j.ress.2008.10.008, 2009.

Denamiel, C.: Landslide-Tsurrogate-V1.0 (GMD-v1.0), [software] [data], Zenodo, https://doi.org/10.5281/zenodo.17519408, 2025.

Denamiel, C., Šepić, J., Huan, X., Bolzer, C., and Vilibić, I.: Stochastic Surrogate Model for Meteotsunami Early Warning System in the Eastern Adriatic Sea, Journal of Geophysical Research: Oceans, 124, 8485–8499, https://doi.org/10.1029/2019JC015574, 2019.

Denamiel, C., Huan, X., Šepić, J., and Vilibić, I.: Uncertainty Propagation Using Polynomial Chaos Expansions for Extreme Sea Level Hazard Assessment: The Case of the Eastern Adriatic Meteotsunamis, Journal of Physical Oceanography, 50, 1005–1021, https://doi.org/10.1175/JPO-D-19-0147.1, 2020.

Denamiel, C., Huan, X., and Vilibić, I.: Conceptual Design of Extreme Sea-Level Early Warning Systems Based on Uncertainty Quantification and Engineering Optimization Methods, Frontiers in Marine Science, 8, https://doi.org/10.3389/fmars.2021.650279, 2021.

Ernst, O. G., Mugler, A., Starkloff, H.-J., and Ullmann, E.: On the Convergence of Generalized Polynomial Chaos Expansions, ESAIM: Mathematical Modelling and Numerical Analysis, 46, 317–339, https://doi.org/10.1051/m2an/2011045, 2012.

Escalante, C., de Luna, T. M., and Castro, M. J.: Non-Hydrostatic Pressure Shallow Flows: GPU Implementation Using Finite Volume and Finite Difference Scheme, Applied Mathematics and Computation, 338, 631–659, https://doi.org/10.1016/J.AMC.2018.06.035, 2018.

Escalante, C., Fernández-Nieto, E. D., Garres-Díaz, J., Morales de Luna, T., and Penel, Y.: Non-Hydrostatic Layer-Averaged Approximation of Euler System with Enhanced Dispersion Properties, Computational and Applied Mathematics, 42, 177, https://doi.org/10.1007/s40314-023-02309-7, 2023.

Fernández-Nieto, E. D., Garres-Díaz, J., Mangeney, A., and Narbona-Reina, G.: A Multilayer Shallow Model for Dry Granular Flows with the -Rheology: Application to Granular Collapse on Erodible Beds, Journal of Fluid Mechanics, 798, 643–681, https://doi.org/10.1017/JFM.2016.333, 2016.

Feuillet, N., Jorry, S., Crawford, W. C., Deplus, C., Thinon, I., Jacques, E., Saurel, J. M., Lemoine, A., Paquet, F., Satriano, C., Aiken, C., Foix, O., Kowalski, P., Laurent, A., Rinnert, E., Cathalot, C., Donval, J. P., Guyader, V., Gaillot, A., Scalabrin, C., Moreira, M., Peltier, A., Beauducel, F., Grandin, R., Ballu, V., Daniel, R., Pelleau, P., Gomez, J., Besançon, S., Geli, L., Bernard, P., Bachelery, P., Fouquet, Y., Bertil, D., Lemarchand, A., and der Woerd, J. V.: Birth of a Large Volcanic Edifice Offshore Mayotte via Lithosphere-Scale Dyke Intrusion, Nature Geoscience 2021 14:10, 14, 787–795, https://doi.org/10.1038/s41561-021-00809-x, 2021.

Formaggia, L., Guadagnini, A., Imperiali, I., Lever, V., Porta, G., Riva, M., Scotti, A., and Tamellini, L.: Global Sensitivity Analysis through Polynomial Chaos Expansion of a Basin-Scale Geochemical Compaction Model, Computational Geosciences, 17, 25–42, https://doi.org/10.1007/s10596-012-9311-5, 2013.

Fornberg, B. and Sloan, D. M.: A Review of Pseudospectral Methods for Solving Partial Differential Equations, Acta Numerica, 3, 203–267, https://doi.org/10.1017/S0962492900002440, 1994.

Garres-Díaz, J., Bouchut, F., Fernández-Nieto, E. D., Mangeney, A., and Narbona-Reina, G.: Multilayer Models for Shallow Two-Phase Debris Flows with Dilatancy Effects, Journal of Computational Physics, 419, 109 699, https://doi.org/10.1016/J.JCP.2020.109699, 2020.

Gerstner, T. and Griebel, M.: Numerical Integration Using Sparse Grids, Numerical Algorithms, 18, 209–232, https://doi.org/10.1023/A:1019129717644, 1998.

Giraldi, L., Le Maître, O. P., Mandli, K. T., Dawson, C. N., Hoteit, I., and Knio, O. M.: Bayesian Inference of Earthquake Parameters from Buoy Data Using a Polynomial Chaos-Based Surrogate, Computational Geosciences, 21, 683–699, https://doi.org/10.1007/s10596-017-9646-z, 2017.

Gonzalez-Vida, J. M., Macías, J., Castro, M. J., Sanchez-Linares, C., Asuncion, M. D. L., Ortega-Acosta, S., and Arcas, D.: The Lituya Bay Landslide-Generated Mega-Tsunami - Numerical Simulation and Sensitivity Analysis, Natural Hazards and Earth System Sciences, 19, 369–388, https://doi.org/10.5194/NHESS-19-369-2019, 2019.

Guillas, S., Sarri, A., Day, S. J., Liu, X., and Dias, F.: Functional Emulation of High Resolution Tsunami Modelling over Cascadia, The Annals of Applied Statistics, 12, 2023–2053, https://doi.org/10.1214/18-AOAS1142, 2018.

Heiss, F. and Winschel, V.: Likelihood Approximation by Numerical Integration on Sparse Grids, Journal of Econometrics, 144, 62–80, https://doi.org/10.1016/j.jeconom.2007.12.004, 2008.

Holtz, M.: Sparse Grid Quadrature in High Dimensions with Applications in Finance and Insurance, vol. 77 of *Lecture Notes in Computational Science and Engineering*, Springer, Berlin, Heidelberg, ISBN 978-3-642-16003-5 978-3-642-16004-2, https://doi.org/10.1007/978-3-642-16004-2, 2011.

Jaynes, E. T.: Information Theory and Statistical Mechanics, Physical Review, 106, 620–630, https://doi.org/10.1103/PhysRev.106.620, 1957.

Jop, P., Forterre, Y., and Pouliquen, O.: A Constitutive Law for Dense Granular Flows, Nature 2006 441:7094, 441, 727–730, https://doi.org/10.1038/nature04801, 2006.

Knio, O. M. and Le Maître, O. P.: Uncertainty Propagation in CFD Using Polynomial Chaos Decomposition, Fluid Dynamics Research, 38, 616–640, https://doi.org/10.1016/j.fluiddyn.2005.12.003, 2006.

Lemoine, A., Pedreros, R., Filippini, A. G., Paquet, F., Thinon, I., Legendre, Y., Mompelat, P. J.-M., Arpaia, L., and Boulhaya, F.: Scénarios d'impact de Tsunamis Pour Mayotte, Final report RP-69869-FR, 2020.

Leone, F., Gherardi, M., Péroche, M., Lagahé, E., Aumond, P., Siliezar Montoya, J., Idaroussi Tsima, F., Poulain, P., Le Friant, A., Mangeney, A., Hachim Mogne, S., and Roudier, V.: Mayotte se prépare au risque tsunami: modélisations, alerte, évacuation, sensibilisation, EchoGéo, 64, https://doi.org/10.4000/echogeo.25078, 2023.

Linares-Sánchez, C., Castro-Diaz, M. J., and Gonzalez-Vida, J. M.: Simulación Numérica de Tsunamis Generados Por Avalanchas Submarinas: Aplicación al Caso de Lituya-Bay, Ph.D. thesis, University of Malaga, 2011.

Liu, P. L.-F., Woo, S.-B., and Cho, Y.-S.: Computer Programs for Tsunami Propagation and Inundation, Cornell University, 25, 49, 1998.

Løvholt, F., Harbitz, C. B., and Haugen, K. B.: A Parametric Study of Tsunamis Generated by Submarine Slides in the Ormen Lange/Storegga Area off Western Norway, in: Ormen Lange–an Integrated Study for Safe Field Development in the Storegga Submarine Area, edited by Solheim, A., Bryn, P., Berg, K., Sejrup, H. P., and Mienert, J., pp. 219–231, Elsevier, Oxford, ISBN 978-0-08-044694-3, https://doi.org/10.1016/B978-0-08-044694-3.50023-8, 2005.

Løvholt, F., Pedersen, G., Harbitz, C. B., Glimsdal, S., and Kim, J.: On the Characteristics of Landslide Tsunamis, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 373, 20140 376, https://doi.org/10.1098/rsta.2014.0376, 2015.

Løvholt, F., Glimsdal, S., and Harbitz, C. B.: On the Landslide Tsunami Uncertainty and Hazard, Landslides, 17, 2301–2315, https://doi.org/10.1007/s10346-020-01429-z, 2020.

Luo, Y., Chen, W., Su, Z., Shi, X., Luo, J., Qu, X., Chen, Z., and Lin, Y.: Deep Learning Network for NMR Spectra Reconstruction in Time-Frequency Domain and Quality Assessment, Nature Communications, 16, 2342, https://doi.org/10.1038/s41467-025-57721-w, 2025.

Lynett, P. and Liu, P. L.-F.: A Two-Layer Approach to Wave Modelling, Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, 460, 2637–2669, https://doi.org/10.1098/rspa.2004.1305, 2004.

Lynett, P. J., Wu, T.-R., and Liu, P. L. F.: Modeling Wave Runup with Depth-Integrated Equations, Coastal Engineering, 46, 89–107, https://doi.org/10.1016/S0378-3839(02)00043-1, 2002.

Ma, G., Shi, F., and Kirby, J. T.: Shock-Capturing Non-Hydrostatic Model for Fully Dispersive Surface Wave Processes, Ocean Modelling, 43–44, 22–35, https://doi.org/10.1016/j.ocemod.2011.12.002, 2012.

Macías, J., Escalante, C., and Castro, M. J.: Multilayer-HySEA Model Validation for Landslide-Generated Tsunamis-Part 1: Rigid Slides, Natural Hazards and Earth System Sciences, 21, https://doi.org/10.5194/nhess-21-775-2021, 2021a.

Macías, J., Escalante, C., and Castro, M. J.: Multilayer-HySEA Model Validation for Landslide-Generated Tsunamis-Part 2: Granular Slides, Natural Hazards and Earth System Sciences, 21, https://doi.org/10.5194/nhess-21-791-2021, 2021b.

Mangeney, A., Bouchut, F., Thomas, N., Vilotte, J. P., and Bristeau, M. O.: Numerical Modeling of Self-Channeling Granular Flows and of Their Levee-Channel Deposits, Journal of Geophysical Research: Earth Surface, 112, 2017, https://doi.org/10.1029/2006JF000469, 2007.

Marbœuf, A., Mangeney, A., Le Friant, A., Castro, M. J., Fernández-Nieto, E., Lucas, A., Poulain, P., Moatty, A., Silver, M., Pedreros, R., Lemoine, A., and Løvholt, F.: Rheology and Bathymetry Effects Captured by a Multilayer Landslide-Tsunami Model, Geophysical Journal International, 242, ggaf232, https://doi.org/10.1093/gji/ggaf232, 2025.

Martin, H. A., Peruzzetto, M., Viroulet, S., Mangeney, A., Lagrée, P. Y., Popinet, S., Maury, B., Lefebvre-Lepot, A., Maday, Y., and Bouchut, F.: Numerical Simulations of Granular Dam Break: Comparison between Discrete Element, Navier-Stokes, and Thin-Layer Models, Physical Review E, 108, 054 902, https://doi.org/10.1103/PHYSREVE.108.054902, 2023.

Mercury, N., Lemoine, A., Doubre, C., Bertil, D., Woerd, J. V. D., Hoste-Colomer, R., and Battaglia, J.: Onset of a Submarine Eruption East of Mayotte, Comoros Archipelago: The First Ten Months Seismicity of the Seismo-Volcanic Sequence (2018–2019), Comptes Rendus - Geoscience, 354, 105–136, https://doi.org/10.5802/CRGEOS.191/FILE/ATTACH/CRGEOS-191-SUPPL.PDF, 2022.

Najm, H. N., Debusschere, B. J., Marzouk, Y. M., Widmer, S., and Le Maître, O. P.: Uncertainty Quantification in Chemical Systems, International Journal for Numerical Methods in Engineering, 80, 789–814, https://doi.org/10.1002/nme.2551, 2009.

Patterson, T. N. L.: The Optimum Addition of Points to Quadrature Formulae, Mathematics of Computation, 22, 847–856, https://doi.org/10.1090/S0025-5718-68-99866-9, 1968.

Poulain, P., Friant, A. L., Pedreros, R., Mangeney, A., Filippini, A. G., Grandjean, G., Lemoine, A., Fernández-Nieto, E. D., Diaz, M. J. C., and Peruzzetto, M.: Numerical Simulation of Submarine Landslides and Generated Tsunamis: Application to the on-Going Mayotte Seismo-Volcanic Crisis, Comptes Rendus. Géoscience, 354, 361–390, https://doi.org/10.5802/crgeos.138, 2022.

Poulain, P., Friant, A. L., Mangeney, A., Viroulet, S., Fernandez-Nieto, E., Diaz, M. C., Peruzzetto, M., Grandjean, G., Bouchut, F., Pedreros, R., and Komorowski, J. C.: Performance and Limits of a Shallow-Water Model for Landslide-Generated Tsunamis: From Laboratory Experiments to Simulations of Flank Collapses at Montagne Pelée (Martinique), Geophysical Journal International, 233, 796–825, https://doi.org/10.1093/GJI/GGAC482, 2023.

Pouliquen, O. and Forterre, Y.: Friction Law for Dense Granular Flows: Application to the Motion of a Mass down a Rough Inclined Plane, Journal of Fluid Mechanics, 453, 133–151, https://doi.org/10.1017/S0022112001006796, 2002.

Rasmussen, C. E. and Williams, C. K. I.: Gaussian Processes for Machine Learning, The MIT Press, ISBN 978-0-262-25683-4, https://doi.org/10.7551/mitpress/3206.001.0001, 2005.

Roeber, V., Cheung, K. F., and Kobayashi, M. H.: Shock-Capturing Boussinesq-type Model for Nearshore Wave Processes, Coastal Engineering, 57, 407–423, https://doi.org/10.1016/J.COASTALENG.2009.11.007, 2010.

Roger, J.: Potential Tsunami Hazard Related to the Seismic Activity East of Mayotte Island, Comoros Archipelago, Science of Tsunami Hazards, 38, 2019.

Roger, J. H. M., Bull, S., Watson, S. J., Mueller, C., Hillman, J. I. T., Wolter, A., Lamarche, G., Power, W., Lane, E., Woelz, S., and Davidson, S.: A Review of Approaches for Submarine Landslide-Tsunami Hazard Identification and Assessment, Marine and Petroleum Geology, 162, 106 729, https://doi.org/10.1016/J.MARPETGEO.2024.106729, 2024.

Salmanidou, D. M., Beck, J., Pazak, P., and Guillas, S.: Probabilistic, High-Resolution Tsunami Predictions in Northern Cascadia by Exploiting Sequential Design for Efficient Emulation, Natural Hazards and Earth System Sciences, 21, 3789–3807, https://doi.org/10.5194/nhess-21-3789-2021, 2021.

Saltelli, A.: Sensitivity Analysis for Importance Assessment, Risk Analysis, 22, 579–590, https://doi.org/10.1111/0272-4332.00040, 2002.

Savage, S. B. and Hutter, K.: The Motion of a Finite Mass of Granular Material down a Rough Incline, Journal of Fluid Mechanics, 199, 177–215, https://doi.org/10.1017/S0022112089000340, 1989.

Scala, A., Lorito, S., Sánchez, C. E., Romano, F., Festa, G., Abbate, A., Bayraktar, H. B., Castro, M. J., Macías, J., and Gonzalez-Vida, J. M.: Assessing the Optimal Tsunami Inundation Modeling Strategy for Large Earthquakes in Subduction Zones, Journal of Geophysical Research: Oceans, 129, e2024JC020 941, https://doi.org/10.1029/2024JC020941, 2024.

Smolyak, S.: Quadrature and Interpolation Formulas for Tensor Products of Certain Classes of Functions, in: Soviet Math. Dokl., vol. 4, pp. 240–243, 1963.

Sobol′, I. M.: Global Sensitivity Indices for Nonlinear Mathematical Models and Their Monte Carlo Estimates, Mathematics and Computers in Simulation, 55, 271–280, https://doi.org/10.1016/S0378-4754(00)00270-6, 2001.

Soize, C. and Ghanem, R.: Physical Systems with Random Uncertainties: Chaos Representations with Arbitrary Probability Measure, SIAM Journal on Scientific Computing, 26, 395–410, https://doi.org/10.1137/S1064827503424505, 2004.

Sraj, I., Mandli, K. T., Knio, O. M., Dawson, C. N., and Hoteit, I.: Uncertainty Quantification and Inference of Manning's Friction Coefficients Using DART Buoy Data during the Tōhoku Tsunami, Ocean Modelling, 83, 82–97, https://doi.org/10.1016/j.ocemod.2014.09.001, 2014.

Sudret, B.: Global Sensitivity Analysis Using Polynomial Chaos Expansions, Reliability Engineering & System Safety, 93, 964–979, https://doi.org/10.1016/j.ress.2007.04.002, 2008.

Sultan, N., Jouet, G., Riboulot, V., Terzariol, M., Garziglia, S., Cattaneo, A., Giraudeau, J., and Jorry, S.: Sea-level fluctuations control the distribution of highly liquefaction-prone layers on volcanic-carbonate slopes, Geology, 51, 402–407, https://doi.org/10.1130/G50785.1, 2023.

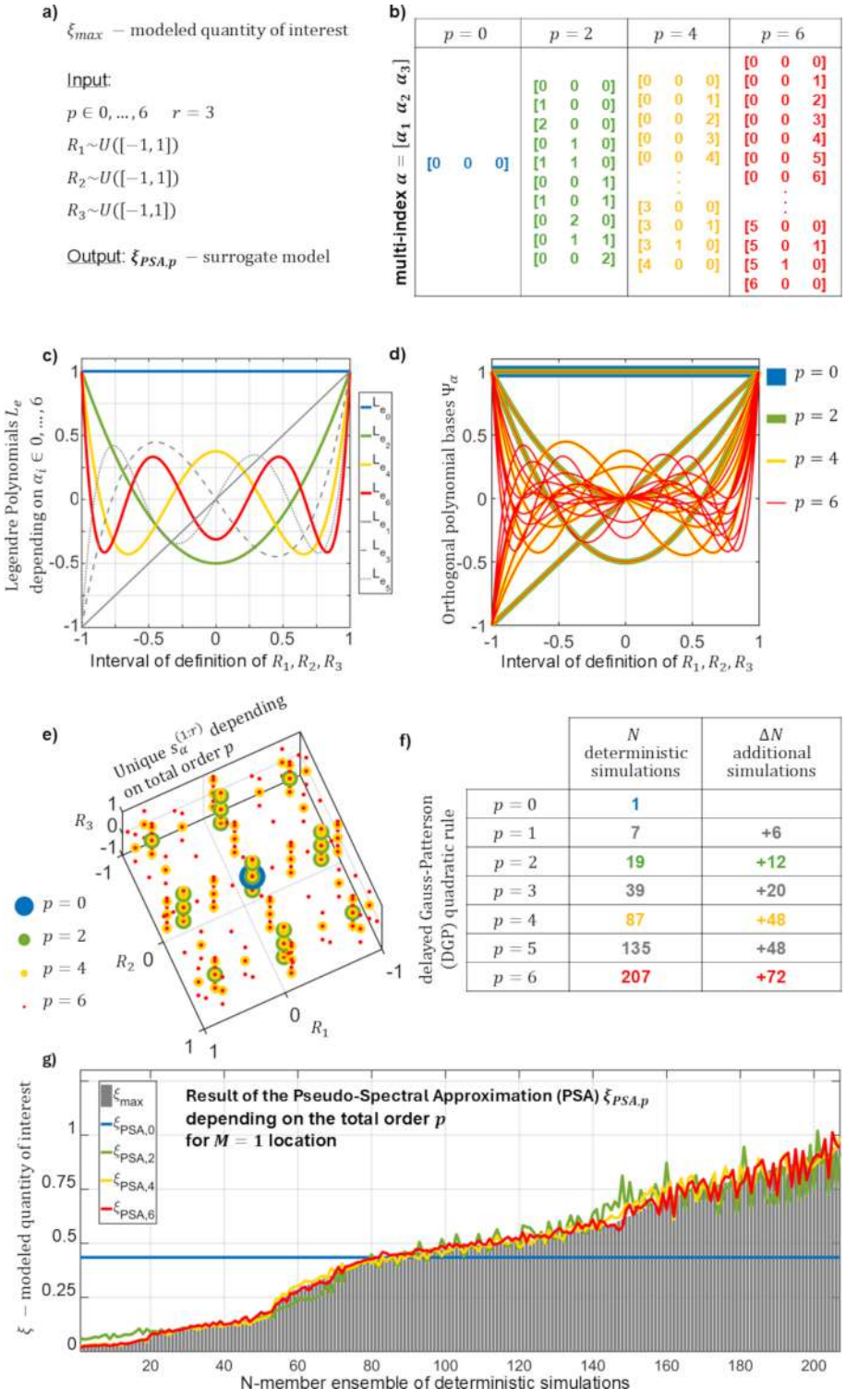
Trefethen, L. N.: Is Gauss Quadrature Better than Clenshaw–Curtis?, SIAM Review, 50, 67–87, https://doi.org/10.1137/060659831, 2008.

von Toussaint, U.: Bayesian Inference in Physics, Reviews of Modern Physics, 83, 943–999, https://doi.org/10.1103/RevModPhys.83.943, 2011.

Wang, M., Wan, Z., and Huang, Q.: A New Uncertain Analysis Method for the Prediction of Acoustic Field with Random and Interval Parameters, Shock and Vibration, 2016, 3693 262, https://doi.org/10.1155/2016/3693262, 2016.

Watts, P., Grilli, S. T., Kirby, J. T., Fryer, G. J., and Tappin, D. R.: Landslide Tsunami Case Studies Using a Boussinesq Model and a Fully Nonlinear Tsunami Generation Model, Natural Hazards and Earth System Sciences, 3, 391–402, https://doi.org/10.5194/NHESS-3-391-2003, 2003.

Xiu, D. and Karniadakis, G. E.: The Wiener–Askey Polynomial Chaos for Stochastic Differential Equations, SIAM Journal on Scientific Computing, 24, 619–644, https://doi.org/10.1137/S1064827501387826, 2002.



	N deterministic simulations	ΔN additional simulations
$p = 0$	1	
$p = 1$	7	+6
$p = 2$	19	+12
$p = 3$	39	+20
$p = 4$	87	+48
$p = 5$	135	+48
$p = 6$	207	+72

Figure 2. Schematic overview of the main mathematical concepts used in Landslide-Tsurrogate v1.0: (a) definition of the problem including quantity of interest, stochastic variables and total order of the gPCE; (b) multi-indices, (c) Legendre polynomials, (d) orthogonal polynomial bases, (e) sets of abscissas and (f) number of deterministic simulations needed, depending on the total order of the gPCE; (g) comparison of the quantity of interest with the surrogate model results depending on the total order of the gPCE.

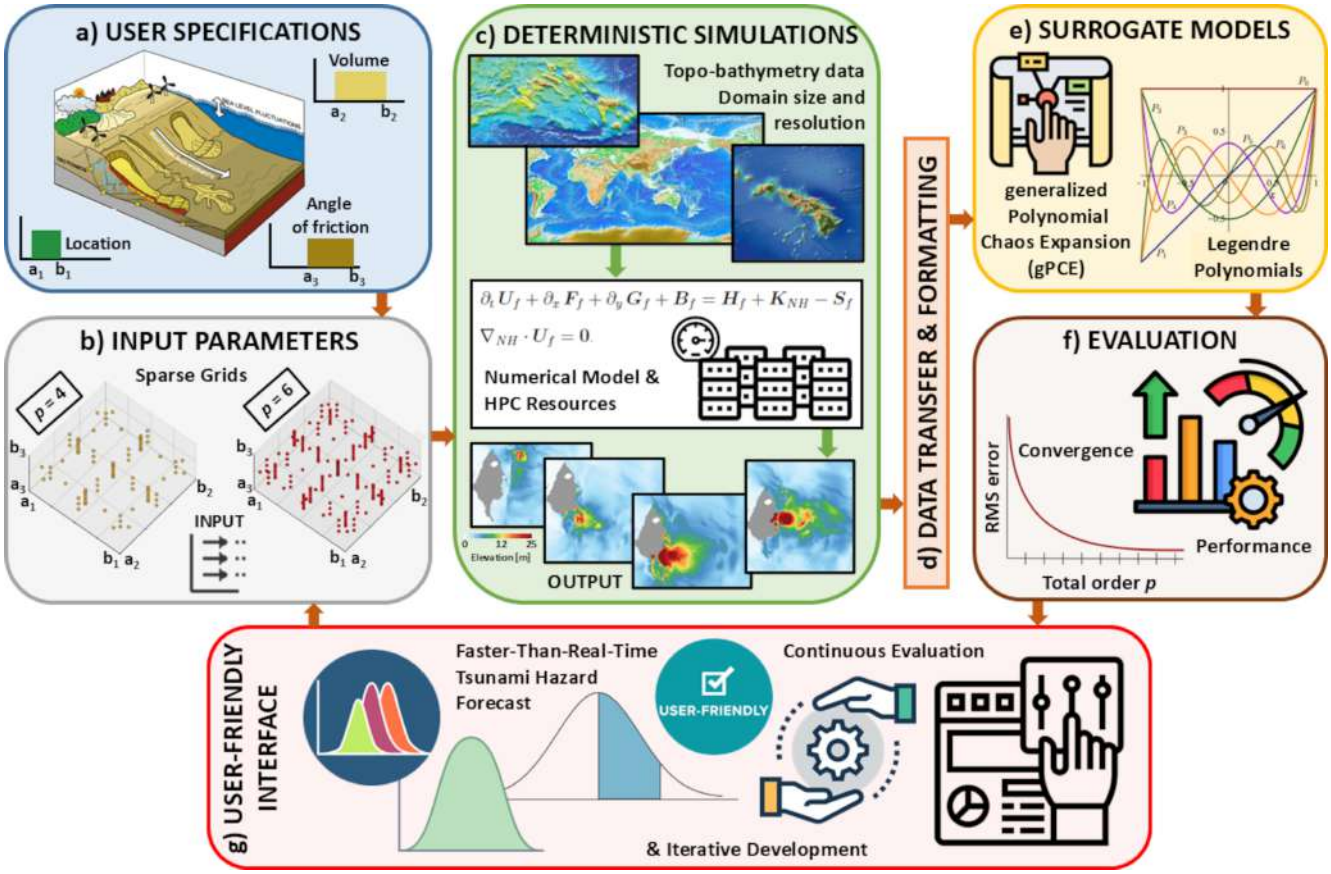


Figure 3. Overview of the seven Landslide-Tsurrogate v1.0 building blocks enabling faster-than-real-time Probabilistic Tsunami Forecast (PTF) based on surrogate models. These building blocks are: (a) **Step 1: User Specifications**, defining the stochastic variables; (b) **Step 2: Input Parameters**,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

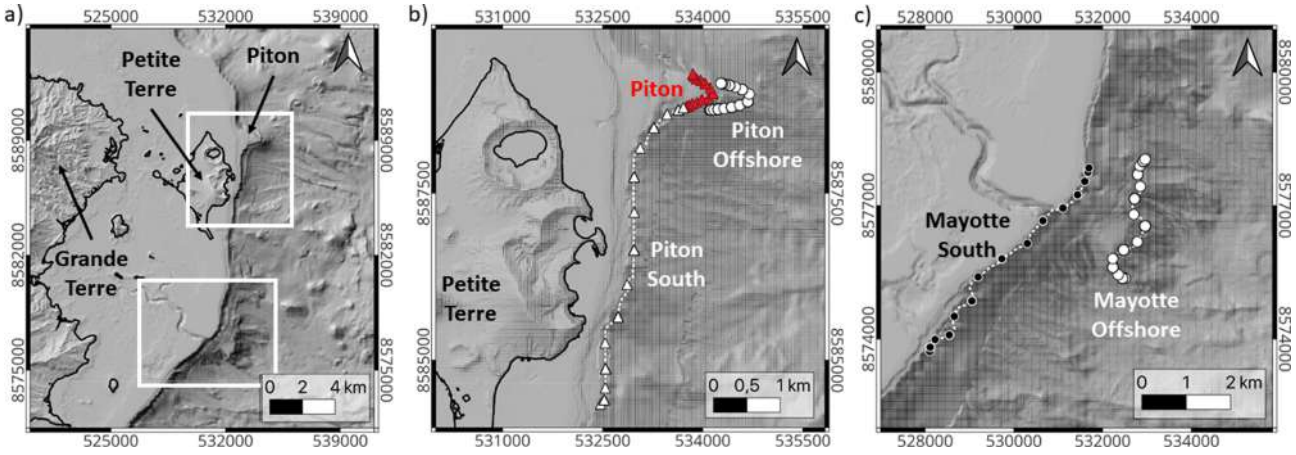


Figure 4. (a) Computational domain and 30m resolution topo-bathymetry of Mayotte. Landslides locations used to build the surrogate models along (b) the Piton (red dots), Piton Offshore (white dots) and South Piton (white triangles) zones and (c) the Mayotte South (black dots) and Mayotte Offshore (white dots) zones. Dashed white lines represent the isolines at steepest slopes.

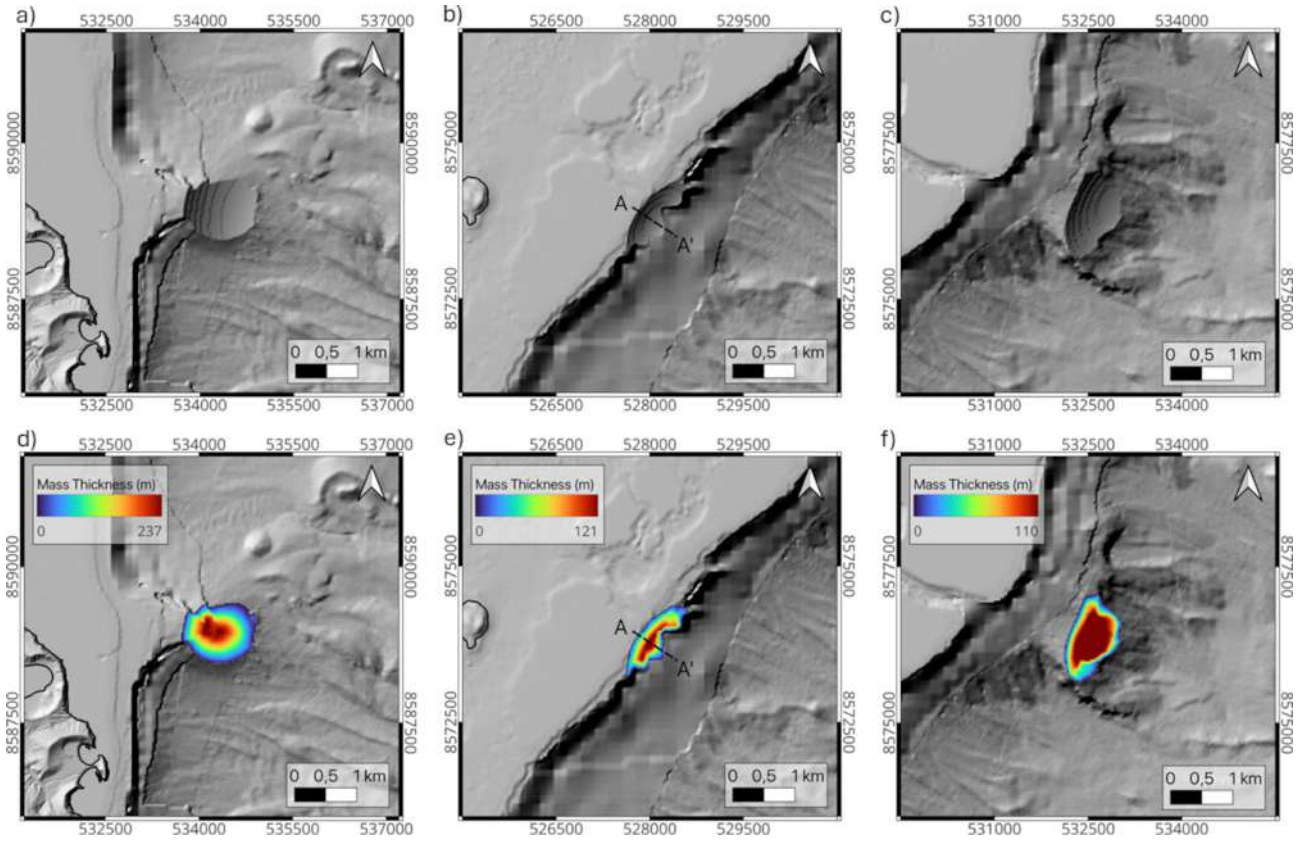


Figure 5. Illustration of three failure surfaces (panels a , b and c) and the corresponding mass thicknesses (panels d, e and f) created by the algorithm, superimposed to the topo-bathymetry at $30\mathrm{m}$ resolution: (a) and (d) simulation 124 from Piton Offshore ($V = 88.84\ 10^6\mathrm{m}^3$), (b) and (e) simulation 0 from Mayotte South ($V = 25.5\ 10^6\mathrm{m}^3$) with the cross section AA' displayed in Fig. 6 and (c) and (f) simulation 114 from Mayotte Offshore ($V = 71.84\ 10^6\mathrm{m}^3$).

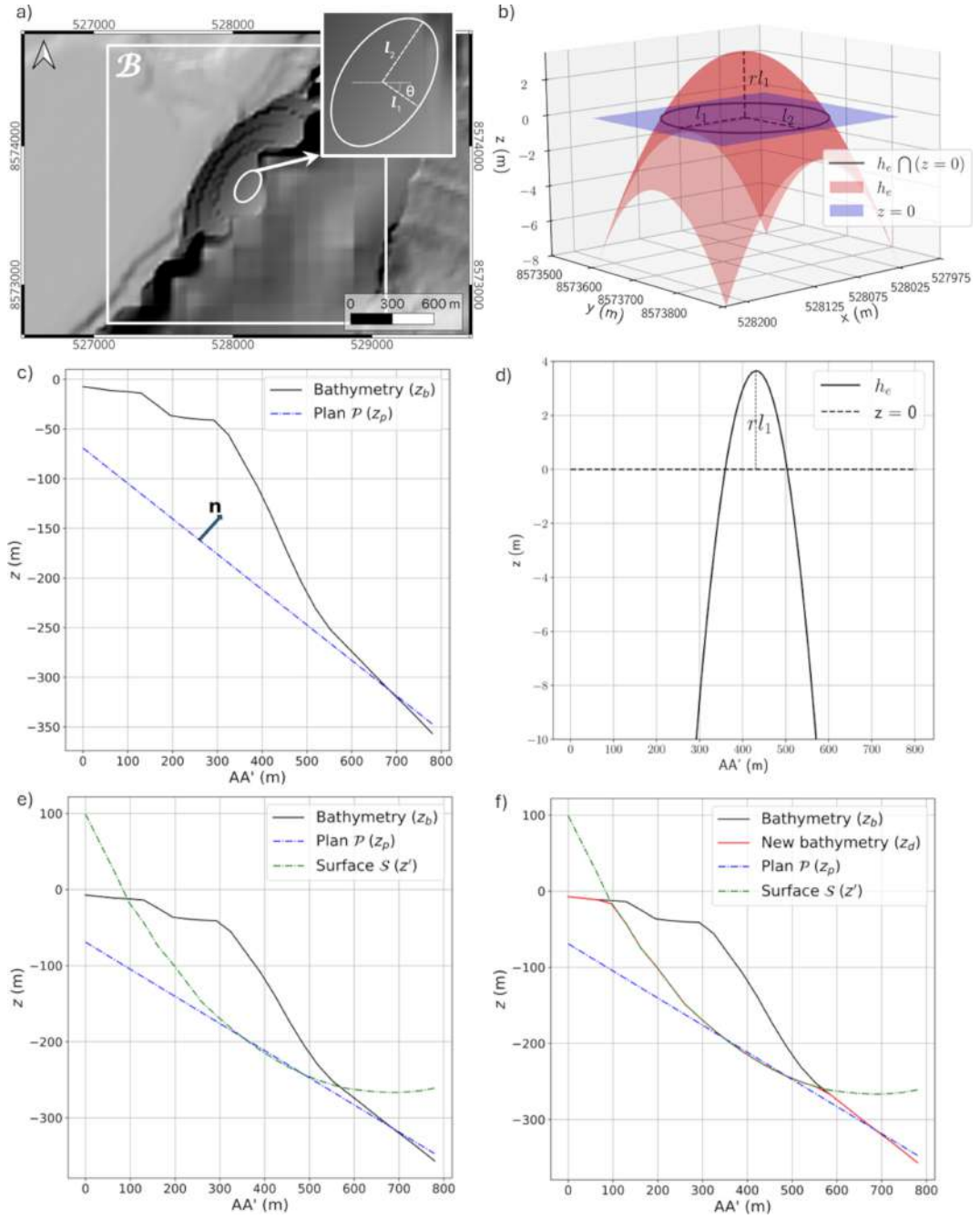


Figure 6. Submarine failure surface of simulation 0 from the Mayotte South zone. (a) the 30 m bathymetry with the failure surface, the box $\mathcal{B}$ from step 1, the ellipsoidal footprint, its semi-axis l_1, l_2 and the rotation angle θ involved in step 3. (b) 3D visualization of the paraboloid h_e from step 3, the plane $z = 0$ and the ellipsoidal footprint formed by the intersection. (c) - (d) Different steps of the algorithm plotted on the cross section AA' shown in Figure 5.

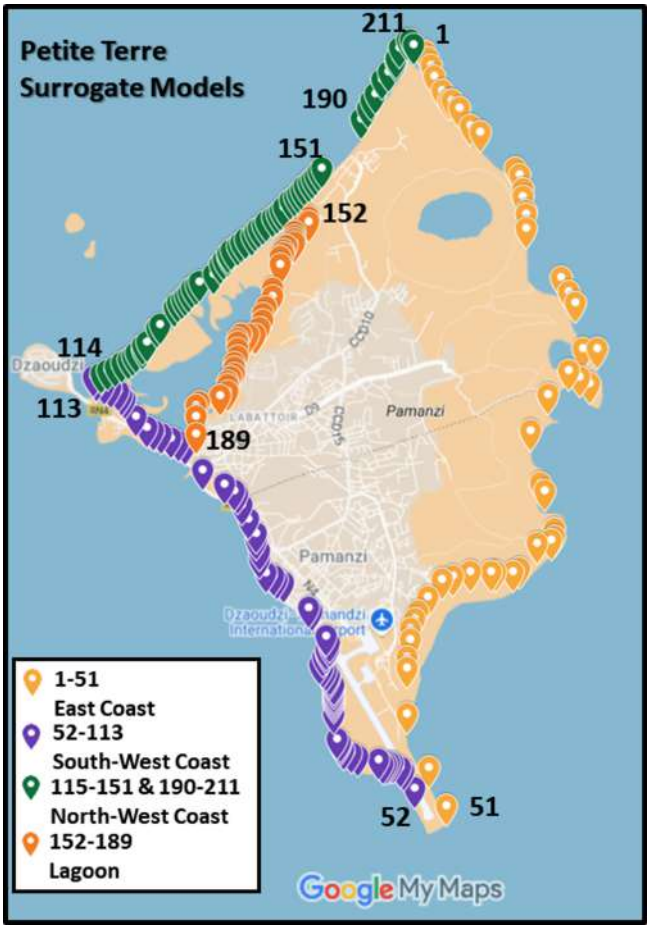


Figure 7. Locations and associated numbers of the 211 surrogate models built along the east, south-west, north-west coasts and the lagoon of Petite Terre. Map data: ©2026 Google.

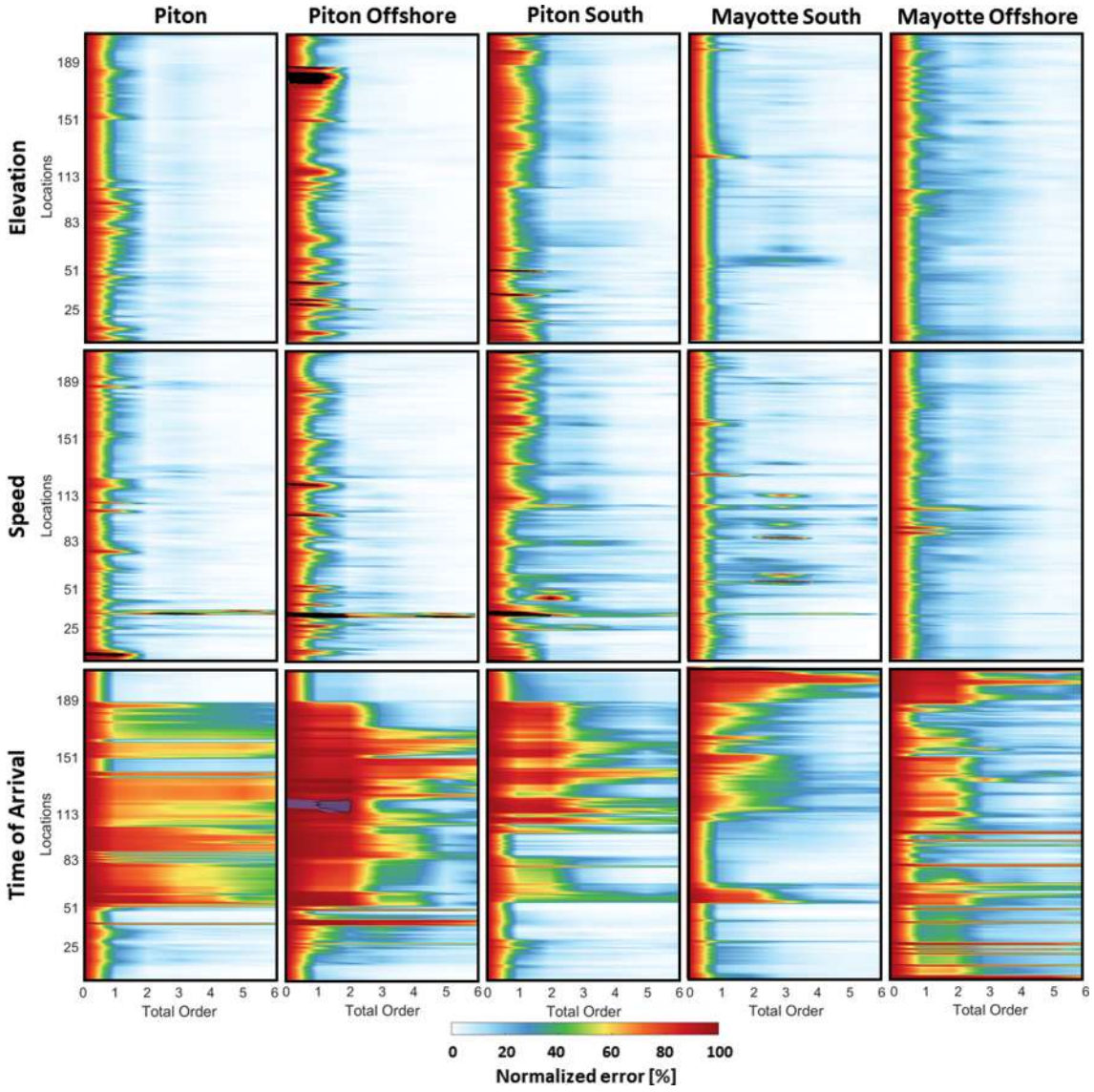


Figure 8. Convergence of the 211 Petite Terre surrogate models of maximum elevation, maximum speed and time of arrival of the landslide-tsunamis. For each of the Piton, Piton Offshore, Piton South, Mayotte South and Mayotte Offshore zones, the convergence is presented as the normalized error [Eq. (11)] in function of the total order of the truncated gPCE decomposition.

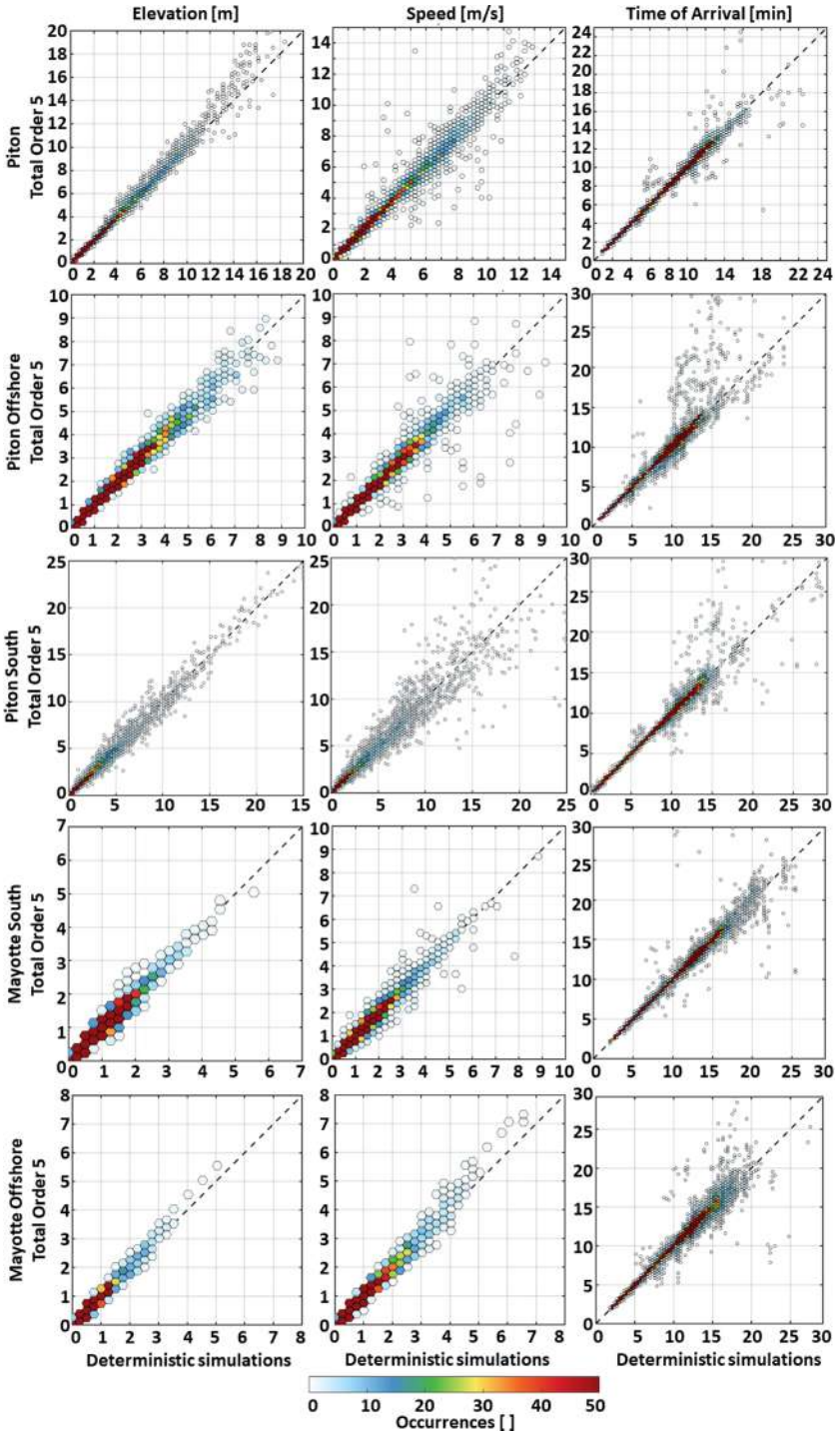


Figure 9. Accuracy of the 211 Petite Terre surrogate models of maximum elevation, maximum velocity and time of arrival of the landslide-tsunamis. For each of the Piton, Piton Offshore, Piton South, Mayotte South and Mayotte Offshore zones, the accuracy is presented as scatter plots of the surrogate model results derived from the gPCE truncated at the total order $P_t = 5$ (based on 135 simulations) as a function of the 72 independent deterministic simulations needed to reach a total order $p = 6$. The dashed black line represents the 1:1 reference line.

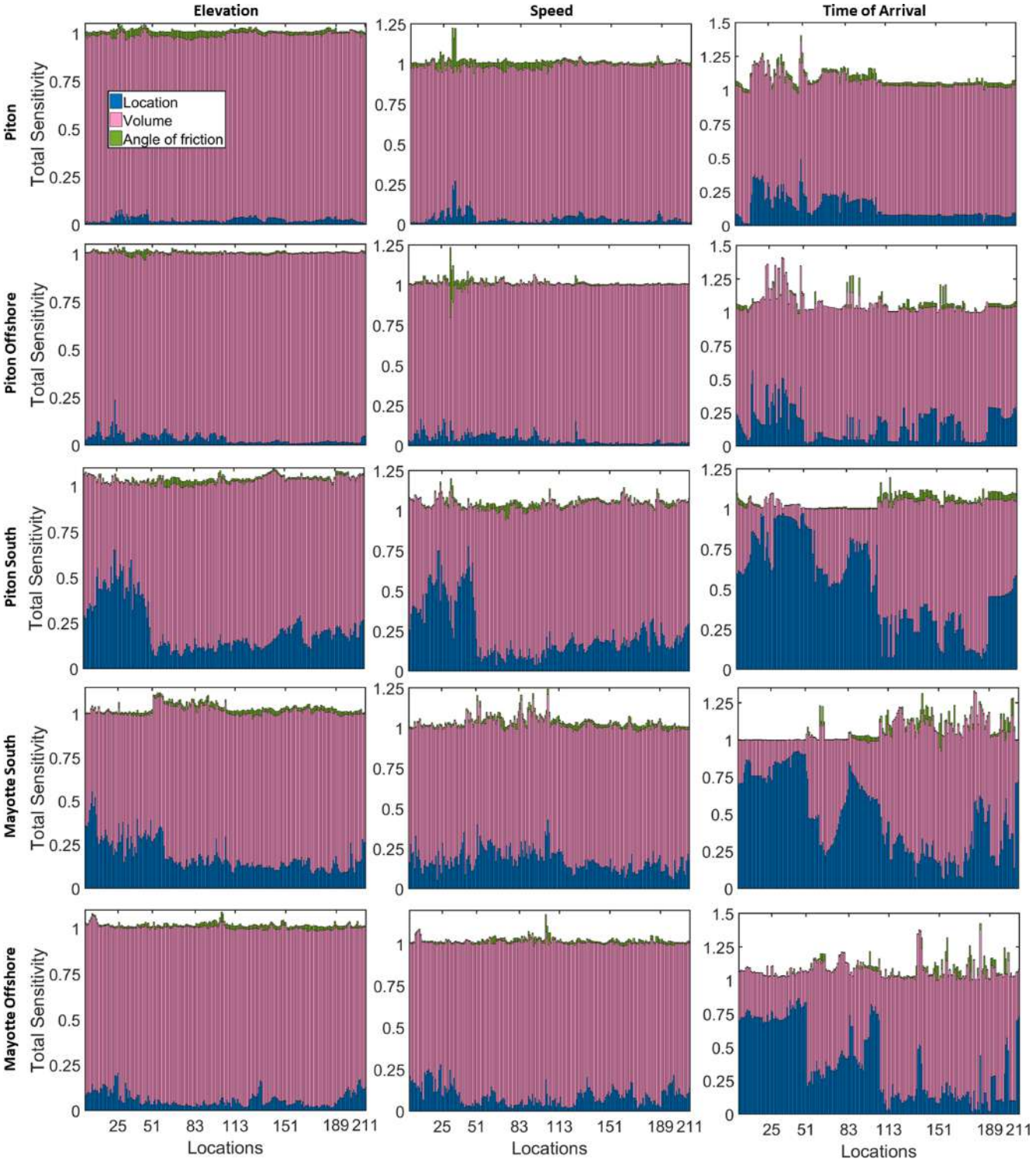


Figure 10. Total sensitivity of the 211 Petite Terre surrogate models of maximum elevation, maximum speed and time of arrival of the submarine landslide-tsunamis. For each of the Piton, Piton Offshore, Piton South, Mayotte South and Mayotte Offshore zones, the total sensitivity is presented as bar plots.

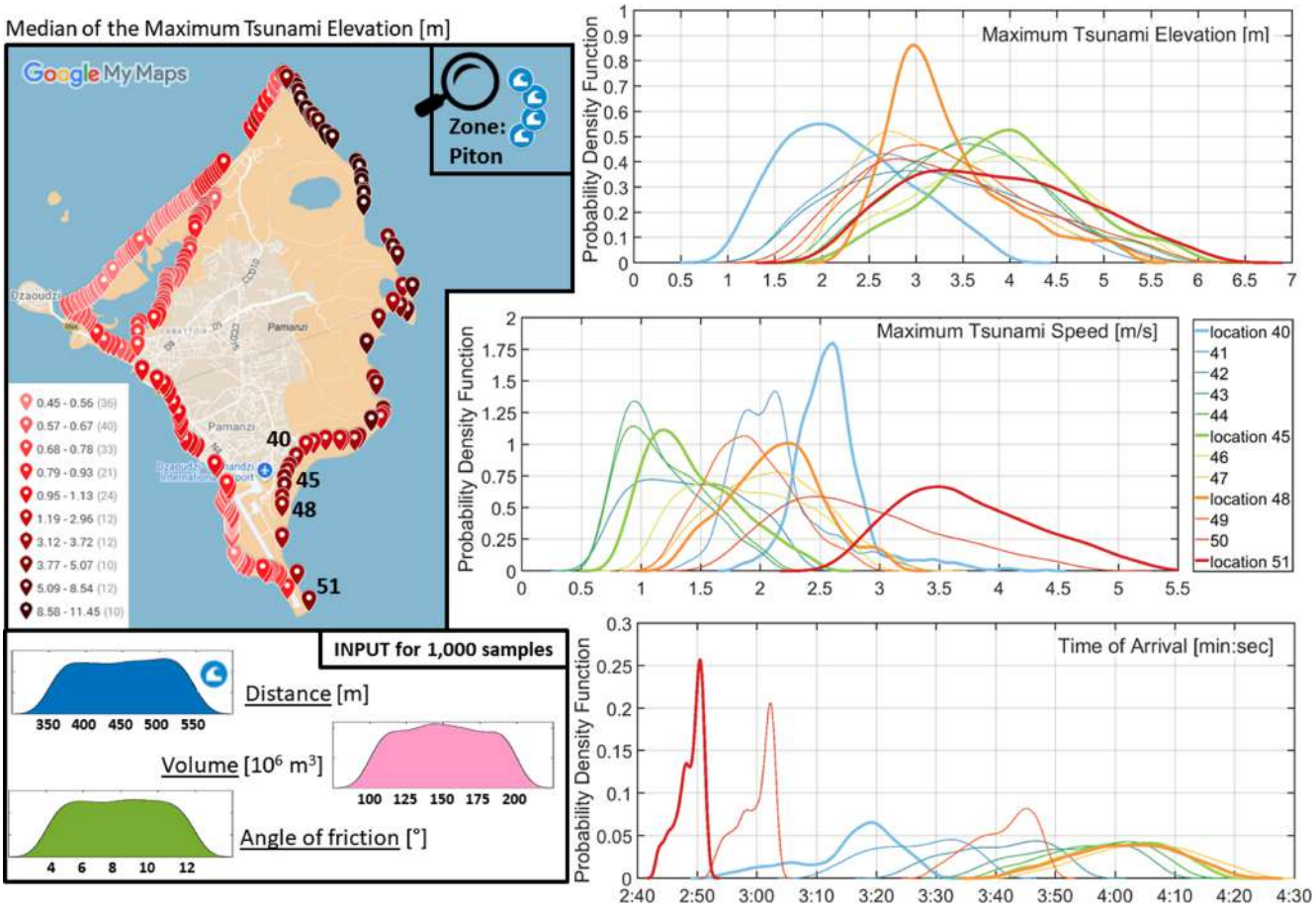


Figure 11. Landslide Probabilistic Tsunami Forecast (PTF) derived from $1,000$ conditions (or samples) for the Petite Terre Airport area and considering submarine landslides with volumes of $V = 100\tilde{}200\ 10^6 \mathrm{m}^3$ located at the tip of the Piton zone. The map represents the median of the maximum tsunami elevation distributions along the Petite Terre coastline while the probabilistic density distributions are displayed based on the surrogate models of maximum elevation, maximum speed and time of arrival located along the eastern coast of the airport area. Map data: ©2026 Google.

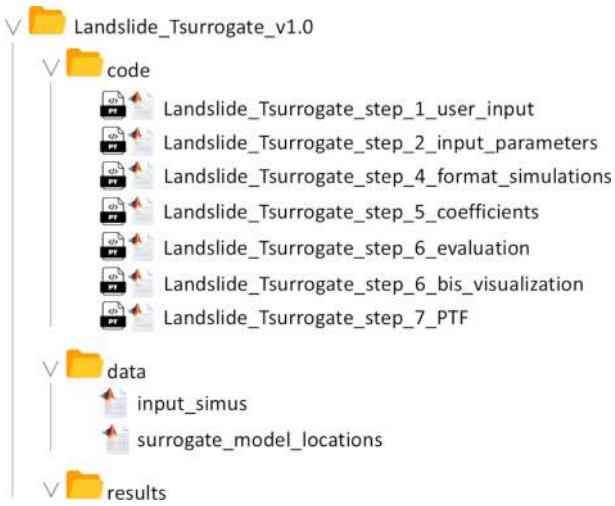


Figure A1. Structure of the Landslide-Tsurrogate v1.0 software.

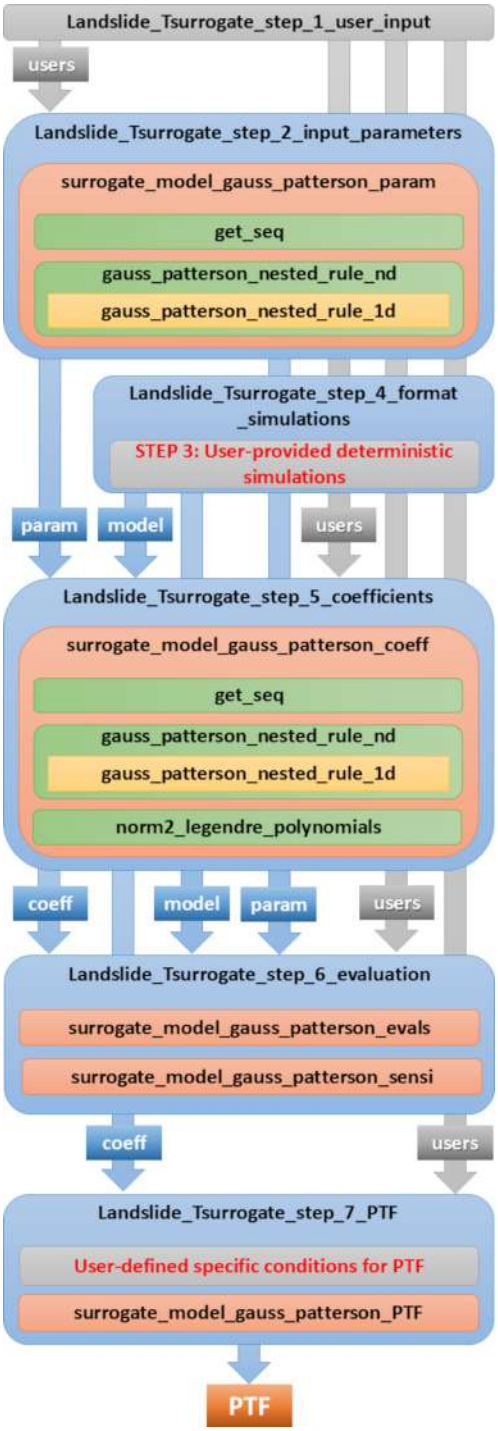


Figure A2. Workflow pipeline of the main programs and associated functions in Landslide-Tsurrogate v1.0 for the generation of the surrogate models. User-oriented programs are shown in Blue boxes, core building-block functions in Orange, and supporting functions in Green and Yellow. Functions or programs that require user editing are highlighted in Grey. The input/output flow between components is represented by arrows, labeled with the names of the variables saved in the corresponding files.

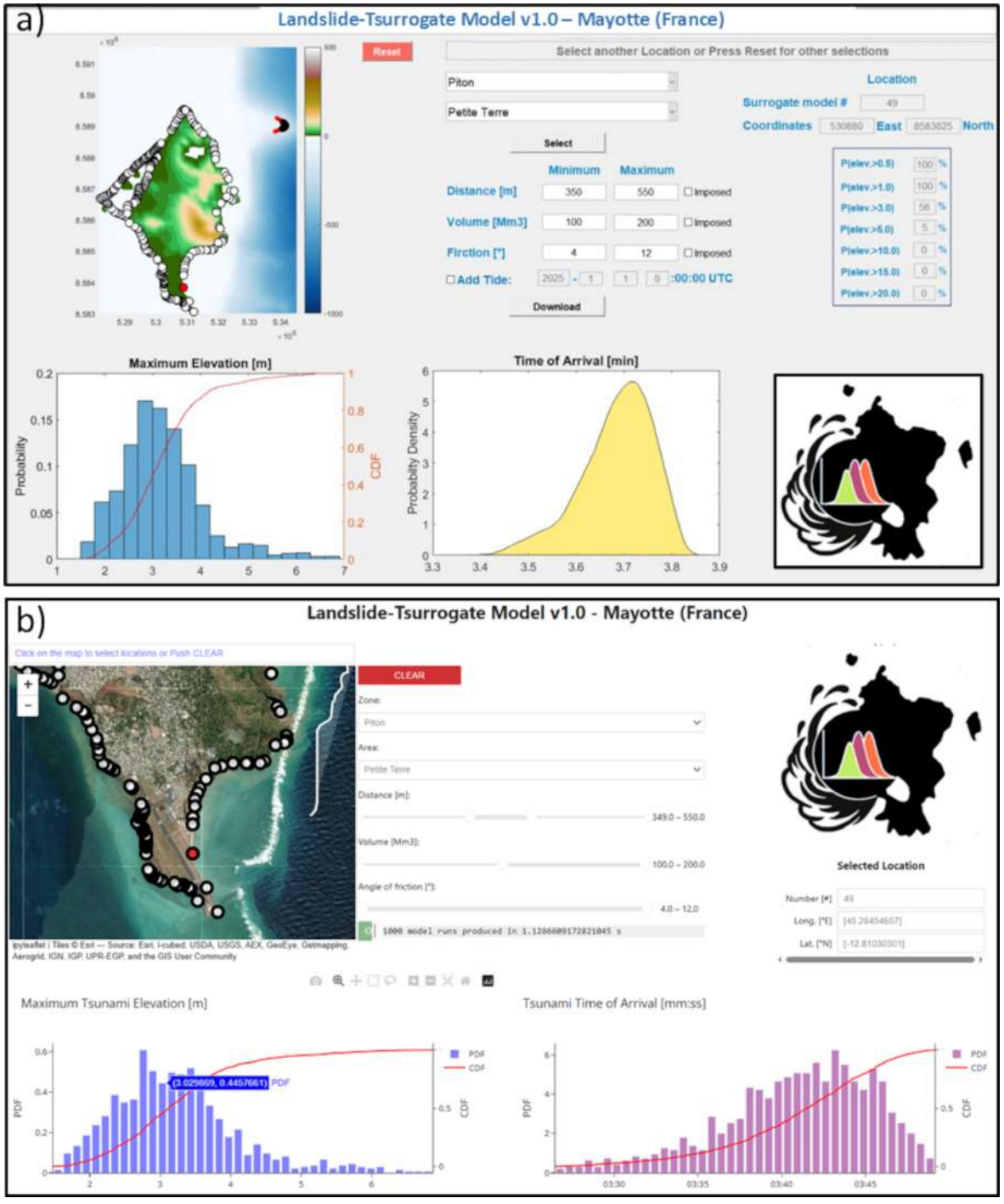


Figure A3. Example of Landslide Probabilistic Tsunami Forecast (PTF) user-friendly interfaces for the Mayotte test case: (a) developed in MATLAB and distributed as program files for experts or as a Windows installer (.exe) for non-experts and (b) developed in Python and distributed as Jupyter Notebook for experts or deployed as a web application through Voilà for non-experts. Map sources: Esri, i-cubed, USDA, AEX, GeoEye, Getmapping, Aerogrid, IGN, IGP, UPR-EGP, and the GIS User Community.