JGR | Machine Learning and Computation

RESEARCH ARTICLE
10.1029/2025JH001100

Stochastic Emulation for Efficient Onshore Probabilistic Tsunami Hazard Assessment

Naveen Ragu Ramalingam[1,2], Alice Abbate[3,4], Erlend Briseid Storrøsten[1], Gareth Davies[5], Stefano Lorito[4], Manuela Volpe[4], Steven J. Gibbons[1], Fabrizio Romano[4], and Finn Løvholt[1]

[1]Norwegian Geotechnical Institute, Oslo, Norway, [2]University School for Advanced Studies - IUSS Pavia, Pavia, Italy, [3]Deutsches GeoForschungsZentrum, Potsdam, Germany, [4]Istituto Nazionale di Geofisica e Vulcanologia, Roma, Italy, [5]Geoscience Australia, Canberra, ACT, Australia

Abstract Machine learning is emerging as a promising strategy for modeling tsunami inundation at reduced computational cost. To enable probabilistic outputs that capture emulator uncertainty, we employ an ensemble of stochastic encoder-decoder emulators with dropout-based stochastic forward pass. Training and testing is conducted on an extensive Mediterranean Sea tsunami simulation data set comprising multiple earthquake source regions with variable magnitude, location, and focal mechanism, including near-field events with local coseismic deformation. For the two target sites Catania and Siracusa in Sicily, we achieve good performance for relatively sparse training sets of only a few hundred events. We demonstrate the potential of our emulators to accurately model inundation from a range of tsunami sources needed to perform probabilistic tsunami hazard assessment (PTHA), reducing the computational cost of PTHA by a factor of 30–100, and estimating the uncertainty introduced by the emulator. Comparing probabilistic hazard maps and curves from our emulation with those from the full simulation ensemble and importance sampling Monte Carlo approach, we demonstrate the efficacy of emulators in capturing the spatial distribution and magnitude of inundation needed in overland tsunami hazard assessment.

Plain Language Summary Probabilistic Tsunami Hazard Analysis (PTHA) requires simulating tens of thousands of tsunami scenarios, with high-resolution inundation modeling being the most computationally expensive step. Machine learning (ML) offers a way to predict inundation outcomes using only a limited number of full simulations for training, greatly reducing computational cost. In this study, we develop an ML model to predict detailed tsunami inundation maps from offshore wave data and local ground deformation caused by earthquakes. The model is trained on simulated Mediterranean tsunamis and tested at two sites in Eastern Sicily, using earthquakes that generate offshore tsunami waves and, in some cases, local ground shifts. Despite using only a few hundred training events, in testing it accurately predicted flood depths across a wide range of earthquake scenarios. The approach not only estimates expected inundation but also provides a range of possible outcomes and their likelihoods, improving uncertainty quantification. By incorporating ML into PTHA, we can efficiently evaluate thousands of potential tsunami events, making large-scale hazard and risk assessment feasible. This advancement enables more cost-effective and timely analysis, supporting better disaster preparedness, planning, and resilience for coastal communities exposed to tsunami hazards.

Key Points:

- Enabling up to 100× faster inundation modeling for probabilistic tsunami hazard analysis using uncertainty aware emulation
- Extensive testing at two Mediterranean sites using large simulation ensembles with complex EQ sources and near-field effects
- Static deformation inputs for local events, asymmetric loss function for extremes, and stratified data selection for efficient training

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

N. Ragu Ramalingam,
naveen.ragu@ngi.no

Citation:

Ragu Ramalingam, N., Abbate, A., Briseid Storrøsten, E., Davies, G., Lorito, S., Volpe, M., et al. (2026). Stochastic emulation for efficient onshore probabilistic tsunami hazard assessment. *Journal of Geophysical Research: Machine Learning and Computation*, *3*, e2025JH001100. https://doi.org/10.1029/2025JH001100

Received 31 OCT 2025
Accepted 28 MAR 2026

Author Contributions:

Conceptualization: Naveen Ragu Ramalingam, Erlend Briseid Storrøsten, Gareth Davies, Stefano Lorito, Steven J. Gibbons, Fabrizio Romano
Data curation: Naveen Ragu Ramalingam, Manuela Volpe
Formal analysis: Naveen Ragu Ramalingam, Erlend Briseid Storrøsten
Investigation: Naveen Ragu Ramalingam, Erlend Briseid Storrøsten, Gareth Davies, Stefano Lorito, Fabrizio Romano

1. Introduction

Tsunamis pose significant threats to coastal communities, with the damage caused primarily by their inundation impacts resulting in loss of life, destruction of infrastructure, and severe economic consequences as illustrated by the 2004 Sumatra-Andaman, 2010 Maule, Chile and 2011 Tohoku, Japan tsunami events (Fritz et al., 2011; Ghobarah et al., 2006; Mori et al., 2011). For optimal disaster preparedness, analysis of tsunami inundation and uncertainty quantification is fundamental to address within the local hazard context. In probabilistic applications dealing with long-term tsunami hazard or acute alert situations, source uncertainty should be addressed by considering sufficiently large ensembles of scenarios (Behrens et al., 2021; Geist & Parsons, 2006; Grezio et al., 2017). This should be accompanied by high-resolution numerical simulations, which in most situations involve solving the nonlinear shallow water equations. However, the computational resources and time needed for such simulations can be prohibitive (Davies et al., 2022; Lorito et al., 2015; Selva et al., 2021). To deal with these constraints, the size of the ensemble of scenarios has often been restricted based on clustering, statistical

Methodology: Naveen Ragu Ramalingam, Erlend Briseid Storrøsten, Gareth Davies, Stefano Lorito
Project administration: Naveen Ragu Ramalingam
Resources: Naveen Ragu Ramalingam, Stefano Lorito, Manuela Volpe
Software: Naveen Ragu Ramalingam
Supervision: Erlend Briseid Storrøsten, Gareth Davies, Stefano Lorito
Validation: Naveen Ragu Ramalingam, Alice Abbate, Steven J. Gibbons
Visualization: Naveen Ragu Ramalingam, Steven J. Gibbons
Writing – original draft: Naveen Ragu Ramalingam
Writing – review & editing: Alice Abbate, Erlend Briseid Storrøsten, Gareth Davies, Stefano Lorito, Steven J. Gibbons, Fabrizio Romano

sampling, and other reduction techniques, which by themselves introduce an additional source of uncertainty, that is, seldom quantified (Baba et al., 2022; Davies et al., 2022; Goda et al., 2023; Lorito et al., 2015; Volpe et al., 2019; Williamson et al., 2020). When feasible, larger numbers of scenarios can still be considered to reduce this uncertainty using High-Performance Computing (HPC) (Gibbons et al., 2020; Tonini et al., 2021). However, access to HPC infrastructure remains limited and costly, particularly outside well-funded research environments and in many tsunami-prone regions, constraining the widespread application of large-ensemble, high-resolution probabilistic tsunami hazard assessment (PTHA).

Tsunami emulators act as a surrogate for numerical model simulators, providing approximate yet computationally inexpensive predictions, particularly when evaluating a large number of scenarios. To be efficient, an emulator should use only a limited number of numerical simulations for its training, while introducing a reasonably bounded additional uncertainty. Their design must align with the specific location, available training data, resources, and application requirements. Statistical emulators offer fast approximations with error estimates that quantify prediction uncertainty (Fukutani et al., 2023; Salmanidou et al., 2021; Tozato et al., 2022). Large tsunami simulation data sets can be used to train machine learning (ML) models (e.g., such as neural networks) which in turn are applied as emulators (Makinoshima et al., 2021; Mulia et al., 2022; Núñez et al., 2022). Tsunami emulators have primarily been designed for rapid early warning applications, applied for near-field tsunamis at sites proximal to the earthquake source region, for example, Cascadia (Liu et al., 2021), Nankai (Fauzi & Mizutani, 2020), Tohoku (Makinoshima et al., 2021; Mulia et al., 2022), Chile (Núñez et al., 2022).

Previous studies (Briseid Storrøsten et al., 2024; Ragu Ramalingam et al., 2025) assessed the feasibility of site-specific ML emulators to predict high-resolution inundation maps and coastal wave time series using simulated tsunami waveforms from deep-offshore point locations as input. Briseid Storrøsten et al. (2024) used an ensemble of 28,000 tsunami scenarios originally defined for PTHA (Gibbons et al., 2020) originating from subduction zones distant to the inundation site of interest. Their emulator could greatly reduce the numerical expense of inundation ensemble simulations, using a relatively small training data set (a few hundred earthquake scenarios), and they investigated how the predictive performance depends on the ML model settings such as the number of kernels, the latent space, and the type of loss function. For simplicity, inundation was considered along only one stretch of coastline and the study purposefully omitted crustal earthquakes that produced coseismic deformation at the coast of interest. Subsequently, Ragu Ramalingam et al. (2025) employed an ensemble of probabilistic emulators based on variational encoding to quantify uncertainty in the output. Here, we extend these approaches towards a general tsunami inundation emulation with uncertainty quantification, capable of evaluating a broad range of earthquake sources.

We introduce emulators designed specifically and optimized for rigorous uncertainty quantification in PTHA workflows. They are developed for two coastal regions of Catania and Siracusa in Eastern Sicily, Italy (Figure 1), leveraging a pre-existing, extensive database of inundation simulations previously used for PTHA (Abbate et al., 2025; Gibbons et al., 2020; Tonini et al., 2021). Both regions are prone to tsunami inundation from both subduction and crustal earthquakes in the Mediterranean Sea, including local crustal earthquake sources (Basili et al., 2021; Tonini et al., 2021). Tsunami simulation results are sensitive to the choices made in rupture modeling and may exhibit significant differences depending on whether a planar or three-dimensional fault surface is used, whether the slip distribution is homogeneous or heterogeneous, and the range of fault-rupture parameters employed (Gibbons et al., 2022; Tonini et al., 2020). Different source features give rise to variations in tsunami generation, which in turn govern how the waves interact with bathymetry and coastal morphology along their propagation paths. In our simulation data set, different sources are represented with varying complexity, with near-field events characteristically generating shorter wavelengths and arrival times than for far-field events. These varied rupture modeling approaches can generate different input waveforms. Accommodating the full range of complexity is of paramount importance when building a tsunami emulator, as any tendency of the emulator to overfit on certain waveform features can affect its generalization and prediction accuracy. To this end, we integrate a compound set of sources shown in Figures 2a and 2b).

The quality of emulator predictions depends strongly on the training data set, whose selection and size are typically constrained by the computational resources available for carrying out expensive high-resolution simulations (de la Asunción, 2024; Gibbons et al., 2020) or the availability of existing data sets (Fauzi & Mizutani, 2020; Kamiya et al., 2022; Liu et al., 2021). Previous approaches mainly focused on considering the event variability across the earthquake magnitude ranges (Makinoshima et al., 2021; Núñez et al., 2022). Alternatively,

10.1029/2025JH001100

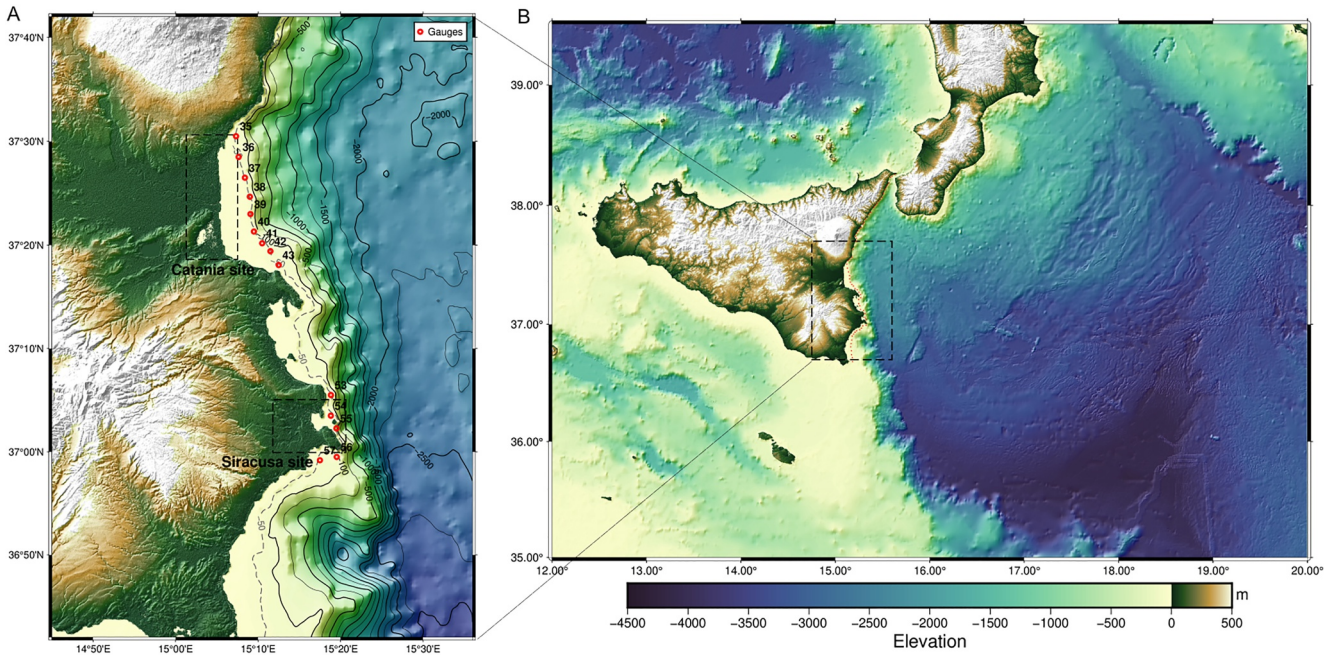

Figure 1. (a) Map of the model region showing the two inundation sites of Catania and Siracusa. The input gauge locations have a depth of approximately 50 m, and are marked on the map with following ranges: Catania from 35 to 44, and Siracusa from 53 to 58. The dashed extent marks the coverage of the inundation grids. (b) The coastal study region and seafloor bathymetry in a wider context.

Green's function summation technique can be used to efficiently generate offshore time-series and analyze the variability of input wave heights as a way to reduce the number of full inundation simulations (Mulia et al., 2022), which is conceptually similar to clustering analysis (e.g., Lorito et al. (2015); Volpe et al. (2019)). In this study, we propose a stratified selection method that enhances the training data selection process by exploiting multiple variables commonly available from offshore PTHA that also serve as proxies for inundation, avoiding the need for full inundation simulations while still accounting for the influence of multiple sources and inundation mechanisms.

Catania and Siracusa are characterized by very different coastal morphology. The Bay of Catania is an open beach with a smooth coastal shelf, resulting in gradual shoaling and inundation over large areas for relatively low amplitude tsunami waves offshore. In contrast, Siracusa is a closed bay with steep slopes onshore and offshore, leading to much greater inundation heights but with a smaller area inundated. The coastline of Siracusa is exposed to inundation both from south and east. Resonance effects (Bressan & Tinti, 2016) may influence the inundation characteristics at different locations within the bay. Further, local earthquakes have affected this region in 1169, 1693, and 1908 (Scicchitano et al., 2007) and cause onshore deformation that significantly influences the inundation area and flow depths for an event. The onshore deformation from local earthquake events can cause uplift, subsidence, or a combination of both along different sections of the coastline. Recognizing and integrating such variability is crucial for developing a robust database of tsunami sources, propagation, and inundation, that serves as training data for an emulator capable of reliably predicting inundation across diverse scenarios and sites.

Our emulator accounts for local earthquake sources that can cause significant coastal subsidence or uplift along the target coastline, as exemplified in the recent 2024 Noto Earthquake (Heidarzadeh et al., 2024). As an analog to using ground displacement from GNSS data for real-time forecasting (Rim et al., 2022), we incorporate static local deformation fields as an additional input parameter to the emulator (Figure 6). The consideration of local topography changes as a dependent variable introduces several challenges compared to emulating tsunami inundation from just offshore time series. These include the need for suitable encoding for the spatially distributed deformation fields, the coupling between local deformation, tsunami shoaling and inundation, and incorporation of their associated uncertainties into the emulator predictions. We enhance the emulation predictions with uncertainty quantification based on a framework applied in Ragu Ramalingam et al. (2025). This uncertainty quantification framework leverages an ensemble of models trained with different partitions of training data

10.1029/2025JH001100

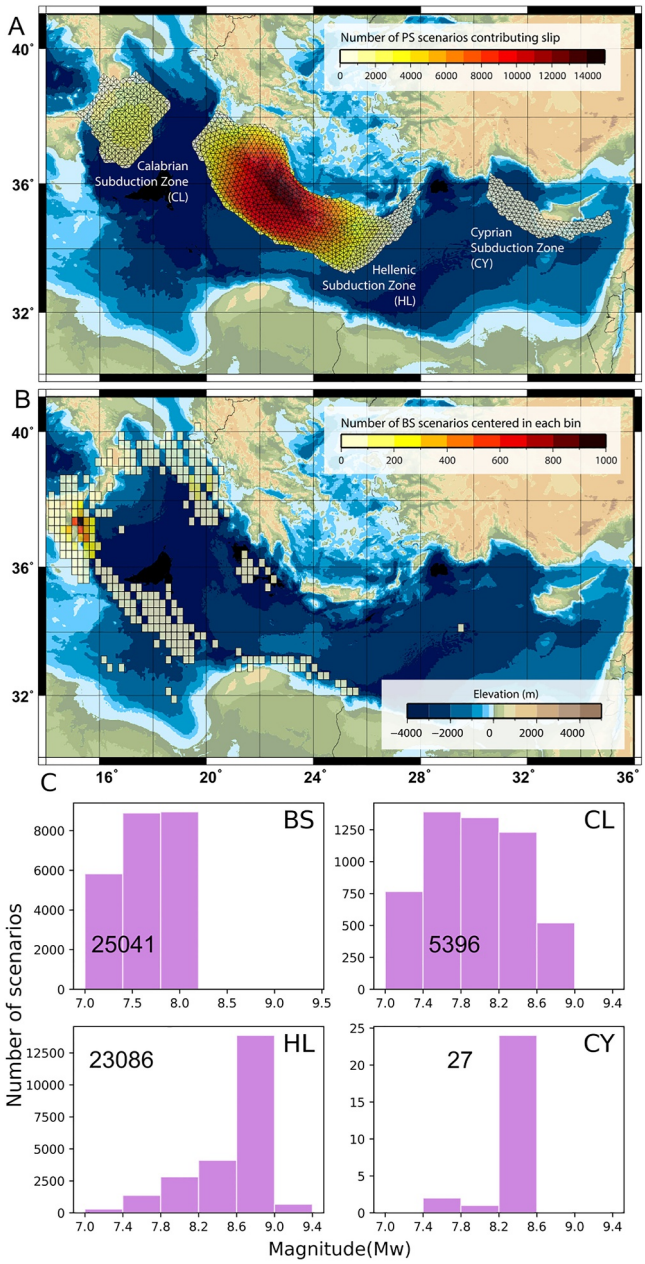

Figure 2. (a) Distribution of subduction-type (PS) earthquake sources. The Calabrian (CL), Hellenic (HL), and Cyprian (CY) subduction zones are shown as triangular meshes, colored by the number of PS scenarios with non-zero slip in each segment. (b) Distribution of crustal (BS) earthquake scenarios significantly contributing to tsunami hazard in Sicily. BS sources are shown as rectangles, colored by the number with epicenters per cell. Most BS sources are near the coast, with local events varying in size and distant ones typically large. (c) Magnitude (M_w) distribution of scenarios across source types: CL sources (closest to the target shoreline) span a broad range, while HL and CY (more distant) mainly contribute large events.

(similar to cross validation) and Bayesian methods (Nemani et al., 2023) to address uncertainties arising from limited training data and model parameter variability, thus improving the robustness and reliability of emulation methods.

Through a demonstrator of a local high resolution PTHA, we compare the emulation results with those obtained using conventional HPC techniques (Gibbons et al., 2020) and a well-established statistical sampling approach called stratified importance sampling (SIS) (Abbate et al., 2025; Davies et al., 2022). The remainder of the paper is organized as follows: Section 2 describes the methodology and enhancements to the emulator, Section 3 presents the emulation results and comparisons with other approaches, and Section 4 discusses the implications and future directions.

2. Data and Methods

2.1. Tsunami Simulation Dataset: Multiple Sources, Mechanisms and Inundation Sites

The machine learning experiments in this work are supported by a simulation data set of 53,550 synthetic events selected from the North-East Atlantic and Mediterranean Tsunami Hazard Model NEAMTHM18 (Basili et al., 2021; Gibbons et al., 2020) summarized in Figures 2 and 3 and Table S1 in Supporting Information S1. These events were selected using a hazard disaggregation procedure (Gibbons et al., 2020) for the two sites of Catania and Siracusa, in eastern Sicily. This data set includes simulations for both subduction earthquakes (Predominant Seismicity–PS) in the Calabrian, Hellenic, and Cyprian arcs and intra-plate crustal earthquakes (Background Seismicity–BS) (see Figure 2 panels A and B). The tsunami sources are represented using stochastic heterogeneous slip for earthquake events on the three subduction interfaces and by rectangular uniform slip for crustal-type events (Gibbons et al., 2020; Scala et al., 2020; Selva et al., 2016). These simulations were carried out using the Tsunami-HySEA implementation of the non-linear shallow water equations (de la Asunción et al., 2013) and were executed on Tier 0 high-performance computers under the ChEESE project (Folch et al., 2023). Each scenario required approximately 25 min of wall-clock time, with one GPU allocated per scenario, to complete the full tsunami propagation and inundation simulation. During all scenario simulations, offshore tsunami wave time series are recorded at virtual points of interest (offshore gauges as shown in Figure 1) and inundation is modeled using a nested grid system going down to 10 m spatial resolution. The maximum flow depth is used as the intensity measure (IM) to characterize tsunami inundation hazard from these simulations. Table 3 compares both simulation runtimes and emulator inference times. This simulation data set has also been used to benchmark the SIS approach for inundation PTHA (Abbate et al., 2025).

In the NEAMTHM18 regional hazard model, tens of millions of earthquake scenarios are assessed. For each scenario, seafloor displacement and offshore tsunami wave time series were computed at virtual offshore gauges using linear superposition of elementary sources modeled on coarse grids (Molinari et al., 2016). The scenario annual occurrence rates were derived from a probabilistic earthquake model (Basili et al., 2021). Using linear model-based offshore wave time series, approximate scenario level metrics (such as the maximum offshore wave height) are available a priori for all scenarios. They are used for the disaggregation analysis and scenario selection procedure, before any high resolution onshore inundation modeling is conducted. Although the metrics derived from these linear models may differ slightly from those obtained using nonlinear propagation-inundation simulations, they

10.1029/2025JH001100

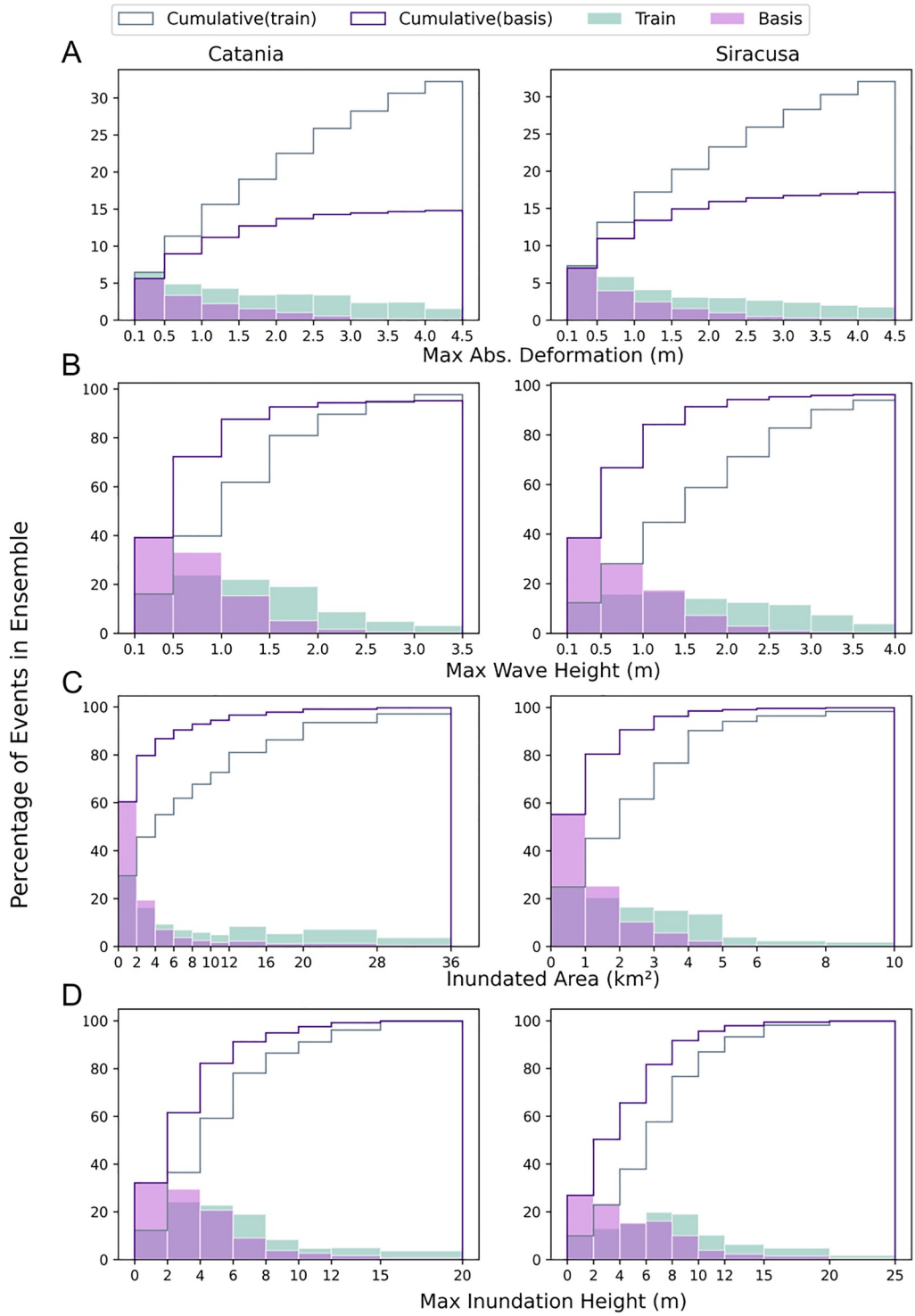

Figure 3. Percentage histograms of tsunami metrics for Catania and Siracusa in the training set (Moderate Training Size with about 3.3% of total scenarios: see text) and full simulation (basis) set. Only absolute deformation (a) and maximum offshore wave height (b) can be used for training selection since inundation area (c) and maximum inundation height (d) are available only after inundation has been simulated for the scenarios. Events generating higher maximum wave heights are more prominently represented in the training set than in the basis set.

are sufficiently informative for event characterization. This enables the identification of representative training subsets of events prior to computationally expensive inundation simulations. This approach preserves the intended computational efficiency of the emulation based PTHA framework, by using coarse resolution offshore propagation for the generation of emulation input.

2.2. Training, Ensemble Strategy and Model Evaluation

Our goal is to develop an inundation emulator on the basis of the offshore waves and local deformation for the entire set of events at relatively low cost. The emulator needs to be trained on a limited number of full inundation simulations (computed at a much higher computational cost). Selecting an appropriate subset of events for which the full simulation is performed is therefore crucial.

Randomly choosing training events from a full PTHA scenario database will result in a subset dominated by the more frequently occurring lower magnitude events. Therefore, adequately representing all anticipated levels of inundation severity requires including a disproportionately higher number of large-magnitude events than suggested by their occurrence rates (Briseid Storrøsten et al., 2024). As we do not know a priori the distribution of inundation impact metrics without having first performed the inundation simulations, sampling strategies for choosing events which best cover the anticipated ranges must therefore depend on proxy variables (Abbate et al., 2025; Baba et al., 2022; Davies et al., 2022; Volpe et al., 2019; Williamson et al., 2020), such as source parameters, local deformation, offshore wave heights, or arrival time.

To maximize the utility of the limited samples for emulator training, we adopt a four-fold cross-validation strategy to create an ensemble of models. In this approach, the sampled data set is partitioned into four equally sized subsets. Three subsets are used for training while the remaining one is held out for validation in a typical 75:25 data split. This process is repeated four times, with each subset serving once as the validation set, as shown in Step 1 of Figure 4. Rotating the validation set across folds ensures that the entire sampled subset is utilized for both model fitting and performance assessment, providing a more robust evaluation of generalization during the training phase. This setup enables effective monitoring of overfitting and supports early stopping through performance evaluation on the held-out validation data during each fold. Leveraging the full information available in the sampled subsets through this ensemble approach also forms a core part of the stochastic emulation strategy which, combined with the Monte Carlo (MC) dropout discussed in Section 2.3.2, allows for a comprehensive quantification of predictive uncertainty.

Crucially, our partitioning strategy departs from standard machine learning practices that typically prioritize a large training-to-testing ratio to maximize learning from a limited data pool. Instead, we utilize the remaining portion of the basis set with more than 50,000 events as a final test set. By reserving the vast majority of our data for this final stage, we subject the emulator to a rigorous stress test. This allows us to verify the emulator's performance across nearly the entire event spectrum, ensuring that the model remains reliable and physically consistent even when trained on a small fraction of the total available simulations.

2.2.1. Selection of Training Data Using a Stratified Sampling Procedure

To consider the impact of multiple sources and inundation mechanisms, we refine our selection using variables available from the regional offshore PTHA analysis for each scenario: (a) the maximum absolute deformation across the onshore inundation grids, (b) the maximum offshore wave amplitude on a representative deep water gauge, and (c) type of the earthquake source. To ensure sufficient variability in our tsunami data set, we adopt a stratified sampling approach with emphasis on extreme events (high-magnitude and low-likelihood events). We divide the data into subsets: events with and without local deformation, and further into subduction type (PS) and crustal type (BS). The sampling procedure shown in Figure 5 is described by the following steps:

1. *Deformation split*: Events are divided into two groups, Deformation Events (causing $>$ 10 cm of absolute deformation on land grids), and Non-Deformation Events (events causing $<$ 10 cm of deformation).
2. *Preserve train and test sets* Both Deformation and Non-Deformation events are randomly split into two halves for training and testing.
3. *Percentile based sampling for deformation events*:
   (a) Deformation Events are separated into two groups based on the 90th percentile of maximum onshore deformation:

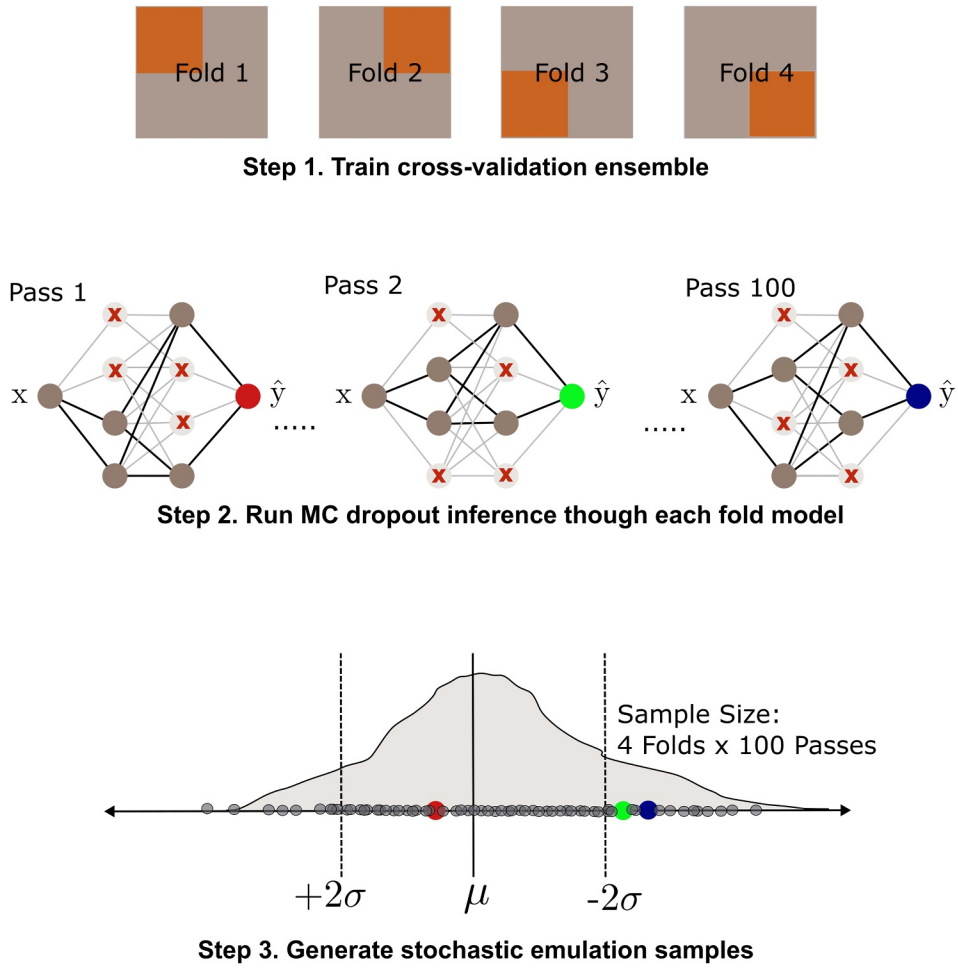

Figure 4. Schematic of the three-step stochastic emulation workflow. Step 1: Train a 4 fold cross-validation ensemble to improve generalization. Step 2: Apply Monte Carlo (MC) dropout inference across these ensemble models to estimate uncertainty. Step 3: Generate stochastic emulation samples for probabilistic prediction combining above steps.

- *Group A*: Consists of events with maximum onshore deformation *below* the 90th percentile (i.e., small and moderate)
- *Group B*: Consists of events with maximum onshore deformation *above* the 90th percentile (i.e., extremes)

(b) The events within each group are divided into 10 uniform bins by deformation magnitude, and n events are sampled from each bin.

(c) The same procedure in Step 3(a) and (b) is repeated, but using the maximum offshore wave height as the sampling parameter.

4. *Percentile based sampling for non-deformation events*:
   (a) BS Events are sampled according to their maximum offshore wave height, following Step 3(a) and (b).
   (b) PS Events are sampled in the same way as BS events, but picking $3n$ events per bin (To balance PS events and BS events).
5. *Combine events* Finally, events selected through steps 3 and 4 are merged after removing any duplicates.

We assess the accuracy of the emulator trained with different numbers of events to identify an optimal training strategy across the two regions (Catania and Siracusa) and to quantify the trade-offs between data set size, predictive performance, and uncertainty. We thus sampled $n = 2, 5, 9, 18, 42$ and 100 scenarios per bin using the above procedure, producing six increasing training sets sizes corresponding to 0.4%, 1%, 1.8%, 3.3%, 7% and 13% of the total 53,550 events referred to as the basis set. After removing duplicates, we obtained 225, 529, 892, 1658,

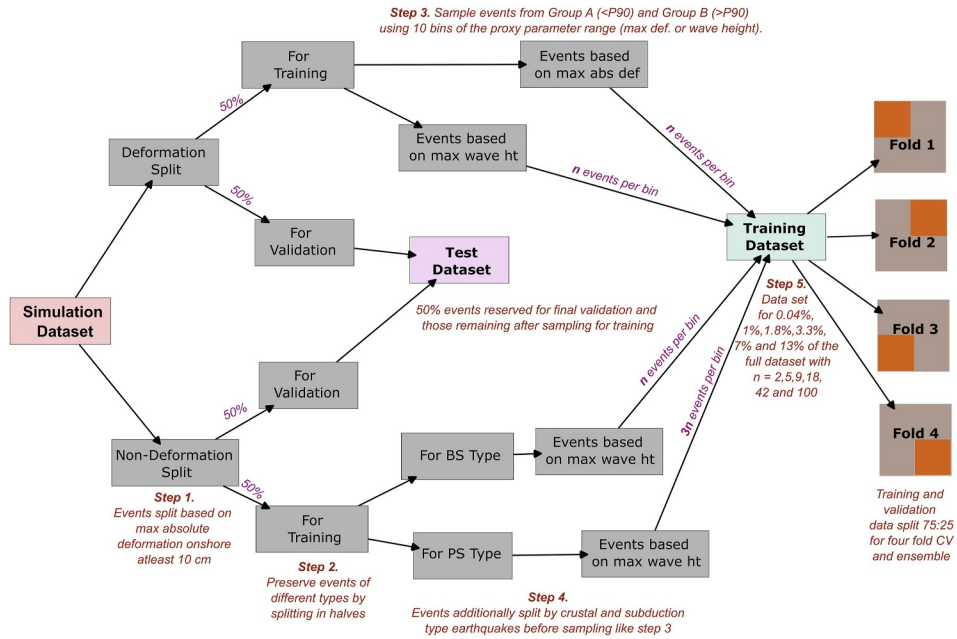

Figure 5. Flowchart illustrating the data partitioning and stratified sampling procedure used to randomly select tsunami scenarios for training, validation and final testing of the emulator. Events are divided by deformation magnitude and then further sampled based on percentiles of maximum absolute onshore deformation and maximum offshore wave height using bin-based sampling within each group. From this representative subset, a 75:25 train and validation split is implemented via four-fold cross-validation to develop the model ensemble. The remaining simulation scenarios (more than 50,000 events) from the original basis set are held out entirely to serve as a rigorous high volume final testing.

3454 and 7071 events for Catania and 228, 550, 961, 1773, 3669 and 6941 events for Siracusa, respectively. This approach captures the diversity of events, providing a robust and varied training set for our machine learning models.

Our stratified selection procedure emphasizes the extreme range of the inputs to better capture large and rare events. The training set containing about 3.3% of the total number of events will be referred to as the Moderate Training Size (MTS). This size was chosen as a reasonable number of events that can be simulated during a local tsunami hazard study and is used to evaluate the challenges we may encounter when dealing with limited training data sets. Figure 3 highlights the class imbalance of events between the basis set and the MTS training set shown in comparison for the two inundation sites. The ridge and box plots for all the training set sizes and the basis set are displayed in Figures S1 and S2 in Supporting Information S1.

2.3. Learning Framework

2.3.1. Data Structure

Beyond training set sizes, other factors can significantly affect the complexity and training of the machine learning emulator. These include the specification of input and output data, what they describe, and the level of detail. Key considerations are:

- *Input Location and Quantity*: The challenge of modeling varies according to where and how many input points are considered. For instance, inputs from deep ocean sensors (Mulia & Satake, 2020; Rabinovich & Eblé, 2015) differ in complexity from those at offshore points with 50–100 m depth or at low resolution directly onshore.
- *Temporal Information*: The duration of available data varies across applications (e.g., real-time sensor network may provide only 20–30 min of observation before prediction, whereas model based approximation can offer several hours of event data as input). This variation can influence how completely the event can be characterized.

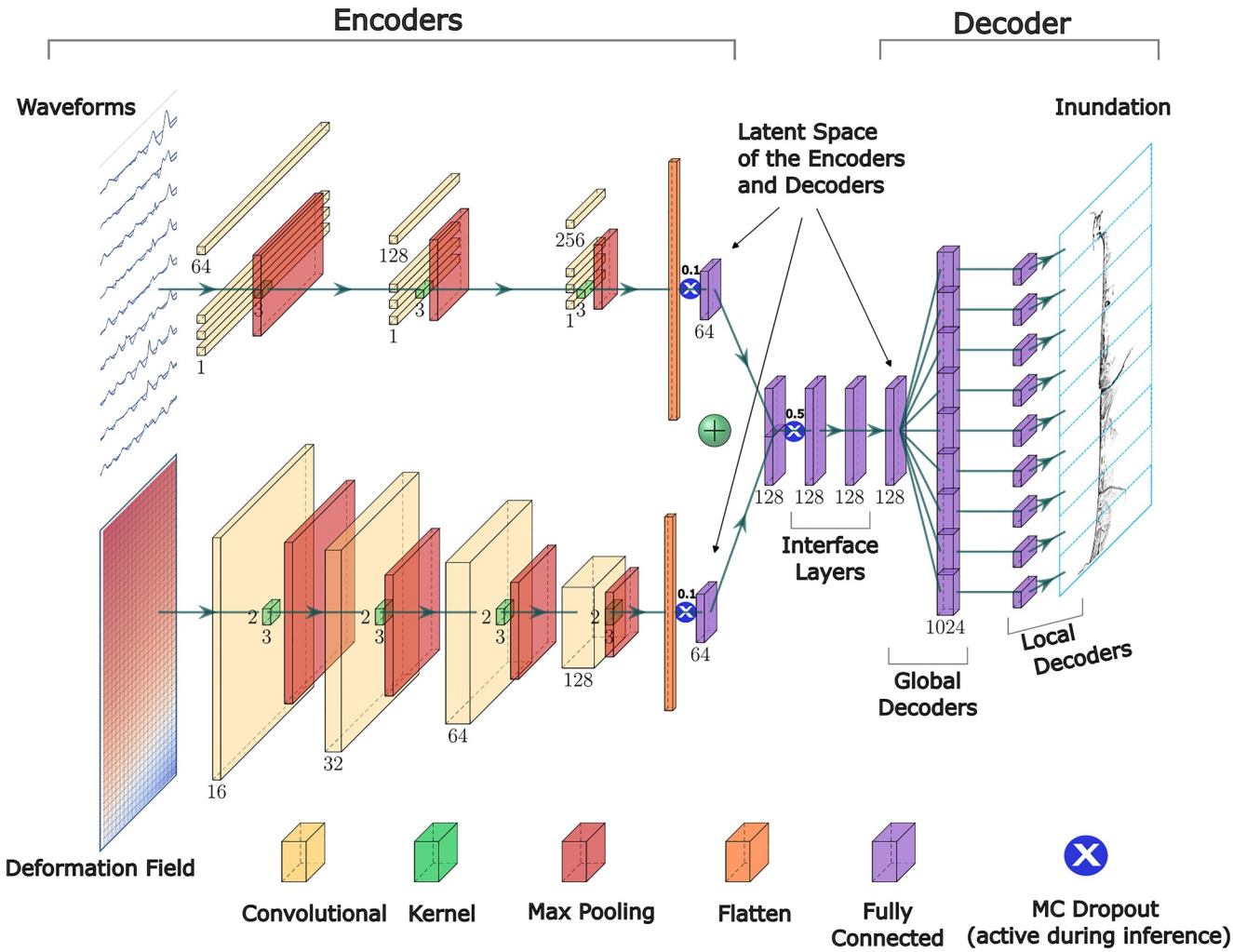

Figure 6. Model architecture of the encoder-decoder emulator presented for the Catania implementation. The flow of data is represented by the initial inputs in the form of waveforms at 9 offshore gauges and the local deformation field that pass through convolutional layers which encode and concatenate them into a combined latent space. This is decoded through linear layers of the global and local decoder, to split them into segments focusing on inundation prediction for smaller regions of Catania.

- *Prediction Locations*: We need to predefine the prediction locations, this can range from a single site of importance or multiple locations. The number of locations together with the complexity of local topography, impacts the model parameter optimization given limited training data.
- *Additional Factors*: Inundation can be driven by various mechanisms and modulated by multiple factors such as local deformation, tides, and long-term sea-level rise. In our work we account for local earthquake induced deformation. When supported by additional data streams, more factors can also be incorporated into emulation frameworks.

We process simulated tsunami waveforms consisting of 480 data points (covering the 4 hr post earthquake with 30 s sampling) from the 9 gauge sites closest to Catania and the 5 closest to Siracusa located at about 50 m depth for the respective inundation predictions shown in Figure 1a. To account for the local deformation field, we use the initial displacement available on the 10 m inundation grids with dimensions 911 × 2222 for Catania and 948 × 1300 for Siracusa. The predictions are made on a predefined subset of locations from the inundation grids, determined from the maximum inundation envelope available from the training set of size with 1.8% of the total number of events. (We are only interested in locations that are "dry" prior to the tsunami and that can conceivably be inundated.) This results in 579,439 and 224,124 prediction points for Catania and Siracusa respectively. Refer Table S2 in Supporting Information S1 for a summary on the data structure and dimension of the inputs and outputs.

2.3.2. Model Architecture

The stochastic emulator model is based on the Encoder-Decoder (ED) architecture. Encoders compress the high dimensional waveform and local deformation input data into a reduced latent space (LS) representation, while decoders reconstruct the high resolution inundation output data from this compressed form. This architecture allows the model to learn complex mappings between inputs and outputs assuming they can be effectively represented by the lower-dimensional LS, crucial for accurate inundation predictions. We used two streams of convolutional neural network layers (Gu et al., 2018), each to encode offshore waveforms and local deformation inputs. One-dimensional and two-dimensional CNNs are employed to capture temporal and spatial patterns in the offshore waveforms and deformation fields respectively. They extract hierarchical features, reducing the dimensionality of the input data, while preserving important information, that is, mapped to the LS.

The LS serves as a lower-dimensional representation of both the inputs and outputs, capturing the essential features of the tsunami waveforms, deformation, or inundation. The decoder in our emulator features a two-part structure, both of which are composed of fully connected layers with nonlinear activation functions. The first part consists of a global decoder, which processes the concatenated representations from the LS capturing the broader patterns of inundation. The second part includes multiple local decoders that independently process 64 different segments of the input flow depth map, effectively splitting it into smaller sections. This segmentation (Kim et al., 2017) helps to limit the input and output dimensions, allowing the model to focus on the precise reconstruction of smaller regions and features. The ability to split and decode multiple parts is particularly beneficial for providing attention to local spatial patterns in the inundation. By enabling specific reconstruction pathways for each region, the split network reduces the number of parameters and improves the overall accuracy and efficiency of the emulator. Figure 6 provides a detailed representation of the emulator architecture.

Several hyperparameters were configured based on initial tests, pretraining procedures and previous models (Briseid Storrøsten et al., 2024; Ragu Ramalingam et al., 2024, 2025). These include the number of channels for the convolutional and fully connected layers of the encoders and decoder layers, size of latent space, number of inundation segments, train-validation split size. Other hyperparameters like batch size, maximum training epochs, learning rate and loss function were adjusted during training iterations. We balance batch size (Smith et al., 2018) to ensure event variability for each batch given the hardware memory constraints, which typically depends on the input and output size.

Early stopping is applied, which halts training when performance on the validation set ceases to improve. The initial learning rate was an important hyperparameter, which sets the step size for parameter change during training iteration. The Adam optimizer (Kingma & Ba, 2017) is used for adaptive learning rate adjustments during training. The summary of the hyperparameter settings for the different models are made available in Table S2 in Supporting Information S1.

To prevent overfitting and improve generalization, we apply dropout (Srivastava et al., 2014), using a rate of 0.1 after the encoding layers and 0.5 after the concatenation of the encoder latent spaces (Figure 6). These values were selected based on prior modeling experience and limited trial-and-error experimentation. Although a systematic ablation study was not performed and is beyond the scope of the present work, the chosen values were found to provide a reasonable balance between regularization and predictive performance.

In addition to the regularization effect during training, we retain these dropout layers during inference to implement Monte Carlo (MC) dropout (Gal & Ghahramani, 2016), enabling stochastic emulation by generating different outputs for the same input on each forward pass (Figure 4). For each input scenario, we perform 100 such stochastic forward passes per model in the cross-validation ensemble, effectively treating the network as an approximate Bayesian ensemble that captures uncertainty in the emulator's parameters. The resulting predictions can then be summarized as a normal distribution, characterized by its mean and 2-sigma confidence bounds, providing a quantitative measure of the epistemic uncertainty in emulation.

2.3.3. Loss Function

Mean squared error (MSE) is one of the most commonly used loss function for regression tasks in ML. It averages the squared differences between predicted and actual values, thereby penalizing larger values more strongly. Misprediction may occur due to limited training data, imbalanced event types and spatial variability. This is particularly true for very small and very high flow depth values, as observed in previous studies during inference

Table 1
Summary of Evaluation Metrics and Formulas

Measure	Formula	Focus
L2Norm	$\sqrt{\sum_{i=1}^{n} y_i^2}$	Magnitude of event
MSE	$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$	Prediction error
Coefficient of Determination (R^2)	$1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$	Correlation strength
Goodness of Fit (G)	$1 - \frac{2\sum_{i=1}^{n}(y_i\,\hat{y}_i)}{\sum_{i=1}^{n} y_i^2 + \sum_{i=1}^{n} \hat{y}_i^2}$	Check for similarity
False Positive Rate (FPR)	$\frac{FP}{FP+TN}$	False inundation
False Negative Rate (FNR)	$\frac{FN}{FN+TP}$	Missed inundation
True Positive Rate (TPR)	$\frac{TP}{TP+FN}$	Detecting inundation
True Negative Rate (TNR)	$\frac{TN}{TN+FP}$	Detecting no inundation

on unseen data sets (Fukutani & Motoki, 2025; Mulia et al., 2022; Ragu Ramalingam et al., 2024). To address imbalance in training information and fit a broader range of values, we developed an asymmetric loss function that specifically penalizes under-prediction by introducing an additional weighting factor for smaller depth predictions. This is crucial in PTHA, where accurately modeling rare, high-impact tail events is essential.

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N}\begin{cases}(1+y_i)(y_i-\hat{y}_i)^2, & \text{if } \hat{y}_i < y_i,\\ (y_i-\hat{y}_i)^2, & \text{otherwise.}\end{cases} \tag{1}$$

where y_i is the true depth, and $\hat{y}_i$ is the predicted depth for the i-th sample location.

2.4. Emulation Evaluation

To assess the performance of the tsunami inundation emulators we applied several evaluation indicators whose formulation is listed in Table 1. These include the *L2Norm* for measuring the event intensity that accounts for both the area and depths of inundation, *MSE* for prediction error, the coefficient of determination (R^2) for correlation or dependence, goodness of fit (G) for scale invariant prediction error, the correlation coefficient (r) for linear relationships and use of confusion matrix scores for evaluating the binary classification if a location is flooded or not.

For a simplified evaluation, we apply the metrics (G, R^2, and *MSE*) to the mean predictions from the ensemble, restricting the analysis on grid cells with reference or predicted flow depth of at least 20 cm across all unseen events. These metrics are computed both at the scenario level and location-wise, with the summation performed accordingly over the set of scenarios or pixel locations. Additionally, we assess the misfit at control points by evaluating the model's ability to classify a flow depth threshold of 20 cm using a confusion matrix (see Figures S8 and S9 in Supporting Information S1). Finally, to quantify prediction error across the simulation data set, we calculate the cumulative *MSE* for all events with inundation affecting more than 5,000 pixels (each with 10 m grid resolution).

2.4.1. Emulation Uncertainty and Comparison With SIS for Local PTHA

A local PTHA quantifies the probability of exceedance of tsunami intensity measures, such as flow depth, at high spatial resolution to capture local variability from an ensemble of possible events. We generate probabilistic hazard curves and maps using both the full set of high-resolution tsunami simulations and the corresponding emulator predictions, based on the annual rates of each scenario. There is also uncertainty associated with the rate

of occurrence of a given scenario. This epistemic uncertainty is represented by multiple realizations or logic tree branches, indexed by $j = 1, 2, \ldots, R$, each providing an alternative estimate of the seismicity or source characteristics.

These PTHA products are computed for multiple flow depth thresholds at each grid cell, consistent with previous studies in this region (Abbate et al., 2025; Basili et al., 2021; Gibbons et al., 2020), using annual occurrence rates derived directly from the NEAMTHM18 regional hazard model (Basili et al., 2021). If $IM(\mathbf{x})$ denotes an inundation metric (e.g., the flow depth) at the location $\mathbf{x}$, and im denotes a threshold of that inundation metric, then the annual exceedance rate (AER) for this inundation metric, for the j'th scenario rate model, is given by

$$\lambda^{(j)}(IM(\mathbf{x}) \geq im) = \sum_{i=1}^{N} \lambda_i^{(j)} \cdot \mathbf{I}(IM_i(\mathbf{x}) \geq im) \tag{2}$$

where $\lambda_i^{(j)}$ is the annual occurrence rate assigned to event i from the j'th scenario rate model, and $\mathbf{I}(\cdot)$ is the indicator function, equal to 1 if the flow depth from event i at location $\mathbf{x}$ exceeds the threshold im, and 0 otherwise. To account for the full range of epistemic uncertainty, we compute exceedance rates across all $R = 1000$ realizations and extract summary statistics P_p, such as the mean and confidence intervals

$$\lambda_{P_p}(IM(\mathbf{x}) \geq im) = P_p\left\{\lambda^{(1)}(IM \geq im), \ldots, \lambda^{(R)}(IM \geq im)\right\} \tag{3}$$

Given the large simulation data set at hand, the emulation error for PTHA is assessed using post-hoc empirical uncertainty analysis (Tibshirani, 1996) applied to the mean of the stochastic emulation results. This analysis characterizes the "true" approximation error of the model by using the full set of 53,550 simulations as a ground-truth reference.

To estimate the errors, the predictions are aggregated into flow depth bins with interval limits of 0.2, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 7.5 m, and an open ended upper interval exceeding 7.5 m. For each location and depth bin, the error is calculated as the difference between the mean prediction and reference values for all predicted events falling in that bin out of the 53,550 scenarios. Because the errors were empirically observed to be approximately symmetric around the mean, they are modeled as a normal distribution. Figures S10 in Supporting Information S1 illustrates this behavior for a representative control point, showing the distribution of depth prediction errors within a given bin interval.

The resulting 1-σ bounds serve as an empirical confidence interval for the emulation error. These bounds are applied to the mean probabilistic hazard curve to derive the post-hoc uncertainty. This comprehensive evaluation serves as a benchmark to validate the internal epistemic uncertainty captured by the stochastic emulation using ensemble and MC dropout (described in Section 2.3.2). By projecting these observed residuals back onto the hazard curves (Figure 17b) at control locations, we can assess if the emulation ensemble uncertainty is consistent with its actual performance across the entire hazard and probability range.

In addition to the full simulation results, we employ SIS (Abbate et al., 2025; Davies et al., 2022) as a statistically rigorous baseline for comparison against results from our stochastic emulation framework. SIS selects a random subset of PTHA scenarios for inundation simulations using stratified sampling across magnitude bins, with importance weights defined by the arrival time, wave amplitude, and mean occurrence rate. This tends to better resolve the hazard caused by events with large waves, particularly those with near-field sources and those that are more likely to occur. The method has been empirically validated at the current study sites (Abbate et al., 2025) and is applied here to the same set of candidate PTHA scenarios available for the emulator training. A total of 1645 and 1788 scenarios are sampled as in Abbate et al. (2025) for Catania and Siracusa regions, to match the MTS (3.3% of the total simulation data set). The resulting hazard curves and exceedance maps computed with the emulator for the MTS size are compared with those from the full inundation simulation and SIS in Figures 17c and 18.

10.1029/2025JH001100

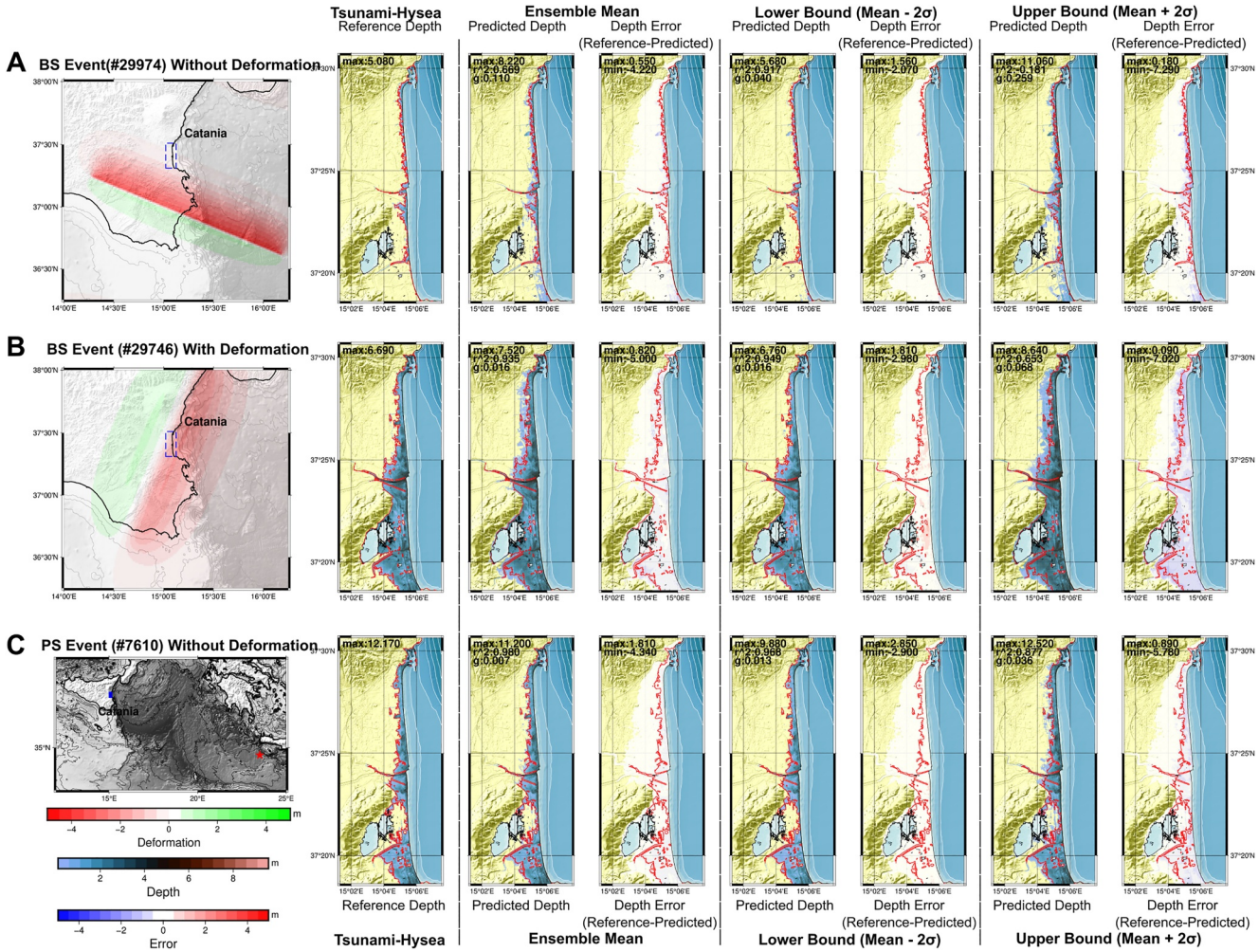

Figure 7. Sample prediction from the test set at Catania. (a) BS event without significant local deformation. (b) BS event with local deformation causing significant subsidence at the inundation site.(c) PS event on the Hellenic Arc without significant local deformation. For each event, we show the reference inundation depth computed from high-resolution inundation simulation, the corresponding emulator prediction, and the spatial differences between the two. This is repeated for the emulation sample mean and for the two extremes at ±2-sigma (lower and upper bounds). The red line outlines the inundation area from the simulation (reference). The white regions in the depth error indicate locations where the emulator either correctly predicts no inundation or accurately reproduces flow depths with minimal error. These results correspond to predictions over all for 579,439 locations of 10 m grid resolution.

3. Results

3.1. Predictive Accuracy for Unseen Events

To evaluate generalization performance, we tested our emulators trained with the MTS data for Catania and Siracusa against the full simulation data set of >50,000 scenarios. Figures 7 and 8 show representative examples of inundation depth predictions for Catania and Siracusa, using the mean of the emulation samples for events both with or without local deformation. The emulator accurately distinguishes between inundated and non-inundated zones, producing flow depth predictions that closely match those obtained from the numerical simulations. However, the emulator shows higher *MSE* for complex cases involving large local uplift or subsidence, especially when the deformation intersects and passes through the inundation zone (Figure 9). While the mean prediction shows conservative shallow inundation in marginal areas outside the true inundation extent, the upper and lower bounds capture both extent and depth variability.

A broader view of error distribution is shown in Figure 10, where event intensity (expressed as the logarithm of *L2Norm* of simulated inundation) is compared with the logarithm of *MSE*. Prediction performance is consistent

10.1029/2025JH001100

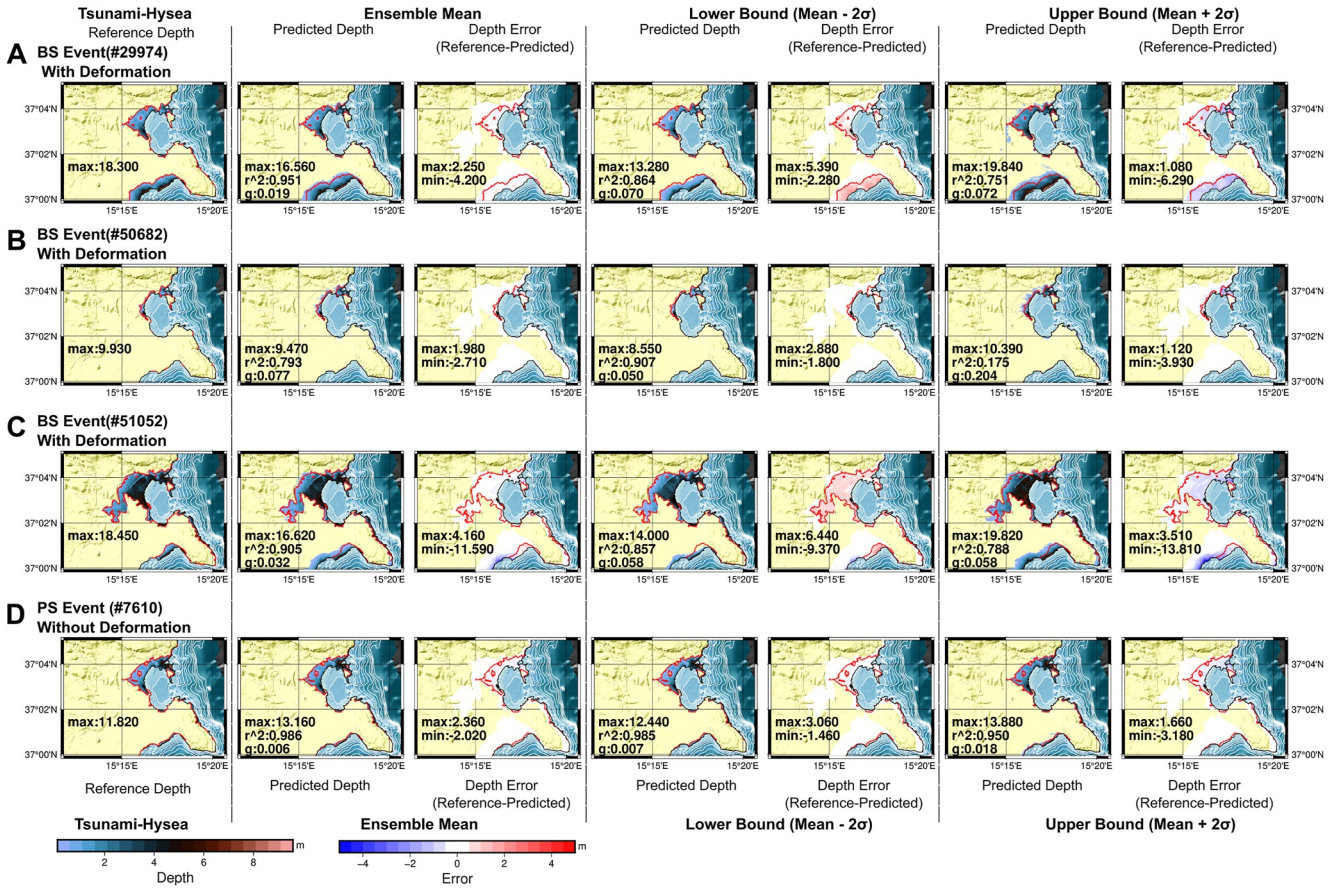

Figure 8. Similarly to Figure 7, we show sample predictions from the test set at Siracusa for four events: (a, b) BS events with significant local deformation causing subsidence and uplift. (c) BS event (#51052) with partial subsidence and uplift. (d) PS event on the Hellenic Arc without local deformation.

for events with moderate to large inundation, with *MSE* increasing approximately linearly with inundation magnitude.

However, for events with very small inundation ($\log L2Norm < -1$), a persistent baseline error is observed. This arises primarily at the inundation boundaries, where the emulator's predictions fluctuate around the wet-dry threshold. In instances where the numerical simulation yields zero depth but the emulator predicts a small depth, a residual is recorded. Because these errors are concentrated at the edge of the flooded area, they represent a conservative estimate of the inundation extent. This behavior is illustrated in Figure 11, which shows the distribution of predicted inundation depth for a no-inundation event at Site 2. The residuals reflect the model's sensitivity at the transition zone, which remains well contained within the quantified uncertainty intervals.

Further, for an emulator trained with smaller training sets, the distribution of predictions may be wide, and using the mean as the representative estimate can result in spurious small inundation depth (see Figure 11–Train size: 1%). This misinterpretation introduces bias in downstream analyses, such as hazard estimation. This highlights the benefit of considering the full distribution of emulated samples rather than relying only on their mean, particularly for smaller inundation events.

Furthermore, while Panel b of Figure 10 shows the performance over more than 50,000 scenarios, such comprehensive testing is typically not available in real-world operational settings. In practice, performance is often assessed using a small subset similar to the validation set represented by the orange markers in Panel A from the training data. This highlights the importance of ensuring representative and well-balanced testing data when deploying emulators.

10.1029/2025JH001100

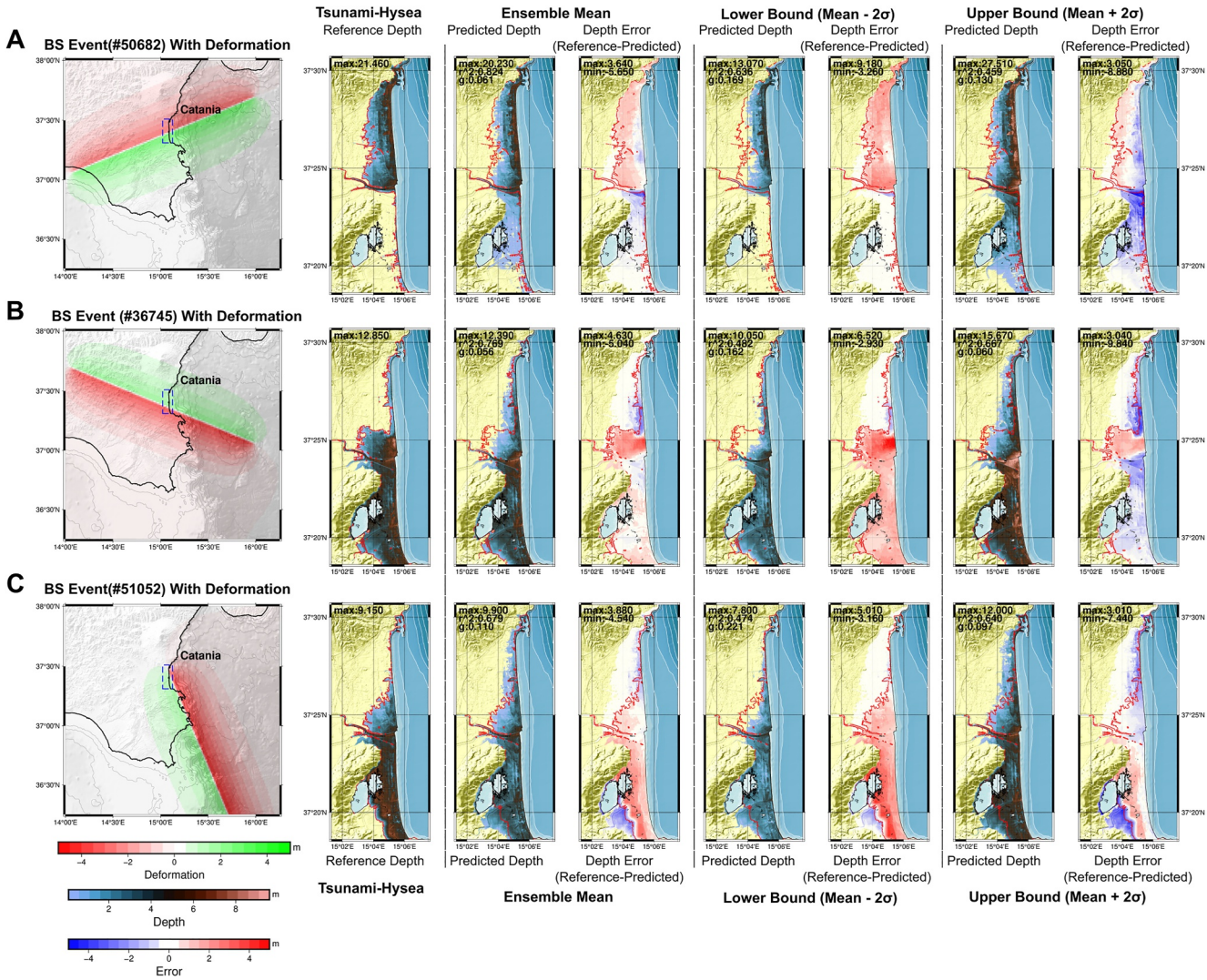

Figure 9. Prediction of BS events #50682, #36745 at Catania, both among the highest *MSE* events. These events feature partial subsidence and uplift at the inundation site. Emulators using local deformation field as additional input are able to emulate such events, but show significant misfit in the predicted depths.

3.2. Effect of Training Set Size on Model Performance

Figure 12 shows the impact of training size on model accuracy for the cross-validation ensemble, analyzing the learning curves for 25% validation portion averaged across epochs. For Catania, we observe that the emulator reaches a performance plateau with relatively few scenarios (a fraction between 1% and 3.3% of the full data set), and it improves further when larger training sizes (7%–13% of the full data set) are considered, although with decreases in returns. In contrast, for the Siracusa site, the predictive accuracy of the emulator improves consistently with increasing training size. This difference is likely due to the greater topographic complexity and resulting variability in inundation patterns for Siracusa, which benefit from a larger and more diverse training set for the emulator to generalize effectively. Interestingly, the SIS method (Abbate et al., 2025) also achieved higher accuracy and precision in representing inundation for Catania than for Siracusa when using a relatively small number of scenarios, a difference that the authors had attributed to the same cause.

Figure 13 and the full ensemble Mean squared error (train, validation and test) listed in Table 2 shows that a convergence point is reached around 7%; adding more data beyond this adds limited increased accuracy. A sharp drop in error between 1.8% and 3.3% indicates a "knee point" in the learning process, corresponding to a marked reduction in both overestimation and underestimation, particularly for smaller-magnitude events. At 3.3% MTS set, we identify a cost-effective balance between emulator performance and the computational effort required for

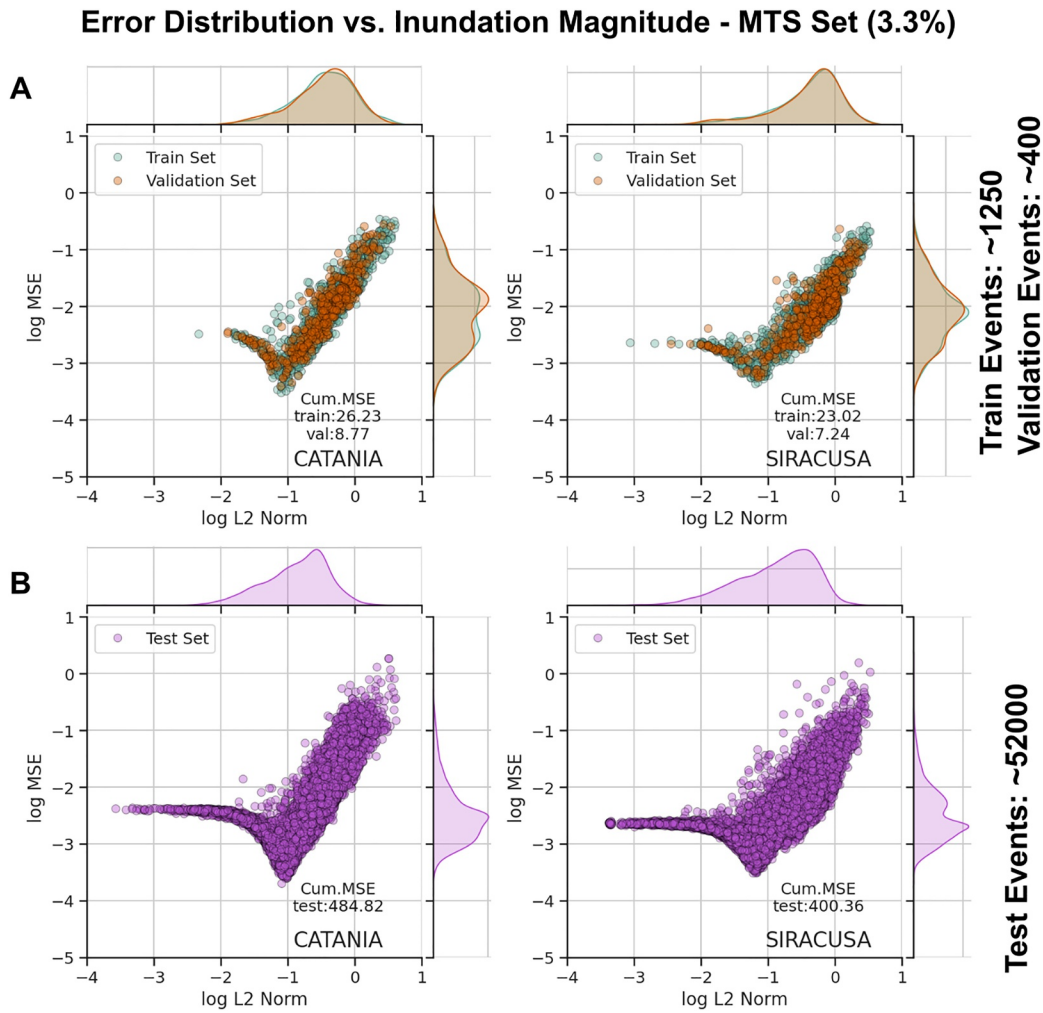

Figure 10. Performance evaluation of the Moderate Training Size Set (3.3%) emulator for the two inundation sites. (a) Scatter plot comparing the error (logarithm of Mean squared error (MSE)) against event inundation magnitudes (logarithm of the *L2Norm*). Although a CV ensemble was employed, whereby all events are utilized for training and validation. The results here are categorized here as 75% train set and 25% validation set to illustrate the performance assessment under limited data conditions during a CV fold. (b) Same as (a), but calculated for the test set consisting of all the remaining simulated events not used during ensemble training. In practice, such a large number of unseen events for evaluation is rarely available. Note: For unit consistency, 1/2 logarithm of *MSE* is directly comparable to logarithm of *L2Norm*; see Table 1 for the formulation.

training data generation, reduces false positives and overestimation, particularly for low-inundation events. Emulation quality becomes practically useful at training sizes as low as 1%, with significant performance gains achieved up to 3.3%. Additional figures illustrating the emulator's performance across all training sizes for both Catania and Siracusa are provided in Figures S3 and S4 in Supporting Information S1.

3.2.1. Sensitivity to Earthquake Origin, Source, Local Effects, Inundation Size and Prediction Location

The Goodness of fit (G) (Table 1) is used to assess the similarity between simulated and emulated inundation, with values ranging from 0 to 1. Values close to 0 indicate a very close match. The overview of G for the complete test set is shown in Figure 14a and reveals that the emulator performance is generally insensitive to any specific earthquake source location, but performs worse for very small inundation events with earthquake magnitude below M_w 7.5, offshore wave heights less than 30 cm and inundation below a threshold of 5000 pixels (out of the total 579,439 prediction locations). Tsunamis from the Cyprus source region, being very small with inundation at fewer than 3000 pixels, also fall into this category. This aligns with the minimum baseline error, suggesting that any small misprediction leads to relatively high G values for small tsunamis. We further identify all the event

10.1029/2025JH001100

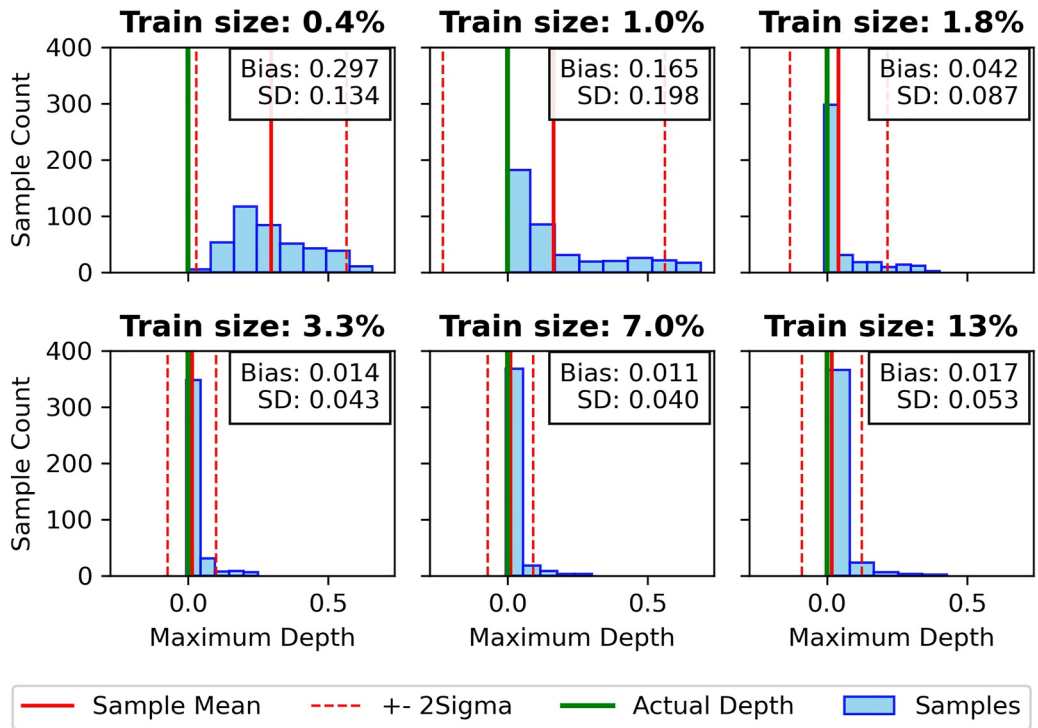

Figure 11. Distribution of predicted maximum inundation depth at Site 2 for a no-inundation event, using emulators trained with different training sizes. Smaller training sizes may lead to under fitting, resulting in overestimated predictions. To reduce storage requirements and make calculations more efficient, we approximate the samples as a normal distribution. This may distort the statistical representation and introduce some bias compared to when using all samples directly as seen in this example.

locations with poor prediction, where G is greater than 0.3 but more than 5000 pixels are inundated, which are all near-field locations (Figure 14b). This suggests that complex local deformation scenarios, where uplift or subsidence directly affects the inundation zone, are more challenging for the emulator to generalize from training data.

Figure 15 examines the performance at the grid (i.e., prediction location) level in relation to the amount of training information. The emulator tends to be accurate (a high correlation R^2) at sites where inundation is common

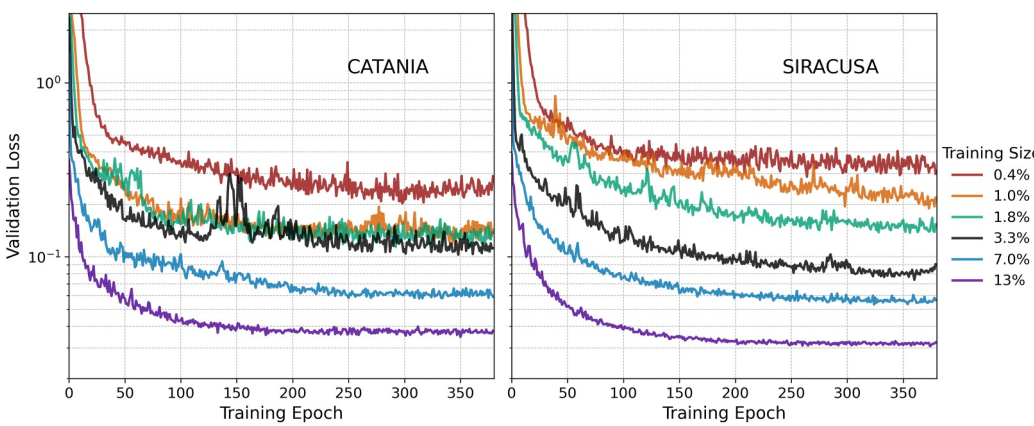

Figure 12. Comparison of validation learning curves across different training sizes for the two inundation sites. Each curve shows the mean calculated from the validation losses computed using the asymmetric loss function Equation 1 for each cross-validation fold, illustrating the learning behavior as a function of training set size. Note that with larger fraction, along with the training sets (75% of the data), the corresponding validation sets (25% of the data) also contain more samples.

10.1029/2025JH001100

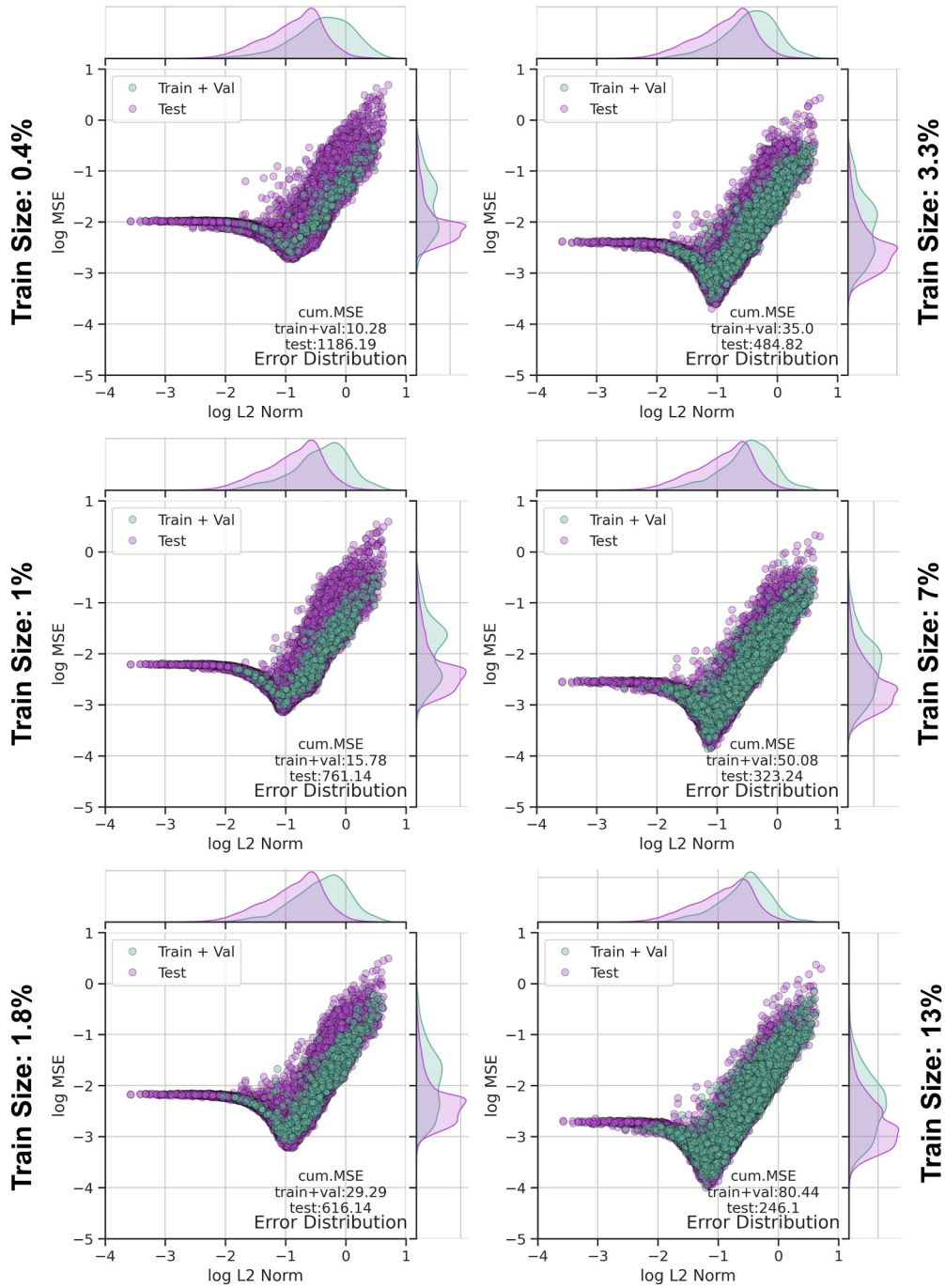

Figure 13. Scatter plot comparing event magnitude (logarithm of the *L2Norm*) with error (logarithm of Mean squared error (MSE)) for varying sampling fractions used for training Catania emulator. Training results represent the combined 75% training and 25% validation events across the CV ensemble folds, and compared against the unseen test set events remaining from the full simulation database. As training size increase, an improvement is observed with the body of the scatter histogram for the test set shifts lower on the log *MSE* scale from near −2 to −3, a lower baseline error for small magnitude events (when logarithm of the *L2Norm* is less than −1) and an overall reduction of the cumulative *MSE* for the test set.

(Panel a–b), suggesting that the local prediction accuracy is strongly dependent on the number of training scenarios inundating the given location. Panel C provides a 2D histogram (heat map) comparing R^2 scores against the frequency of flood occurrence. The trend confirms that grid cells represented by at least 30–100 training data tend to have R^2 values greater than 0.8. This reinforces the importance of spatial coverage and event diversity when

Table 2
Training Time and Full Ensemble MSE for Different Training Fractions at Catania and Siracusa

Region	Training set (%)	Training time (min)	MSE (% change)
Catania	0.4%	29	1196.47 (baseline)
	1.0%	54	776.921 (35.1%)
	1.8%	89	645.43 (46.1%)
	3.3%	190	519.82 (56.6%)
	7%	444	373.32 (68.8%)
	13%	949	326.54 (72.7%)
Siracusa	0.4%	16	1181.24 (baseline)
	1.0%	52	833.27 (29.5%)
	1.8%	98	564.75 (52.2%)
	3.3%	154	430.62 (63.5%)
	7%	309	324.24 (72.6%)
	13%	609	302.60 (74.4%)

Note. The percentage change in *MSE* compared to the baseline of training size 0.4% is shown in parentheses.

selecting training data. Results for infrequently inundated or rarely sampled locations may be used with caution and require a targeted sampling strategy to improve local prediction skills. A more detailed location-based performance breakdown is presented in Figures S6–S9 in Supporting Information S1 checking for bias across different event types (PS, BS, with or without local deformation). We find that the model generalizes well without event-type specific behavior. Misfit evaluations (Figures S8–S9 in Supporting Information S1) at the control locations shown in Figure 15, each with at least 100 training examples, show that inundation depth errors lie within 1 m between the 5th and 95th percentiles and are centered slightly below zero, reflecting the conservative tendency of the model across the prediction.

3.3. Emulating Probabilistic Tsunami Hazard Assessment (PTHA)

We evaluate the emulator's performance for a local PTHA by assessing its ability to reproduce annual exceedance rates for flow-depth thresholds using hazard curves and probabilistic maps. Aleatory uncertainty represented by the full event ensemble and epistemic uncertainty from the earthquake source logic tree are inherited directly from the regional hazard model (Basili et al., 2021) providing the annual occurrence rates for each scenario in the ensemble. Emulation uncertainty is modeled using the stochastic emulation procedure (see Section 2.4.1).

Figure 16 shows the predicted AER for a 0.5 m inundation threshold at POI 2 in Catania across emulators using different training set sizes. The AER, computed using the ensemble mean of the depth predictions, is plotted along with ±2-sigma intervals, representing the confidence bounds of the emulator. As the training size increases, the interval ranges narrow and closer to the true estimate, indicating a reduction in both bias and uncertainty. For further evaluation and comparison, the PTHA emulation results are based on the emulator trained using the MTS Set at 3.3%, as this represents a plausible and generally feasible training size for PTHA applications. Results for other control point locations across Catania and Siracusa are reported in Figure S11 and S12 in Supporting Information S1.

Neural-network instability due to noise or small perturbations in input data has been identified as a potential concern in tsunami prediction applications (Rim et al., 2024). Our use of convolutional encoders, ensemble-based modeling, and Monte Carlo dropout during inference is intended to help mitigate sensitivity to such perturbations and provide an indication of predictive uncertainty. While these techniques cannot eliminate the possibility of adversarial vulnerabilities or instability effects, the resulting variability is explicitly propagated into the predicted uncertainty bounds of the hazard estimate. As a result, instabilities may be potentially identified rather than being overlooked.

3.3.1. Hazard Curve Comparison at Control Points

We applied the emulator to compute hazard curves at two control locations in Catania, as shown in Figure 17. Panel A shows the impact of training size on hazard curve accuracy, strong convergence is observed across all the training sizes, with all curves approaching the reference hazard curve computed using the full simulation data set (solid line). Panel B compares the epistemic uncertainty captured by the stochastic emulation approach against the post-hoc uncertainty estimate. The two uncertainty estimates are in good agreement, with emulation having slightly larger uncertainty in the rarer tail of the distribution, which is appropriate for conservative hazard modeling. For context, we also show the confidence bounds (5th–95th and 16th–84th percentiles) of the reference hazard curve obtained from the seismic rate ensemble, indicating that the emulator's uncertainty remains within acceptable bounds and is smaller than the uncertainty inherent from the seismic source model. Panel C compares emulator-based hazard curves to those obtained using SIS, after sampling approximately the same number of events as used in the emulator's training. While SIS provides an unbiased estimate on average and is relatively more accurate for a broad range of probabilities, it exhibits less reliability and greater variability at very low probabilities (high return period), particularly at inland sites for return periods >100,000 years. This highlights the benefit of using an emulator, which even with limited training sizes generalizes better in sparsely sampled

10.1029/2025JH001100

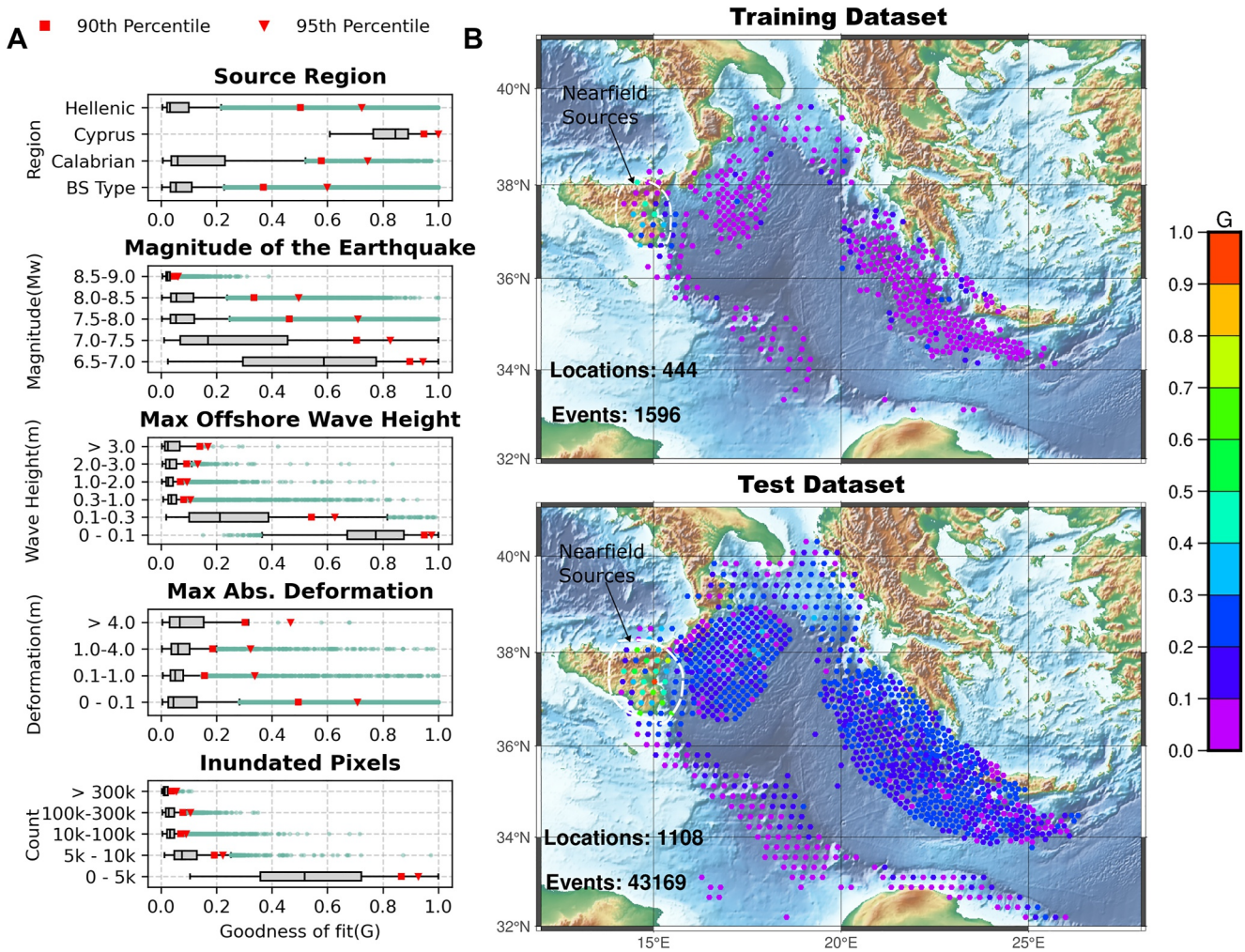

Figure 14. (a) Prediction performance across different earthquake and tsunami parameters for Catania. (b) Map of earthquake origin locations to their performance, only events greater than 10 cm in the offshore gauge and inundation greater than 5,000 pixels are shown for the training set of size 3.3% and the test set. When identifying events with poor performance using $G > 0.3$, all filtered events in the training and test data set are found to be associated with near-field or local sources marked by dashed lines (about 90 events at 27 locations for the test set), indicating that these complex local events require further representation in the training set.

regions and can be used to capture correlation across events and locations. Results for other control point locations across Catania and Siracusa are reported in Figures S13–S15 in Supporting Information S1. These locations each have at least 100–1000 flood occurrences represented in the training set, spanning a range of inundation conditions.

3.3.2. Hazard Map Comparison and Implications

Figure 18 compares PTHA maps computed for Catania region from (a) the full numerical simulation of 53,550 events, (b) the emulator predictions using 3.3% MTS for training, and (c) SIS-based results with equivalent sampling size of about the 3.3% MTS set. The hazard maps classify areas by the annual probability of exceeding depth thresholds of 20 cm, 1 m, 3 m, and 5 m per year, and include probability contours for 10^{-4}, 10^{-5}, 10^{-6}, and 10^{-7} from the full simulation results for ease of comparison. Visual inspection of the maps and associated errors (Figure S17 in Supporting Information S1) indicates close agreement of both the reduced-cost approaches with the aggregated results from the full simulation based PTHA. This comparison provides additional insight into the quality of predictions in regions with limited flood occurrences available for emulator training or sampling, which were not covered by the control locations in the hazard curve analysis of the previous section. The overall extent and gradation of inundation probabilities across the Catania region are well captured, with minor discrepancies

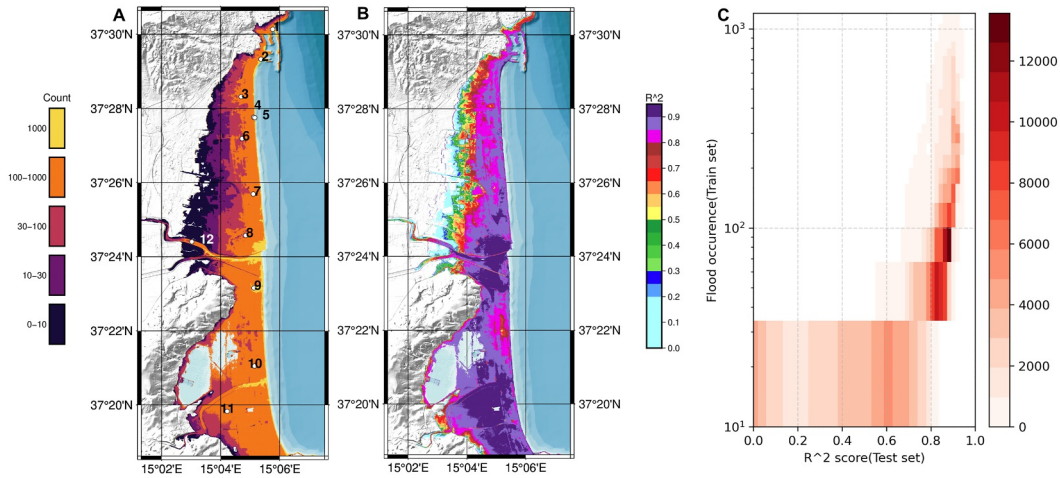

Figure 15. (a) Count of flood occurrence in the training data set at prediction grid. (b) Prediction performance at the inundation grid, quantified by R^2, using the test set. (c) Heat map showing the relationship between R^2 for the all the test events and the number of flood occurrence in the training events.

primarily occurring in inland areas characterized by very low inundation probabilities (marked pink for AER $< 10^{-8}$).

At small thresholds (20 cm and 1 m), the emulator-based hazard maps display spatially coherent patterns with mild underestimation towards the inland limits of inundation. In contrast, the SIS-based maps show some local variability, characterized by alternating zones of slight over- and underestimation. At the higher threshold of 5 m, deviations associated with the extremely low probability of inundation become evident, with the emulator results showing clear underestimation inland. Similar spatial patterns are observed for the Siracusa region (Figures S16 and S18 in Supporting Information S1), but the SIS results exhibit a comparatively stronger tendency to overestimate the probability of exceedance, while the emulator continues to show mild underestimation along the edges of the inundation zones.

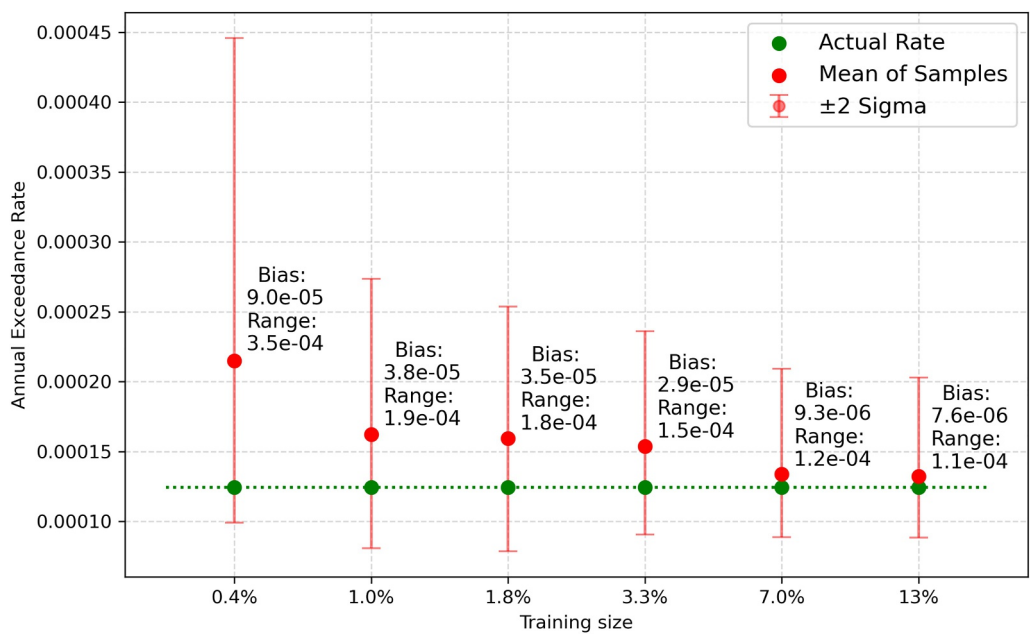

Figure 16. The annual rate of exceedance for an inundation depth of 0.5 m at POI 2. The estimated annual exceedance-rate using the mean of the emulation ensemble with different training sizes shows reduction in the range of the confidence interval bounds given by mean ±2-sigma and converging between the 3.3%–7% training sizes.

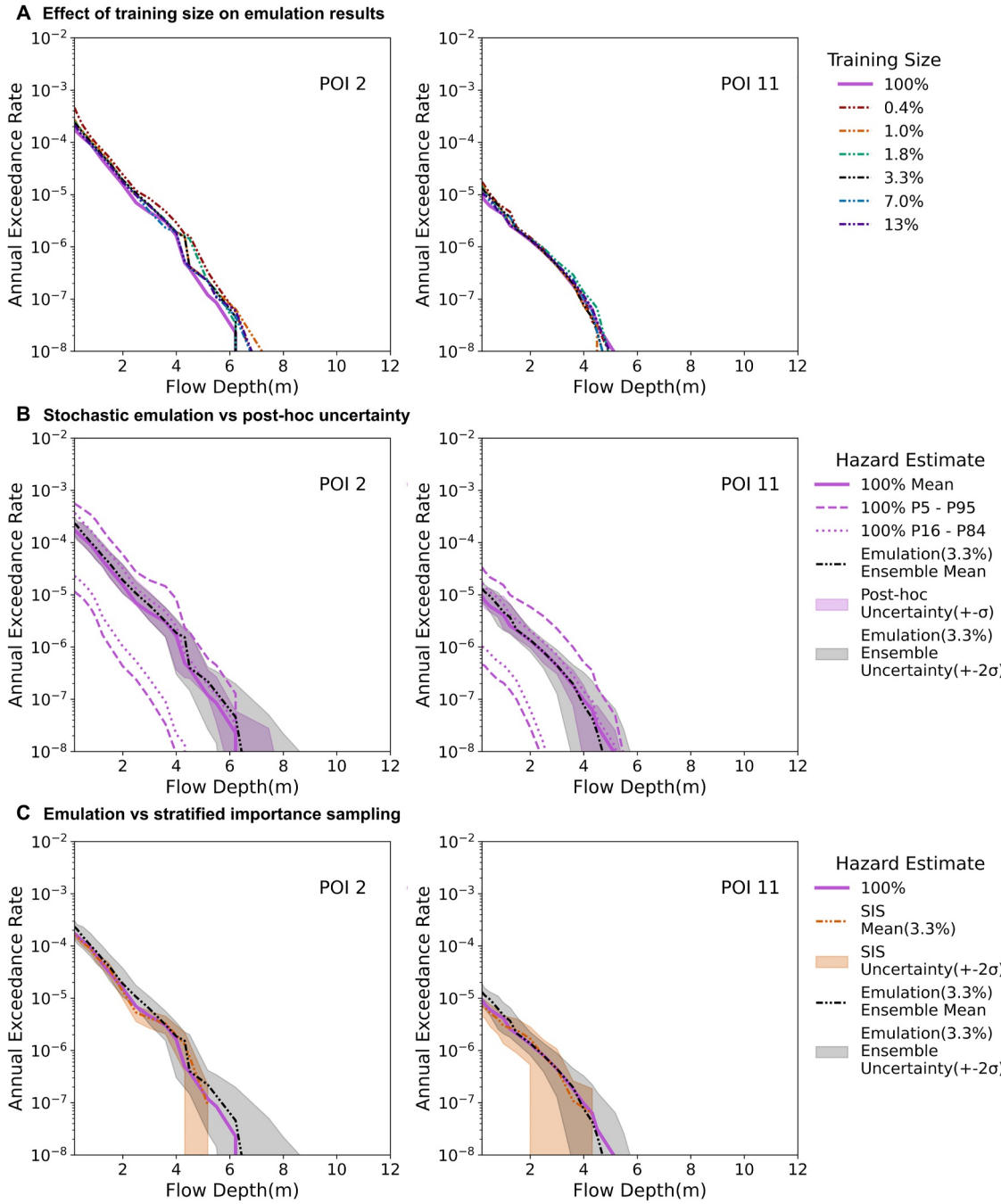

Figure 17. Probabilistic Hazard Curve at control POI 2 and 11 of Catania (see Figure 15 for locations). (a) Emulation results using ensemble mean for different training sizes (0.4%–13% of the data set), showing the effect of training data quantity on hazard curve approximation. Solid line represents results using full simulation data set. (b) Comparison of emulation uncertainty and post-hoc uncertainty estimates when trained with 3.3% of the data. For reference, the confidence interval of the epistemic uncertainty in the seismic rates (5th/95th and 16th/84th percentiles) is also shown, indicating that the uncertainty from the emulator is smaller than that inherent in the seismic rate model. (c) Comparison between emulation results and stratified importance sampling (SIS) results. The apparent divergence in the extreme tail for the SIS results arises from truncation at the highest sampled scenario.

The emulator's strength lies in its ability to generalize from limited training data and to reproduce both coastal and inland low-probability inundation zones. In contrast, SIS provides unbiased mean estimates across a broad range of return periods but may exhibit higher variance and reduced stability in sparsely sampled regions or inundation ranges. The SIS follows Abbate et al. (2025) in emphasizing scenarios with larger offshore waves that are closer to the target site and have larger annual occurrence rates, while optimizing the magnitude bin sampling for

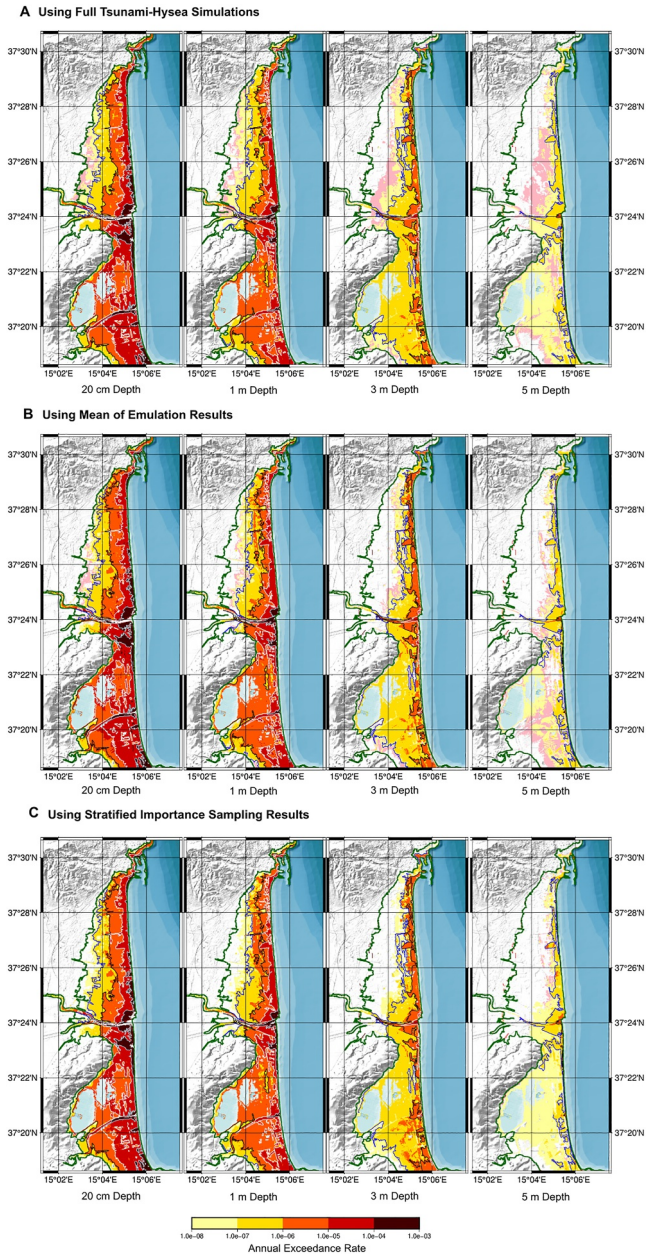

Figure 18. Comparison of the different probability maps for Catania using (a) the full inundation simulation with Tsunami-Hysea for 53,550 events, (b) the mean predictions of the emulator and (c) the results from Stratified Importance Sampling. Exceedance-rate zones are defined by the annual probability of exceeding inundation depths of 20 cm, 1 m, 3 m, and 5 m. Light pink regions marks areas with annual exceedance rate below 10^{-8}. Green boundary marks the predefined region of emulator prediction. Hazard contours corresponding to exceedance probabilities of 10^{-4}, 10^{-5}, 10^{-6}, and 10^{-7} from panel (a) are overlaid on panel (b) and (c) using light blue, white, black, and blue lines respectively for easy comparison.

offshore wave heights from 0.5 to 5.0 m. This tends to under-represent the few very large, rare scenarios that require more targeted treatment. The SIS accuracy could be increased by sampling more scenarios, using additional offshore sites or by potentially redesigning the importance function to better focus on extremes (e.g., leveraging the local deformation fields).

4. Conclusions

This study presents a machine learning-based stochastic emulator designed to efficiently approximate tsunami inundation across a wide range of earthquake and tsunami scenarios for PTHA. The emulator directly incorporates co-seismic deformation inputs and is capable of predicting inundation with quantified uncertainty, offering a scalable and site-adaptable approach for tsunami hazard modeling.

We evaluated the emulator using high-resolution simulation data sets for two distinct coastal regions in Sicily at Catania and Siracusa demonstrating its accuracy and cost-effectiveness. Although the simulation data set was not originally designed for emulator training, the emulator prediction were consistent across a wide range of events. For interface-type earthquake events, which may play a critical role in regions with strong near-field seismicity and complex co-seismic deformation patterns, increasing the representation of such scenarios in training would improve emulator performance by using targeted stratified sampling that prioritizes interface-intersection events, along with testing the sensitivity of the alternative deformation threshold values(set as 10 cm in this work) should be explicitly considered in the design of future training data sets.

Training with just 1%–3.3% of the simulation scenarios produced reliable PTHA outputs, with only marginal gains beyond 7%. In practice, when a precomputed database is not available, a modeler can determine the optimal training size by iteratively increasing the simulation set and testing for convergence on a small pilot set. This process should continue until the reduction in predictive uncertainty meets the specific requirements of the project and its computational constraints. Comparison of emulator-based results against full numerical simulations and SIS confirms its robustness and the conservative handling of uncertainty. The benefit of emulation is most pronounced for extreme inundation cases, where sampling approaches might under-represent low-probability events that affect far inland or rarely inundated zones. Table 3 summarizes the computational savings, with emulation offering a speed-up of 30–200 times compared to full numerical simulation. All machine learning computations including training and inference, were performed on an NVIDIA A100 80 GB GPU.

Importantly, the learning curve analysis highlights notable differences in the training regime for the two inundation sites. While emulation performance converges early for Catania, improvements for Siracusa persist with increasing data, attributed to its more complex coastal morphology. This indicates that optimal training sizes are also location-sensitive and depend on both the variability of scenarios used as inputs, the local spatial characteristics of inundation and other regional effects. For example, locations like Siracusa, dividing the prediction grid spatially (e.g., south and east sectors) could improve emulator generalization by reducing regional complexity. Further robustness to noise and small input perturbations causing neural network instability may be achieved through explicit adversarial testing and the incorporation of input filtering or noise augmentation during training. While the present work primarily focuses on the mean results for the sake of brevity and to establish the baseline performance across training sizes, the

Table 3
Overview of Run Times for Numerical Simulation Using Tsunami-Hysea and Emulation for Generating probabilistic tsunami Hazard Assessment Data for Both Regions of Catania and Siracusa

Region	Method	# Events	Run time (min)
Catania	Simulation (1)	1	25
+ Siracusa	Simulation (Training)	0.4%	5,625
	Simulation (Training)	3.3%	44,325
	Simulation (53,550)	100%	**1,338,750**
Catania	Stochastic Emulation	100%	660
Siracusa	Stochastic Emulation	100%	320
Emulation Time		0.4%	**6,650**
(Data Simulation + Training + Prediction)		3.3%	**45,649**
Overall Gain from Emulation			**30x – 200x**

Note. Emulation time consists of time needed for data generation, training and prediction. Overall efficiency gain is calculated by dividing the simulation time for all 53,550 scenarios by that of the emulation highlighted in bold.

integration of the full stochastic emulation samples or specific quantiles remains a key feature of our framework that can be applied for more refined hazard characterization.

Our findings also suggest the complementary use of emulators and SIS within a hybrid PTHA strategy. Emulator predictions, even when sparsely trained, can guide SIS to better sample under-represented inland scenarios, while SIS estimates can correct emulator outputs in regions well-characterized from offshore proxies. This integration could significantly enhance the coverage, accuracy, and efficiency of future PTHA workflows.

While this work focused on uncertainties due to training size and seismic source representation, further extensions should incorporate other sources of aleatory and epistemic uncertainty, such as those from unknown faults and sources, topography, bathymetry, tidal phase, sea level rise, and structural defenses (Alhamid et al., 2022; Fukutani et al., 2023; Sepúlveda et al., 2020). Our framework, being the first to demonstrate full site-specific stochastic emulation for PTHA, provides a strong foundation to integrate such additional factors.

Finally, the emulator's improved performance across a wide range of inundation intensities marks a significant step forward relative to earlier studies, that focused on a limited set of earthquake source regions and magnitude ranges, and reported underestimation of extreme events (Fukutani & Motoki, 2025; Mulia et al., 2022). However, we acknowledge that further testing across alternative source configurations and geographic settings is required to draw more general conclusions. Future refinements, such as incorporating weighted loss functions or pre-training strategies, could further reduce biases for smaller events and enhance reliability across the entire hazard spectrum. The emulation framework presented in this work can play a vital role in advancing rapid, PTHA needed for many unmodelled regions. At the same time, it may enhance prediction systems for decision-making such as early warning.

Acknowledgments
NRR was supported by a doctoral research scholarship under the "Dipartimenti di Eccellenza" project, funded by the Italian Ministry of Education, University and Research (MIUR) at IUSS Pavia. Part of this work was conducted during a mobility period to NGI, Oslo, funded by the Erasmus + programme, and later as a postdoctoral researcher within the PCTWIN project, jointly funded by the Natural Environment Research Council (NERC), UKRI and the Ministry of Earth Sciences (MoES), Government of India (Grant: NE/Z503496/1). The work presented here includes machine learning computations carried out on the High Performance Computing DataCenter at IUSS, co-funded by Regione Lombardia through the funding programme established by Regional Decree No. 3776 of 3 November 2020. The development and testing of the machine learning emulators was supported by "A Digital Twin for Geophysical Extremes" (DT-GEO) and received funding from Horizon Europe under Grant Agreement No 101058129. We would also like to thank Luigi Cesarini for discussions and support during the development and testing of the machine learning model. GD publishes with the permission of the CEO, Geoscience Australia.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Availability Statement

The processed tsunami simulation data used for machine learning training and testing in this study are available at the Zenodo repository (Ragu Ramalingam, 2024a, 2024b). The code required to reproduce the result in the article is made available through the Zenodo repository (Ragu Ramalingam, 2026).

References

Abbate, A., Davies, G., Lorito, S., Kalligeris, N., Romano, F., Tonini, R., & Volpe, M. (2025). Importance sampling of seismic tsunami sources with near-field emphasis for inundation PTHA: Benchmarking with complete ensembles. *Geophysical Journal International*, *241*(1), 155–169. https://doi.org/10.1093/gji/ggaf034

Alhamid, A. K., Akiyama, M., Ishibashi, H., Aoki, K., Koshimura, S., & Frangopol, D. M. (2022). Framework for probabilistic tsunami hazard assessment considering the effects of sea-level rise due to climate change. *Structural Safety*, *94*, 102152. https://doi.org/10.1016/j.strusafe.2021.102152

Baba, T., Kamiya, M., Tanaka, N., Sumida, Y., Yamanaka, R., Watanabe, K., & Fujiwara, H. (2022). Probabilistic tsunami hazard assessment based on the Gutenberg–Richter law in eastern Shikoku, Nankai subduction zone, Japan. *Earth Planets and Space*, *74*(1), 156. https://doi.org/10.1186/s40623-022-01715-1

Basili, R., Brizuela, B., Herrero, A., Iqbal, S., Lorito, S., Maesano, F. E., et al. (2021). The making of the NEAM Tsunami Hazard Model 2018 (NEAMTHM18). *Frontiers in Earth Science*, *8*, 616594. https://doi.org/10.3389/feart.2020.616594

Behrens, J., Løvholt, F., Sørensen, M., Lorito, S., Rossetto, T., Murphy, S., et al. (2021). Probabilistic tsunami hazard and risk analysis - A review. *on Research Gaps*, *9*(April), 1–28. https://doi.org/10.3389/feart.2021.628772

Bressan, L., & Tinti, S. (2016). Statistical properties of coastal long waves analysed through sea-level time-gradient functions: Exemplary analysis of the Siracusa, Italy, tide-gauge data. *Natural Hazards and Earth System Sciences*, *16*(1), 223–237. https://doi.org/10.5194/nhess-16-223-2016

Briseid Storrøsten, E., Ragu Ramalingam, N., Lorito, S., Volpe, M., Sánchez-Linares, C., Løvholt, F., & Gibbons, S. J. (2024). Machine learning emulation of high resolution inundation maps. *Geophysical Journal International*, ggae151. https://doi.org/10.1093/gji/ggae151

Davies, G., Weber, R., Wilson, K., & Cummins, P. (2022). From offshore to onshore probabilistic tsunami hazard assessment via efficient Monte Carlo sampling. *Geophysical Journal International*, *230*(3), 1630–1651. https://doi.org/10.1093/gji/ggac140

de la Asunción, M. (2024). Prediction of tsunami alert levels using deep learning. *Earth and Space Science*, *11*(3), e2023EA003385. https://doi.org/10.1029/2023EA003385

de la Asunción, M., Castro, M. J., Fernández-Nieto, E. D., Mantas, J. M., Acosta, S. O., & González-Vida, J. M. (2013). Efficient GPU implementation of a two waves TVD-WAF method for the two-dimensional one layer shallow water system on structured meshes. *Computers & Fluids*, *80*, 441–452. https://doi.org/10.1016/j.compfluid.2012.01.012

Fauzi, A., & Mizutani, N. (2020). Machine learning algorithms for real-time tsunami inundation forecasting: A case Study in nankai Region. *Pure and Applied Geophysics*, *177*(3), 1437–1450. https://doi.org/10.1007/s00024-019-02364-4

Folch, A., Abril, C., Afanasiev, M., Amati, G., Bader, M., Badia, R. M., et al. (2023). The EU Center of Excellence for Exascale in Solid Earth (ChEESE): Implementation, results, and roadmap for the second phase. *Future Generation Computer Systems*, *146*, 47–61. https://doi.org/10.1016/j.future.2023.04.006

Fritz, H. M., Petroff, C. M., Catalán, P. A., Cienfuegos, R., Winckler, P., Kalligeris, N., et al. (2011). Field Survey of the 27 February 2010 Chile Tsunami. *Pure and Applied Geophysics*, *168*(11), 1989–2010. https://doi.org/10.1007/s00024-011-0283-5

Fukutani, Y., & Motoki, M. (2025). A neural network-based surrogate model for efficient probabilistic tsunami inundation assessment. *Coastal Engineering*, *200*, 104767. https://doi.org/10.1016/j.coastaleng.2025.104767

Fukutani, Y., Yasuda, T., & Yamanaka, R. (2023). Efficient probabilistic prediction of tsunami inundation considering random tsunami sources and the failure probability of seawalls. *Stochastic Environmental Research and Risk Assessment*, *37*(6), 2053–2068. https://doi.org/10.1007/s00477-023-02379-3

Gal, Y., & Ghahramani, Z. (2016). Dropout as a bayesian approximation: Representing model uncertainty in deep learning. *arXiv*. https://doi.org/10.48550/arXiv.1506.02142

Geist, E. L., & Parsons, T. (2006). Probabilistic analysis of tsunami hazards. *Natural Hazards*, *37*(3), 277–314. https://doi.org/10.1007/s11069-005-4646-z

Ghobarah, A., Saatcioglu, M., & Nistor, I. (2006). The impact of the 26 December 2004 earthquake and tsunami on structures and infrastructure. *Engineering Structures*, *28*(2), 312–326. https://doi.org/10.1016/j.engstruct.2005.09.028

Gibbons, S. J., Lorito, S., de la Asunción, M., Volpe, M., Selva, J., Macías, J., et al. (2022). The sensitivity of tsunami impact to earthquake source parameters and manning friction in high-resolution inundation simulations. *Frontiers in Earth Science*, *9*(January), 1–23. https://doi.org/10.3389/feart.2021.757618

Gibbons, S. J., Lorito, S., Macías, J., Løvholt, F., Selva, J., Volpe, M., et al. (2020). Probabilistic tsunami hazard analysis: High performance computing for massive Scale inundation simulations. *Frontiers in Earth Science*, *8*(December), 1–20. https://doi.org/10.3389/feart.2020.591549

Goda, K., Orchiston, K., Borozan, J., Novakovic, M., & Yenier, E. (2023). Evaluation of reduced computational approaches to assessment of tsunami hazard and loss using stochastic source models: Case study for Tofino, British Columbia, Canada, subjected to Cascadia megathrust earthquakes. *Earthquake Spectra*, *39*(3), 1303–1327. https://doi.org/10.1177/87552930231187407

Grezio, A., Babeyko, A., Baptista, M. A., Behrens, J., Costa, A., Davies, G., et al. (2017). Probabilistic tsunami hazard analysis: Multiple sources and global applications. *Reviews of Geophysics*, *55*(4), 1158–1198. https://doi.org/10.1002/2017RG000579

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., et al. (2018). Recent advances in convolutional neural networks. *Pattern Recognition*, *77*, 354–377. https://doi.org/10.1016/j.patcog.2017.10.013

Heidarzadeh, M., Ishibe, T., Gusman, A. R., & Miyazaki, H. (2024). Field surveys of tsunami runup and damage following the January 2024 Mw 7.5 Noto (Japan sea) tsunamigenic earthquake. *Ocean Engineering*, *307*, 118140. https://doi.org/10.1016/j.oceaneng.2024.118140

Kamiya, M., Igarashi, Y., Okada, M., & Baba, T. (2022). Numerical experiments on tsunami flow depth prediction for clustered areas using regression and machine learning models. *Earth Planets and Space*, *74*(1), 127. https://doi.org/10.1186/s40623-022-01680-9

Kim, J., Park, Y., Kim, G., & Hwang, S. J. (2017). SplitNet: Learning to semantically split deep networks for parameter reduction and model parallelization. In D. Precup & Y. W. Teh (Eds.), *Proceedings of the 34th international conference on machine learning* (Vol. 70, pp. 1866–1874). PMLR. Retrieved from https://proceedings.mlr.press/v70/kim17b.html

Kingma, D. P., & Ba, J. (2017). Adam: A method for stochastic optimization. *arXiv*. Retrieved 2024-08-29, from http://arxiv.org/abs/1412.6980 (arXiv:1412.6980 [cs]).

Liu, C. M., Rim, D., Baraldi, R., & LeVeque, R. J. (2021). Comparison of machine learning approaches for tsunami forecasting from sparse observations. *Pure and Applied Geophysics*, *178*(12), 5129–5153. https://doi.org/10.1007/s00024-021-02841-9

Lorito, S., Selva, J., Basili, R., Romano, F., Tiberti, M., & Piatanesi, A. (2015). Probabilistic hazard for seismically induced tsunamis: Accuracy and feasibility of inundation maps. *Geophysical Journal International*, *200*(1), 574–588. https://doi.org/10.1093/gji/ggu408

Makinoshima, F., Oishi, Y., Yamazaki, T., Furumura, T., & Imamura, F. (2021). Early forecasting of tsunami inundation from tsunami and geodetic observation data with convolutional neural networks. *Nature Communications*, *12*(1), 2253. https://doi.org/10.1038/s41467-021-22348-0

Molinari, I., Tonini, R., Lorito, S., Piatanesi, A., Romano, F., Melini, D., et al. (2016). Fast evaluation of tsunami scenarios: Uncertainty assessment for a Mediterranean Sea database. *Natural Hazards and Earth System Sciences*, *16*(12), 2593–2602. https://doi.org/10.5194/nhess-16-2593-2016

Mori, N., Takahashi, T., Yasuda, T., & Yanagisawa, H. (2011). Survey of 2011 Tohoku earthquake tsunami inundation and run-up. *Geophysical Research Letters*, *38*(7). https://doi.org/10.1029/2011GL049210

Mulia, I. E., & Satake, K. (2020). Developments of tsunami observing systems in Japan. *Frontiers in Earth Science*, *8*, 145. https://doi.org/10.3389/feart.2020.00145

Mulia, I. E., Ueda, N., Miyoshi, T., Gusman, A. R., & Satake, K. (2022). Machine learning-based tsunami inundation prediction derived from offshore observations. *Nature Communications*, *13*(1), 5489. https://doi.org/10.1038/s41467-022-33253-5

Nemani, V., Biggio, L., Huan, X., Hu, Z., Fink, O., Tran, A., et al. (2023). Uncertainty quantification in machine learning for engineering design and health prognostics: A tutorial. *Mechanical Systems and Signal Processing*, *205*, 110796. https://doi.org/10.1016/j.ymssp.2023.110796

Núñez, J., Catalán, P. A., Valle, C., Zamora, N., & Valderrama, A. (2022). Discriminating the occurrence of inundation in tsunami early warning with one-dimensional convolutional neural networks. *Scientific Reports*, *12*(1), 1–20. https://doi.org/10.1038/s41598-022-13788-9

Rabinovich, A. B., & Eblé, M. C. (2015). Deep-Ocean measurements of tsunami waves. *Pure and Applied Geophysics*, *172*(12), 3281–3312. https://doi.org/10.1007/s00024-015-1058-1

Ragu Ramalingam, N. (2024a). Processed dataset used for developing tsunami inundation emulators testing. *Zenodo*. https://doi.org/10.5281/zenodo.13741284

Ragu Ramalingam, N. (2024b). Processed dataset used for developing tsunami inundation emulators training. *Zenodo*. https://doi.org/10.5281/zenodo.13738078

Ragu Ramalingam, N. (2026). Ml4sicilytsunami: Code submission to JGR: Machine learning and computation. *Zenodo*. https://doi.org/10.5281/zenodo.19400308

Ragu Ramalingam, N., Johnson, K., Pagani, M., & Martina, M. L. V. (2025). Advancing nearshore and onshore tsunami hazard approximation with machine learning surrogates. *Natural Hazards and Earth System Sciences*, *25*(5), 1655–1679. https://doi.org/10.5194/nhess-25-1655-2025

Ragu Ramalingam, N., Storrøsten, E. B., Gibbons, S. J., Davies, G., Lorito, S., Volpe, M., et al. (2024). Towards using machine learning emulation for probabilistic inundation mapping: Multiple earthquake sources and near-field effects. *Zenodo*. https://doi.org/10.5281/zenodo.14285109

Rim, D., Baraldi, R., Liu, C. M., LeVeque, R. J., & Terada, K. (2022). Tsunami early warning from global navigation satellite System data using convolutional neural networks. *Geophysical Research Letters*, *49*(20), e2022GL099511. https://doi.org/10.1029/2022GL099511

Rim, D., Suri, S., Hong, S., Lee, K., & LeVeque, R. J. (2024). A stability analysis of neural networks and its application to tsunami early warning. *Journal of Geophysical Research: Machine Learning and Computation*, *1*(4), e2024JH000223. https://doi.org/10.1029/2024JH000223

Salmanidou, D. M., Beck, J., Pazak, P., & Guillas, S. (2021). Probabilistic, high-resolution tsunami predictions in northern Cascadia by exploiting sequential design for efficient emulation. *Natural Hazards and Earth System Sciences*, *21*(12), 3789–3807. https://doi.org/10.5194/nhess-21-3789-2021

Scala, A., Lorito, S., Romano, F., Murphy, S., Selva, J., Basili, R., et al. (2020). Effect of shallow slip amplification uncertainty on probabilistic Tsunami hazard analysis in subduction zones: Use of long-term balanced stochastic slip models. *Pure and Applied Geophysics*, *177*(3), 1497–1520. https://doi.org/10.1007/s00024-019-02260-x

Scicchitano, G., Monaco, C., & Tortorici, L. (2007). Large boulder deposits by tsunami waves along the Ionian coast of south-eastern Sicily (Italy). *Marine Geology*, *238*(1), 75–91. https://doi.org/10.1016/j.margeo.2006.12.005

Selva, J., Lorito, S., Volpe, M., Romano, F., Tonini, R., Perfetti, P., et al. (2021). Probabilistic tsunami forecasting for early warning. *Nature Communications*, *12*(1), 5677. https://doi.org/10.1038/s41467-021-25815-w

Selva, J., Tonini, R., Molinari, I., Tiberti, M., Romano, F., Grezio, A., et al. (2016). Quantification of source uncertainties in Seismic Probabilistic Tsunami Hazard Analysis (SPTHA). *Geophysical Journal International*, *205*(3), 1780–1803. https://doi.org/10.1093/gji/ggw107

Sepúlveda, I., Tozer, B., Haase, J. S., Liu, P. L.-F., & Grigoriu, M. (2020). Modeling uncertainties of bathymetry predicted with satellite altimetry data and application to Tsunami hazard assessments. *Journal of Geophysical Research: Solid Earth*, *125*(9), e2020JB019735. https://doi.org/10.1029/2020JB019735

Smith, S. L., Kindermans, P.-J., Ying, C., & Le, Q. V. (2018). Don't decay the learning rate, increase the batch size. *arXiv*. https://doi.org/10.48550/arXiv.1711.00489

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, *15*(56), 1929–1958.

Tibshirani, R. (1996). A comparison of some error estimates for neural network models. *Neural Computation*, *8*(1), 152–163. https://doi.org/10.1162/neco.1996.8.1.152

Tonini, R., Basili, R., Maesano, F. E., Tiberti, M. M., Lorito, S., Romano, F., et al. (2020). Importance of earthquake rupture geometry on tsunami modelling: The Calabrian Arc subduction interface (Italy) case study. *Geophysical Journal International*, *223*(3), 1805–1819. https://doi.org/10.1093/gji/ggaa409

Tonini, R., Di Manna, P., Lorito, S., Selva, J., Volpe, M., Romano, F., et al. (2021). Testing tsunami inundation maps for evacuation planning in Italy. *Frontiers in Earth Science*, *9*, 628061. https://doi.org/10.3389/feart.2021.628061

Tozato, K., Takase, S., Moriguchi, S., Terada, K., Otake, Y., Fukutani, Y., et al. (2022). Rapid tsunami force prediction by mode-decomposition-based surrogate modeling. *Natural Hazards and Earth System Sciences*, *22*(4), 1267–1285. https://doi.org/10.5194/nhess-22-1267-2022

Volpe, M., Lorito, S., Selva, J., Tonini, R., Romano, F., & Brizuela, B. (2019). From regional to local SPTHA: Efficient computation of probabilistic tsunami inundation maps addressing near-field sources. *Natural Hazards and Earth System Sciences*, *19*(3), 455–469. https://doi.org/10.5194/nhess-19-455-2019

Williamson, A. L., Rim, D., Adams, L. M., LeVeque, R. J., Melgar, D., & González, F. I. (2020). A source clustering approach for efficient inundation modeling and regional Scale probabilistic tsunami hazard assessment. *Frontiers in Earth Science*, *8*, 591663. https://doi.org/10.3389/feart.2020.591663