Natural Hazards
and Earth System
Sciences

Advancing nearshore and onshore tsunami hazard approximation with machine learning surrogates

Naveen Ragu Ramalingam[1], **Kendra Johnson**[2], **Marco Pagani**[2,3], **and Mario L. V. Martina**[1]

[1]University School for Advanced Studies – IUSS Pavia, Pavia, 27100, Italy
[2]Global Earthquake Model (GEM) Foundation, Pavia, 27100, Italy
[3]Institute of Catastrophe Risk Management, Nanyang Technological University, 639798, Singapore

**Correspondence:** Naveen Ragu Ramalingam (naveen.raguramalingam@iusspavia.it)

Abstract. Probabilistic tsunami hazard assessment and probabilistic tsunami risk assessment (PTHA and PTRA) are vital methodologies for computing tsunami risk and prompt measures to mitigate impacts. However, their application across extensive coastlines, spanning hundreds to thousands of kilometres, is limited by the computational costs of numerically intensive simulations. These simulations often require advanced computational resources, like high-performance computing (HPC), and may yet necessitate reductions in resolution, fewer modelled scenarios, or use of simpler approximation schemes. To address these challenges, it is crucial to develop concepts and algorithms for reducing the number of events simulated and more efficiently approximate the needed simulation results. The case study presented herein, for a coastal region of Tohoku, Japan, utilises a limited number of tsunami simulations from submarine earthquakes along the subduction interface to build a wave propagation and inundation database. These simulation results are fit using a machine learning (ML)-based variational encoder–decoder model. The ML model serves as a surrogate, predicting the tsunami waveform on the coast and the maximum inundation depths onshore at the different test sites. The performance of the surrogate models was assessed using a 5-fold cross-validation assessment across the simulation events. Further, to understand their real-world performance and generalisability, we benchmarked the ML surrogates against five distinct tsunami source models from the literature for historic events. Our results found the ML surrogate to be capable of approximating tsunami hazards on the coast and overland, using limited inputs at deep offshore locations and showcasing their potential in efficient PTHA and PTRA.

1 Introduction

Tsunamis are potentially one of the most devastating natural hazards impacting life, property, and the environment. More than 250 000 casualties and USD 280 billion in damages were caused by tsunamis worldwide between 1998 and 2017 (Imamura et al., 2019), with the 2004 Indian Ocean and 2011 Tohoku tsunami events responsible for most of these losses. Tsunami hazard and risk assessments are fundamental in disaster management, as they facilitate the effective management of coastal regions and communities at risk of experiencing a tsunami (Mori et al., 2022). The results of tsunami hazard and risk assessment help plan and prioritise local and regional hazard mitigation efforts like land use and management, engineering protective structures and buildings, tsunami monitoring and early-warning systems, evacuation plans, and emergency response.

The simulation-based tsunami hazard analysis workflow consists of modelling different processes of the tsunami life cycle – generation, propagation, and inundation (Behrens and Dias, 2015; Marras and Mandli, 2021) – as depicted in Fig. 1. Each of these tsunami processes requires forward numerical modelling at different spatial scales, with varying complexity and different numerical schemes. Many of these steps are computationally demanding, and a substantial number of such simulations may be required for tsunami hazard analysis.
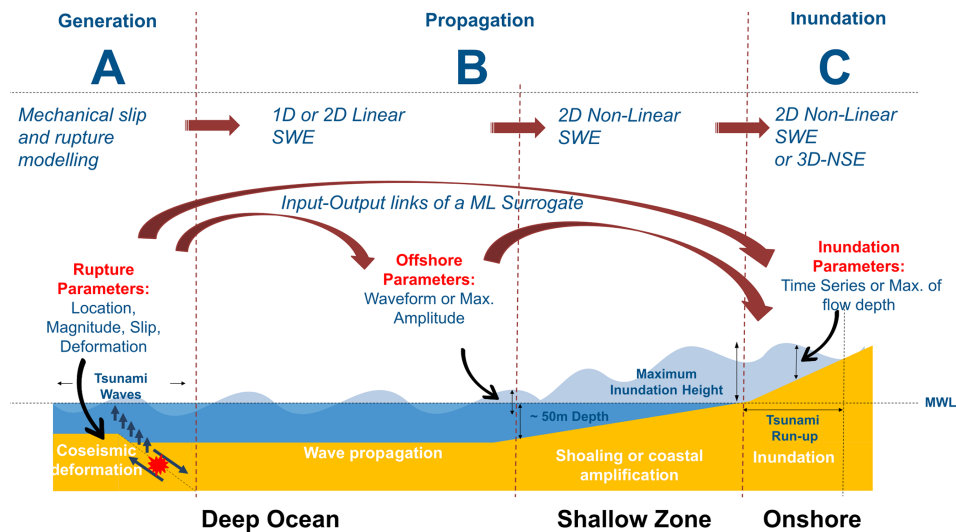
Figure 1. The sequence of a tsunami with deep ocean generation and propagation, shallow coastal-zone shoaling, coastal inundation and interactions with the built environment, and the methods used to model each part of the sequence. The dashed red lines depict the interfaces between the tsunami forward-modelling steps where different models can be coupled or linked.

There are two broadly categorised approaches to tsunami hazard and risk assessment (Mori et al., 2018): deterministic and probabilistic approaches. The deterministic approach aims to study a limited number of large tsunami scenarios, such as historical events or the possible worst-case events. This approach has been most widely used as it requires fewer events to simulate and hence less computational effort. With only a single or limited number of tsunami scenarios modelled, we can estimate a tsunami hazard metric for the given scenarios, such as wave height at an offshore site or inundation depth, run-up, etc. at an onshore location of interest for each scenario. Such results are easy to communicate and are useful in conservative decision-making activities such as evacuation planning. Instead, the probabilistic approach involves modelling a large number of possible tsunami events typically in the range of thousands (Gibbons et al., 2020) to millions (Basili et al., 2021; Davies et al., 2018) to estimate the exceedance rate of the said tsunami hazard metric at a location or region of interest (Geist and Parsons, 2006; Grezio et al., 2017). This approach is complex and computationally demanding but allows for the possibility of exploring different sources of uncertainty and making risk-informed decisions. When linked with fragility and loss models, probabilistic estimates of potential damage or loss of life and property are obtainable (Goda and De Risi, 2017).

The large computational burden from the many simulations needed in probabilistic tsunami modelling can limit their application, especially for onshore tsunami hazard and risk assessment where modelling of inundation processes is vital (Lorito et al., 2015; Grezio et al., 2017). Accurately simulating the tsunami wave shoaling process nearshore and the resulting inundation onshore necessitates the use of a 2D non-linear shallow water equation (NLSWE) model at reso-

lutions finer than at least 100 m on land. For scenarios where modelling turbulent tsunami forces onshore with precision is essential, a 3D Navier–Stokes equation (NSE) model proves to be suitable (Marras and Mandli, 2021). Importantly, both the 2D NLSWE and 3D NSE models come with higher computational costs, exhibiting an exponential factor in comparison to the more computationally efficient 1D linear shallow water equation (LSWE) model commonly employed to model tsunami wave propagation in deeper oceanic regions. Figure 1 represents the offshore-to-onshore tsunami forward-modelling flow and the input–output links of the machine learning (ML) surrogates.

To handle or overcome this computational challenge of scale and the need to model a large number of events, various methods have been adopted (Behrens et al., 2021):

1. reduce the number of events to be modelled, using sampling techniques, clustering, and other selection methods (Lorito et al., 2015; Williamson et al., 2020; Davies et al., 2022);

2. approximate results using methods like Green's function, amplification factors, and reduced complexity models (Molinari et al., 2016; Løvholt et al., 2016; Glimsdal et al., 2019; Gailler et al., 2018; Grzan et al., 2021; Röbke et al., 2021);

3. improve hardware and computation like a nesting of grid domains, adaptive variable grid resolution, parallelisation and GPU-based acceleration, coupled multi-scale modelling, and exascale codes (LeVeque et al., 2011; Shi et al., 2012; Oishi et al., 2015; Macías et al., 2017; Marras and Mandli, 2021; Folch et al., 2023);

4. use surrogates using statistical emulators and ML models (Sarri et al., 2012; Salmanidou et al., 2021; Mulia et al., 2018; Fauzi and Mizutani, 2019; Makinoshima et al., 2021; Fukutani et al., 2023; Mulia et al., 2022).

Running the numerical simulations, especially for modelling the tsunami inundation with an NLSWE model, takes a lot of time and computational resources, e.g. accounting for about 13 600 GPU hours in the local probabilistic tsunami hazard assessment (PTHA) study for Catania by Gibbons et al. (2020) (refer to Table 6 for runtimes in this study). Among the different approaches to reducing such computational and time burdens, surrogates can provide an instantaneous approximation of the output of the numerical simulation. Surrogates are fit (trained) on a set of model inputs and outputs, called training data, to derive a mathematical or statistical relationship between the inputs and outputs; this provides a fast solution to otherwise time-consuming numerical simulations but may introduce some acceptable errors. Surrogates, such as statistical emulators, can be fitted using a small training dataset built using a limited number of simulations. The application of such surrogates in tsunami modelling has been used for uncertainty quantification or sensitivity analysis (Sarri et al., 2012; de Baar and Roberts, 2017; Kotani et al., 2020; Salmanidou et al., 2021; Giles et al., 2021; Tozato et al., 2022) and hazard assessment (Fukutani et al., 2021, 2023; Lee et al., 2023), which are generally difficult to conduct in a brute-force approach where one would need to simulate all possible events explicitly or in a real-time setting where running high-fidelity models in a limited time is difficult. Another type of surrogate used is machine learning models (referred to as ML surrogates herein), which are also trained on the available input–output datasets using a supervised learning framework. Tsunami ML surrogates in comparison to statistical surrogate models, especially those based on deep neural networks, utilise a large training dataset to optimise the model parameters using backpropagation. This may require a much higher number of numerical simulations to create the relevant input–output datasets for training or fitting such a model. Using such a framework for prediction of a real-time tsunami is feasible and widely proposed for faster-than-real-time forecasting and early-warning purposes (Mulia et al., 2018; Fauzi and Mizutani, 2019; Liu et al., 2021; Rodríguez et al., 2022; Kamiya et al., 2022; Mulia et al., 2022; Wang et al., 2023), especially with inputs derived directly from real-time sensors (Makinoshima et al., 2021; Mulia et al., 2022). See Table 1 for a comparison of different surrogate classes.

Models of varying complexity and resolution, as depicted in Fig. 1, are often coupled to accurately simulate the generation and propagation of tsunamis from the deep ocean to the nearshore and inundation onshore (Abrahams et al., 2023; Son et al., 2011). For probabilistic tsunami modelling with a large number of events, the typical approach is to use outputs of a given model level as boundary conditions to the next model in the forward-modelling flow, taking advantage of each model's varying complexity, resolution, and computational efficiency. In this study, a hybrid modelling framework is introduced (see Sect. 2). We suggest using a limited number of full simulations to build or train a tsunami ML surrogate and completing the most computationally demanding phases of tsunami forward modelling for the remaining events as direct final modelling using the ML surrogate. Figure 1 compares the different methods used in traditional tsunami forward-modelling chains and as alternative different connections for the setting up of a surrogate in the form of statistical emulators and ML models.

This framework recognises several key concepts, including reducing the number of events for numerical simulation, modular multi-scale modelling, and the use of ML-based surrogates for hazard approximation. By combining these ideas, the proposed framework provides a comprehensive approach to enhance the modelling process and efficiently achieve the final results. Many of the models and datasets needed in this framework are already available; we discuss how to put them together in a workflow and build the ML-based surrogate for the final tsunami modelling. Further, we conduct experiments to check the skill and usability of an ML surrogate for tsunami hazard approximation nearshore and onshore.

The current challenge lies in the need for an extensive size of the training dataset generally required by ML surrogates in tsunami modelling for training to generate accurate predictions. While they provide instantaneous results, the cost of developing the training dataset for the ML surrogates may outweigh the benefit of instantaneous prediction. When ML models are not appropriately trained to understand the underlying physics or dynamics of the tsunami and use a small training dataset, they may overfit and struggle to generalise well beyond the training data (out-of-training situations) (Seo, 2024).

To overcome this dependence on a large training size, our ML surrogate exploits the ability of the encoder–decoder network in dimensional reduction, feature representation, and sequence-to-sequence transformation for approximating tsunami hazards nearshore and onshore. We use tsunami data from a small set (about 500) of simple earthquake rupture scenarios which may provide a sufficient learning basis for the ML surrogate. We describe how a variational encoder–decoder (VED), a type of neural network model (see Sect. 2.2.2), is trained to take the tsunami waveform at points where offshore depth is 100 m as input and predict the tsunami waveform at nearshore points with depths of 5 m and maximum inundation maps onshore for three different locations along the coastal Tohoku region in Japan. Thus, skipping the computationally demanding modelling of non-linear processes in the shallow water regions nearshore and the inundation processes onshore for a given tsunami event. We check the generalisation ability of this model by testing the prediction error for a set of finite-fault rupture events of his-

Table 1. Comparison of recent work using surrogates for tsunami approximation.

Reference	Method	Simulation size (no. of events)	Input parameter	Output parameter
Statistical surrogate				
Salmanidou et al. (2021)	GPE	60	Seafloor displacement parameters	Max height at coast
Tozato et al. (2022)	GPE, SVD	50	EQ source parameters (slip and rake)	Onshore max inundation depths, impact force
Fukutani et al. (2023)	GPE, SVD	360	EQ magnitude (M_w), Seawall height	Onshore max inundation depths
Gopinathan et al. (2021)	GPE	300	EQ magnitude and location	Max height, velocity at coast
ML surrogate				
Liu et al. (2021)	CNN (VAE)	1300	Short tsunami observation	Long tsunami forecast nearshore
Makinoshima et al. (2021)	CNN	10 000	Short tsunami observation	Onshore point inundation time series
Mulia et al. (2022)	Dense NN	3060	Short tsunami observation (max)	Onshore max inundation depths
Rodríguez et al. (2022)	Dense NN	16 000	EQ source parameter	Max wave height, arrival time at coast
Núñez et al. (2022)	CNN	6776	Offshore observation at 50 m or 100 m depth	Onshore point inundation time series
Cesario et al. (2023)	Regression tree	15 408	EQ source parameter	Max wave height at coast
de la Asunción (2024)	Dense NN	128 000	EQ source parameter	Alert level at coast

GPE: Gaussian process emulator, SVD: singular value decomposition, CNN: convolutional neural network, VAE: variational autoencoder, Dense NN: dense neural network. EQ: earthquake.

toric tsunami scenarios to evaluate the efficacy of this hybrid modelling approach.

2   A framework for approximating tsunami hazard nearshore and onshore at reduced computation cost

Our framework employs a hybrid modelling approach that integrates physics-based numerical simulation and data-driven ML models for tsunami hazard approximation. By combining the strengths of these models, we aim to represent the tsunami hazard for a coastal site of interest as the time series of the tsunami wave height near or along the shore and the max inundation depth onshore within a reasonable computational budget.

This section discusses the various components of the proposed framework as seen in Fig. 2. The first step is the generation of synthetic data which is needed to train the ML surrogate model. This requires the full forward modelling of the tsunami generated from simple earthquake sources discussed in the design of events and recording the tsunami water-level time series for such events at different depths: at 100 m depth

and 5 m depth nearshore, as well as the maximum inundation depth onshore for the event using an NLSWE tsunami model.

The second step is the development of ML models for use as surrogates, whose outputs are the nearshore tsunami height time series and the onshore maximum inundation depth. In the current study, the framework is applied and tested by comparing the prediction of the ML surrogate against results from numerical simulation for a portion of the synthetic events as hold-out testing and historic events for generalisation testing.

2.1   Generation of synthetic data

The characteristics of tsunamis in a region are intricately tied to factors such as earthquake sources and their recurrence, ocean bathymetry, shelf profile, coastal topography, and artificial structures like coastal defences and urban infrastructure. The rarity of tsunami events and the limited availability of comprehensive observational records across the coastal region make it impossible to construct a robust dataset for
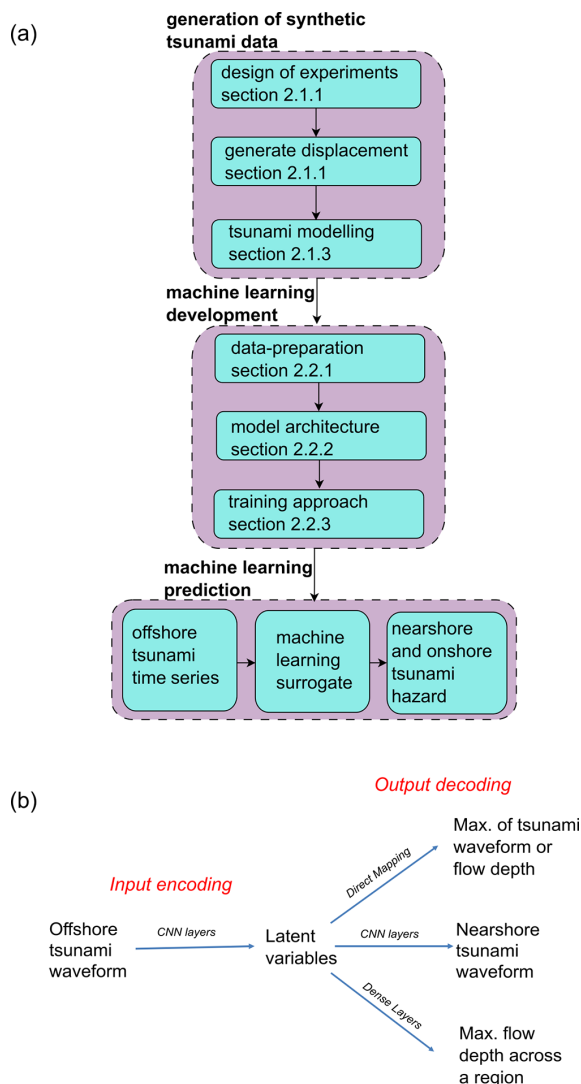
Figure 2. Overall framework for nearshore and onshore hazard approximation using an ML surrogate. **(a)** Flowchart of the procedure. **(b)** Different model structures based on prediction types.

training and testing an ML model using purely historic observations.

Due to this data limitation, we instead attempt to create a diverse dataset for training and testing the ML surrogate to effectively capture the wide spectrum of tsunami dynamics in a region. The idea is to simulate earthquakes of different magnitudes, locations, slip amounts, and rupture geometries, generating tsunami waves of different amplitudes and wavelengths offshore and causing inundation of varying patterns and extents onshore.

Thus, we model a wide variety of earthquake scenarios discussed in Sect. 2.1.2 using a tsunami hydrodynamic model created for the Tohoku region (Sect. 2.1.1), covering three test sites.

2.1.1 Tsunami model and test locations

Three locations with different coastal configurations are identified for this study – Rikuzentakata (enclosed bay), Ishinomaki (shielded), and Sendai (open bay) – where different coastal processes like shoaling, refraction, reflection, and resonance are expected to impact the tsunami nearshore results and provide varied settings for testing the proposed methodology of developing an ML-based surrogate.

For the simulation of the synthetic events, as well as recording their offshore waveform (water-level time series) and onshore flood inundation depths, a tsunami model based on GeoClaw Version 5.7.1 (Clawpack Development Team, 2020) was developed. This GeoClaw model covers the Pacific side of Honshu island and solves the depth-averaged NLSWE using adaptive mesh refinement using rectangular grids in the geographic coordinate system (latitude and longitude) for the base level 0 at a resolution of 0.01215°.

The adaptive mesh refinement ranges from level 1 of resolution 0.006075° for the overall model domain, level 2 of resolution 0.00405° at bathymetric depths around 100 to 850 m, and level 3 of resolution 0.00135° for bathymetric depths less than 100 m as in Fig. 3a. An additional refinement is enforced for the three test sites, resulting in the finest grids of 0.00045° that are roughly equivalent to 50 m in resolution for capturing the inundation onshore as seen in Fig. 3b.

The model uses topographic data from the Japan Cabinet Office project, available in 1350–450–150–50 m resolutions (Japan Cabinet Office, 2016); the final 50 m resolution grids of this dataset were superimposed with COP-DEM (European Space Agency, 2024), a digital surface model (DSM) representing the elevation of the Earth's surface including buildings, infrastructure, and vegetation with a native resolution of 30 m, and a coastal tsunami defence elevation dataset also available from the cabinet project for representing the onshore elevation more realistically for the tsunami inundation modelling.

The virtual gauges are set at depths of 5 and 100 m as shown in Fig. 3b to record the elevation of water level at regular time intervals, and fixed grid monitoring is used to record the maxima of the inundation depths from each event for the three coastal regions for each tsunami simulation. Each tsunami simulation is run for a 6 h duration from the onset of the tsunami. Tides and wave components are ignored, and the initial water-level condition is set at zero mean sea level to consider a still ocean condition.

The model was tested and calibrated using the 2011 Tohoku historical event's gauge (NOWPHAS, 2011) and inundation survey (CEC, 2012) data; see the results available in Figs. S1 and S2 in the Supplement.

2.1.2 Design of experiment (DOE)

The design of the experiment consists of a total of 559 hypothetical events of two categories of rupture, distinguished
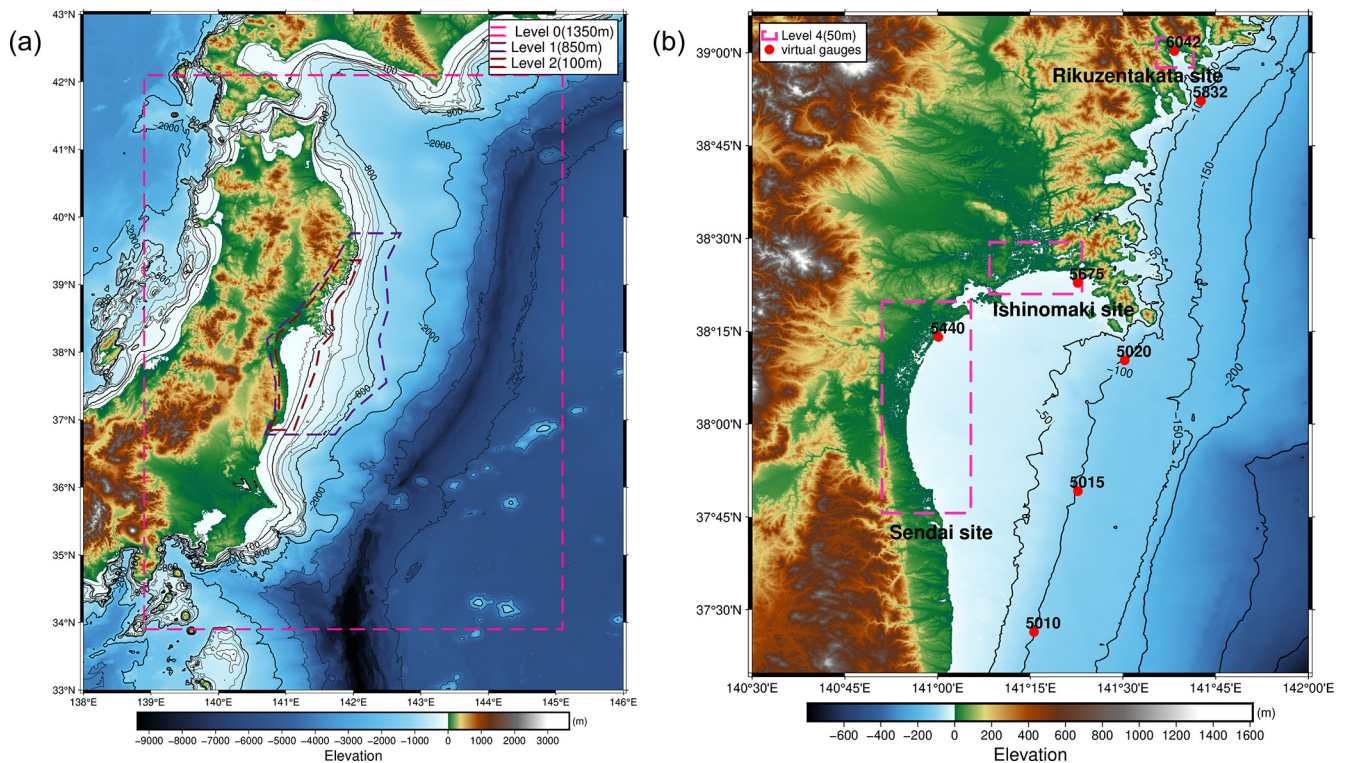
Figure 3. (a) Tsunami model coverage and the adaptive grid resolution system. **(b)** Virtual gauges and inundation sites with high-resolution nested domains.

by their geometry and slip distribution: (a) Type A – ruptures represented by a single rectangular planar surface with homogeneous slip – and (b) Type B – ruptures that combine numerous smaller rectangular planar surfaces (i.e. sub-ruptures), each of which has homogeneous slip, such that the rupture surface can bend and that the slip distribution can be heterogeneous. The number of events in the DOE is constrained by available computational resources, and our goal is to use a feasible number of training events and maximise the efficiency of the surrogate model. The source scenarios are modelled by adapting procedures previously applied by Gusman et al. (2014) and Mulia et al. (2018).

A total of 119 locations were selected as the top centre of the faults for modelling hypothetical tsunamigenic earthquakes of Type A. These events span M_w 7.5–9.0 at an interval of 0.5 and are uniformly distributed over the Tohoku subduction interface (see Fig. 4a and b). This results in 476 potential events (119 locations × 4 magnitudes). The M_w 9.0 events are restricted to locations where the centre of the rupture's top edge is shallower than 16 km. Deeper events cause unrealistic uplift on large inland portions of the study region and are unlikely to cause a tsunami. To ensure realistic modelling and prohibit these events from adversely affecting the quality of the surrogate training, these events were excluded, leaving 383 Type A events.

Multi-fault ruptures of Type B were created using a combination of 6 to 12 planar sub-faults similar to the unit sources used in the NOAA SIFT database (Gica et al., 2008) of length 100 km and width 50 km, as in Fig. 4c. The event magnitudes range M_w 8.68–9.08, and the ruptures are distributed along the shallow section of the Tohoku subduction interface. The bottom centres of the rupture edges are at depths between 17–28 km. The slip distributions are modelled as a skewed normal distribution where the average combined slip value is between 10 and 20 m. The scenarios varied by the number of faults involved: scenarios with 6 faults were assigned with a slip of 10 m, scenarios with 8 faults with slips of 10 and 15 m, 10 faults with a slip of 15 m, and 12 faults with slips of 15 and 20 m. This systematic variation led to a total of 176 different Type B earthquake scenarios.

Information on depth, slip, strike, and dip (see Table 2) is derived from the Slab2 model of the Japan Trench (Hayes et al., 2018), and the rake is always set at 90° (Aki and Richards, 2002). The seafloor deformation is analytically modelled by assuming homogeneous slip for the rupture or sub-ruptures using the Okada solution (Okada, 1985), with the values of rupture length, width, and slip scaled (see Table 2) based on the magnitude of the event (Strasser et al., 2010). We consider that the co-seismic displacement is instantaneous and equivalent to the sea surface displacement generating the tsunami. This initial sea surface displacement
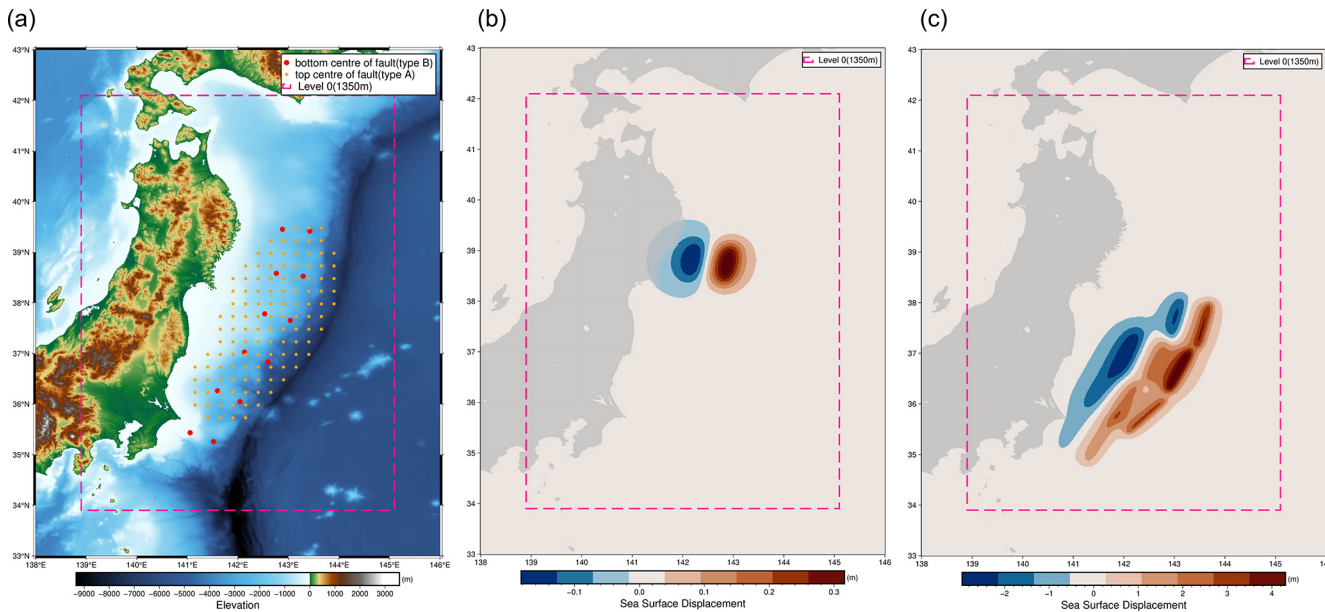
Figure 4. Example initial displacement for the two rupture types used in the design of experiment for the tsunami dataset. **(a)** Fault location for Type A and Type B; **(b)** Type A – event 35; **(c)** Type B – event 33.

Table 2. Earthquake rupture parameters for Type A and Type B; the rake value is always 90°.

Type	M_{w}	Length (km)	Width (km)	Displ. (m)	Depth (km)	Dip (°)	Strike (°)
Type A Min	7.5	81.37	56.29	0.24	10.2	5.54	187.20
Type A Max	9	613.76	189.23	3.30	45.7	17	225.78
Type B Min	8.68	300	100	4.72	17.01	8.37	188.72
Type B Max	9.08	600	100	17.36	28.98	16.53	222.27

is modelled to match the base resolution of the tsunami model at 0.01215° grids.

In summary, the DOE for training the surrogate model considered two main factors: (a) moment magnitude, which determines the profile of displacement (length, width, and slip amount) based on the moment magnitude-area scaling relationship (Strasser et al., 2010), and (b) the location of the events where fault parameters such as depth, dip, rake, and strike are derived from the Slab2 dataset.

2.1.3 Test events

Along with using a random subset of the events from the design of experiments, we additionally model a set of five events to evaluate the performance of the ML model (shown in Fig. 5). This is to test the model over a generalised dataset which is different from the training dataset used to build the model. The numerically simulated tsunamis are known to be sensitive to the earthquake slip, fault geometry, rupture mechanisms, and the discretisation used (Gibbons et al., 2022; Goda et al., 2014), and their resulting wave profile,

direction, and inundation may vary significantly. Although there is significant variation in the characteristics of the different tsunami events generated in the design of experiments to train and test the ML surrogate, we add these additional historic test events with complex heterogeneous slip, events with instantaneous and time-dependent slip displacement, and events located beyond the Tohoku source region used for training, as seen in Fig. 4. These include the following events described below.

1. 2011 Tohoku Earthquake:
   – Test A (instantaneous displacement) using Fujii et al. (2011),
   – Test E (time-dependent displacement) using Yamazaki et al. (2018);

2. 1933 Sanriku Earthquake:
   – Test B (outer rise event with normal faulting) using Okal et al. (2016);

3. 1896 Sanriku Earthquake:

- Test C (outside Tohoku source region considered in DOE) using Satake et al. (2017);

4. 1968 Tokachi-oki Earthquake:

- Test D (outside Tohoku source region considered in DOE) using Riko et al. (2001).

2.2 Machine learning model

Previous studies (Fauzi and Mizutani, 2019; Liu et al., 2021; Makinoshima et al., 2021; Núñez et al., 2022; Mulia et al., 2022; Rim et al., 2022) that focused on using ML models for use in tsunami forecasting and early-warning needs have also used neural networks in the form of convolution and dense NN models. These models are usually trained to take short-duration (20–30 min) inputs from sensors like offshore ocean bottom pressure sensors and geodetic sensors (GNSS, global navigation satellite system) for their tsunami hazard prediction. This is due to the constraints of a short lead time in the event of a tsunami. Furthermore, the tsunami observation network is designed with a given earthquake source region in focus, leading to the design and training of the ML model which is constrained to a very specific regional and source setting.

In contrast, we discuss two ML surrogate models in this work with the main difference being the output of the model: either a time series of the water surface elevation or the maximum inundation depth across a set of onshore locations. The design of the ML model or surrogate for the current framework focuses on the following:

1. The model architecture should be able to train and fit with a relatively small number of events such that it can serve the purpose of solving the computational constraint for its use as a surrogate model.

2. The model has a balanced performance; it can predict the hazard sufficiently well for small-magnitude and large-magnitude events across the domain of interest.

3. The model training is not sensitive to training datasets; that is, it should not overfit the data and should be capable of predicting results for different ruptures as long as the offshore water-level amplitude is available.

4. The model design should be able to connect easily with available outputs from other regional propagation models and be easily replicated across different coastal configurations.

2.2.1 Data preparation

The ML model in this study uses two types of datasets from the tsunami simulations: (a) tsunami waveforms – water-level time series – and (b) maximum inundation depth map. Figure 6 provides a summary of the dataset for the three test

locations – inundation depths, areas, and maximum amplitudes at the virtual gauges.

For each of the events, we process the simulated water-level time series at the virtual gauges and the maximum inundation map in the following steps. All events where the tsunami water level does not cross a set threshold of 0.1 m at the selected deep offshore virtual gauges of 100 m depth and 0.5 m at the shallow nearshore virtual gauge of 5 m depth are ignored as they result in negligible tsunami inundation.

At the instance of a time when this threshold is crossed at the shallow gauge, a simulation time window of 240 min is selected to calculate a uniformly sampled wave amplitude time series with 1024 data points. Many of these local source events cause significant local deformation to the bathymetry, which is captured in the water-level time series. This offset is removed from the time series data at both nearshore and offshore virtual gauges (e.g. Fig. 7) as a preprocessing step for reducing the complexity in the dataset which can help with training the model.

In the case of the maximum inundation maps, all grid points that are flooded across the entire simulation dataset of all the modelled events are selected as a fixed set of prediction locations shown in Fig. 8 for Rikuzentakata, and the maximum inundation for each event at each such location is stored as a 1D array. This yields 6648 grid points for the Rikuzentakata area, 54671 grid points for the Ishinomaki area, and 129 941 grid points for the Sendai area.

2.2.2 Variational encoder–decoder (VED) – model architecture and training approach

Our design of the VED architecture extends from existing statistical and ML models used for tsunami forecasting and rapid modelling. Mulia et al. (2018) used the feature space from the principal component analysis (PCA) method as a form of dimensional reduction to search the closest inundation map from a large simulation database against results from a quick low-resolution simulation and interpolate new high-resolution results using these samples. ML-based models that are trained on tsunami simulation databases, in particular convolutional neural networks (CNNs), have also been used to predict the full time series of the water-level elevation using sparse observational inputs (Liu et al., 2021; Makinoshima et al., 2021; Núñez et al., 2022). In the case of prediction for maximum inundation depth maps, fully connected or dense layers have been used to efficiently map the output across the large set of locations as in Fauzi and Mizutani (2019) and Mulia et al. (2020, 2022).

The deep neural network models used in the above work for tsunami predictions typically need a large number of training examples to learn a useful representation and avoid overfitting the parameters of their different layers. Certain model architectures and training schemes can help them become more learning efficient, generalise better, and provide higher prediction performance. One such approach is
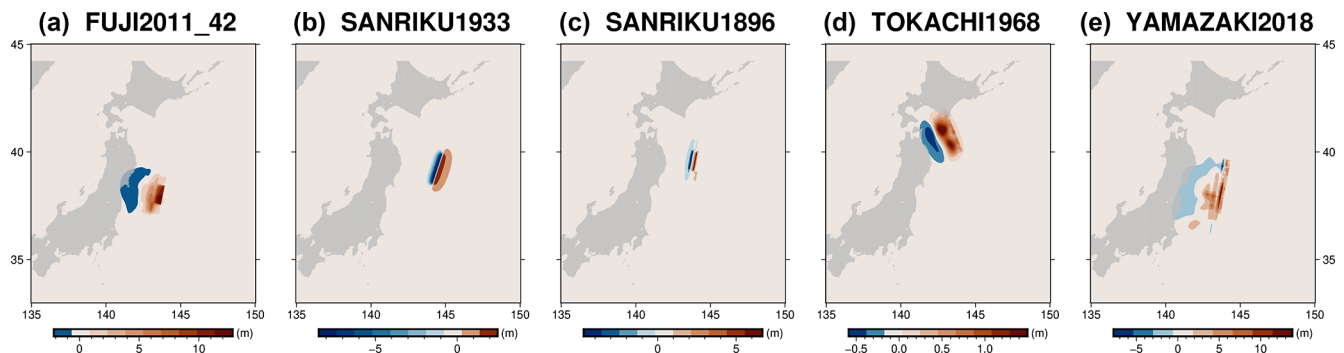
Figure 5. Initial sea surface displacement for the five test events.
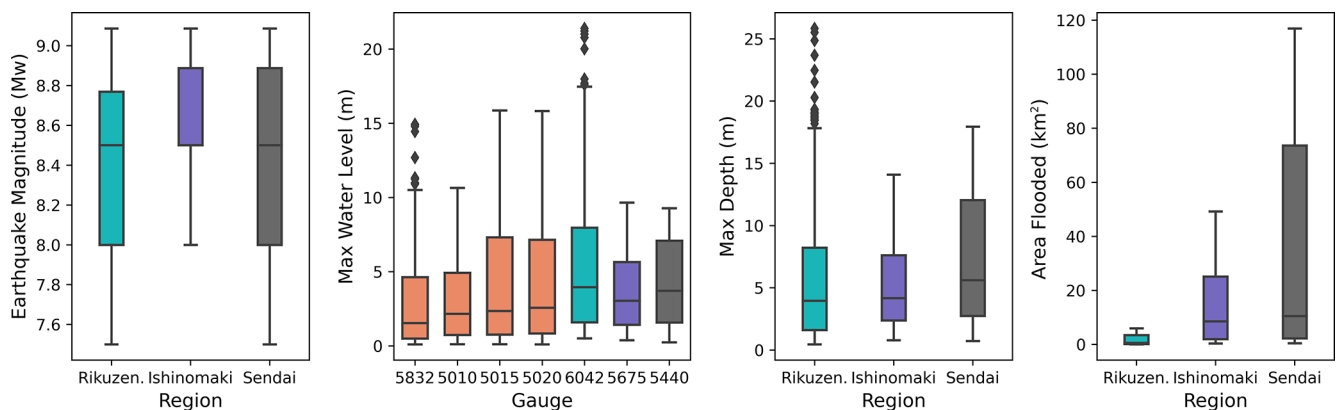
Figure 6. Box plot showing variability in the offshore and nearshore gauges and the inundation properties for the three test sites.

encoding–decoding, a class of supervised ML aimed at training a neural network to learn a lower-dimensional representation of the input that can be used to construct the output. It consists of three parts: encoder, latent variables, and decoder.

$$z = f(w_f, b_f, x), \tag{1}$$
$$y = g(w_g, b_g, z). \tag{2}$$

The encoder function $f(.)$ with its learnable parameters, weights w_f, and biases b_f maps the input data x to a reduced number of latent variables z, and a decoder function $g(.)$ with its learnable parameters, weights w_g, and biases b_g maps the latent variables z back to the high-dimensional data space as the construction of the output y (see Eqs. 1 and 2). The encoder–decoder (ED) is trained to minimise the reconstruction error between the input data and the output of the decoder network using backpropagation optimisation. Similar to Liu et al. (2021), we use convolutional layers and variational encoding (see Fig. 9). The encoding convolutional layers and max pool operations of the encoders perform dimensionality reduction on the time series data at the input gauges similar to a principle component analysis (PCA) of Mulia et al. (2018), and in the inverse process, the decoding transposed convolutional or dense layers perform the necessary transformation to predict the output time series or multi-location inundation depths.

Variational encoding maps the input data to the probabilistic distribution of the latent variables z; this makes VEDs more flexible in capturing the underlying data distribution and in modelling the epistemic uncertainty in the input encoding process. Compared to a deterministic encoder, the latent space is mapped as continuous and smooth variables, making for a more interpretable and structured representation of the data in the encoding latent variables (Liu et al., 2021).

We use two VED models as the ML-based surrogates to predict the water-level time series (waveform) at a gauge location and the flood inundation footprint in the form of the maximum inundation depth. Early in the study, we evaluated the prediction of two inundation parameters – maximum inundation height and maximum inundation depth – with the model predicting the maximum inundation depth better. We do not discuss the result for the maximum inundation height in this study. The models receive their input x in the form of a stack of time series from the selected offshore gauge(s), which is encoded into the latent variables z by four convolutional layers with kernel size of 3 and padding of 1 using a leaky ReLU (0.5, rectified linear unit) activation, which helps
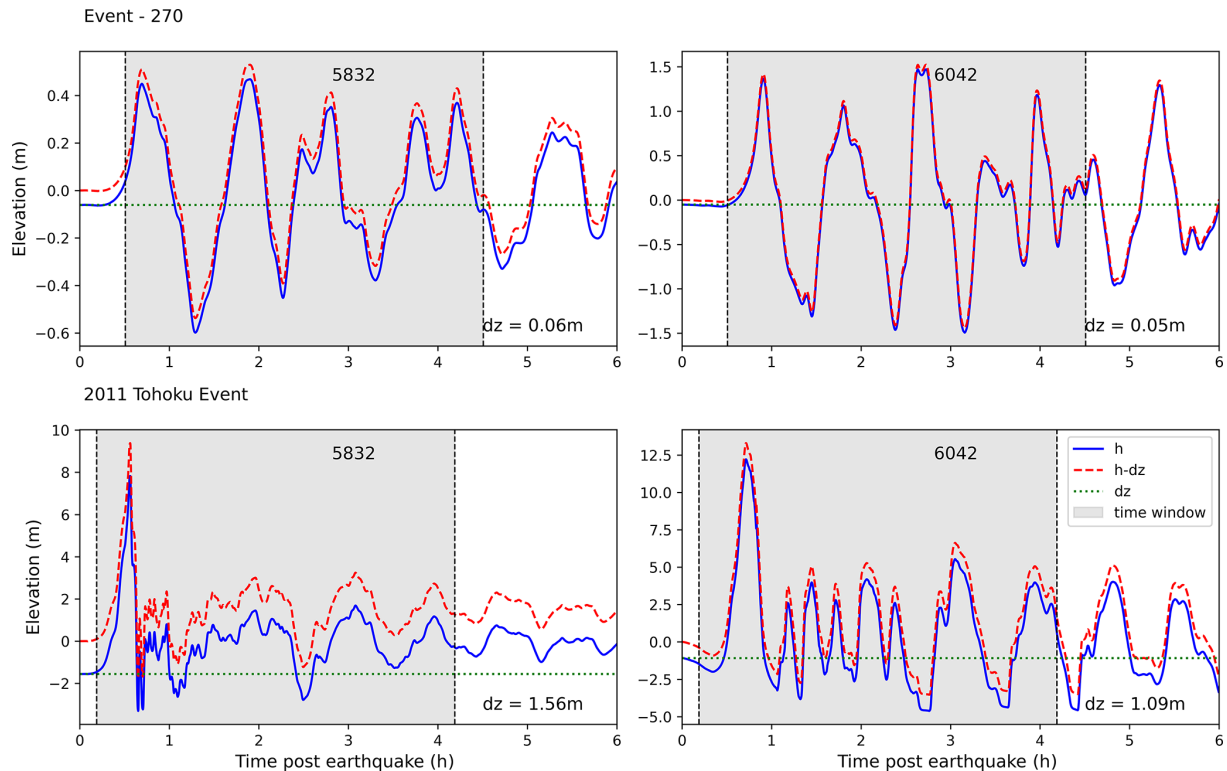
Figure 7. Correction of the waveform by removing deformation, with a 4 h selection window based on an offshore gauge crossing threshold of 10 cm.
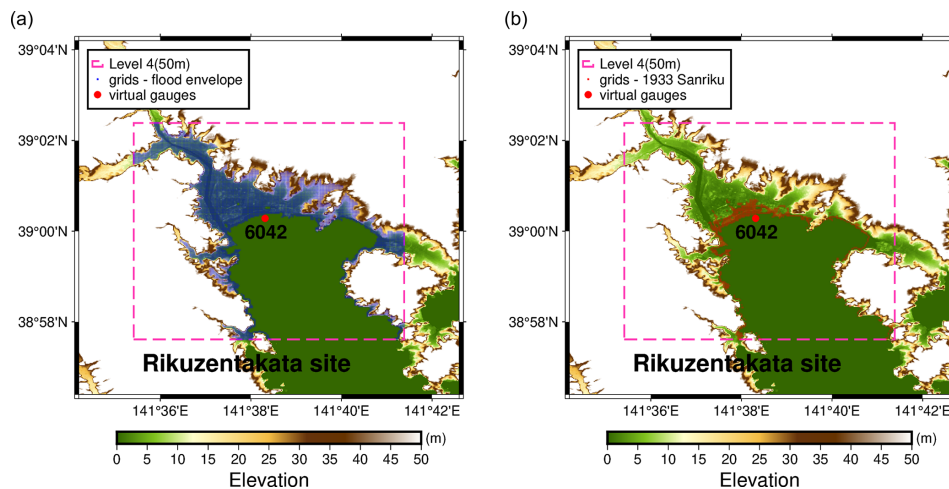
Figure 8. Location of grids affected by inundation: **(a)** within the overall flood envelope from all DOE simulations and **(b)** for a single 1933 Sanriku test event.

introduce non-linearity and prevent a vanishing gradient during training. After each convolutional layer, we perform a max-pooling operation to reduce spatial dimensions on the feature maps. The choices of the number of layers, number of channels, kernel size, and the input data sampling rate are interdependent and were decided by testing multiple configurations; Tables S1 and S2 describe the two surrogate archi-

tectures used in this study. For the encoding in the onshore VED model, we additionally apply batch normalisation after the first max-pooling layer to help stabilise the training as we use a single-batch training, and a dropout layer with a rate of 0.1 is applied after the final max-pooling layer as additional regularisation (refer to other operations in Table S2). The latent variables z are encoded as the mean and σ values
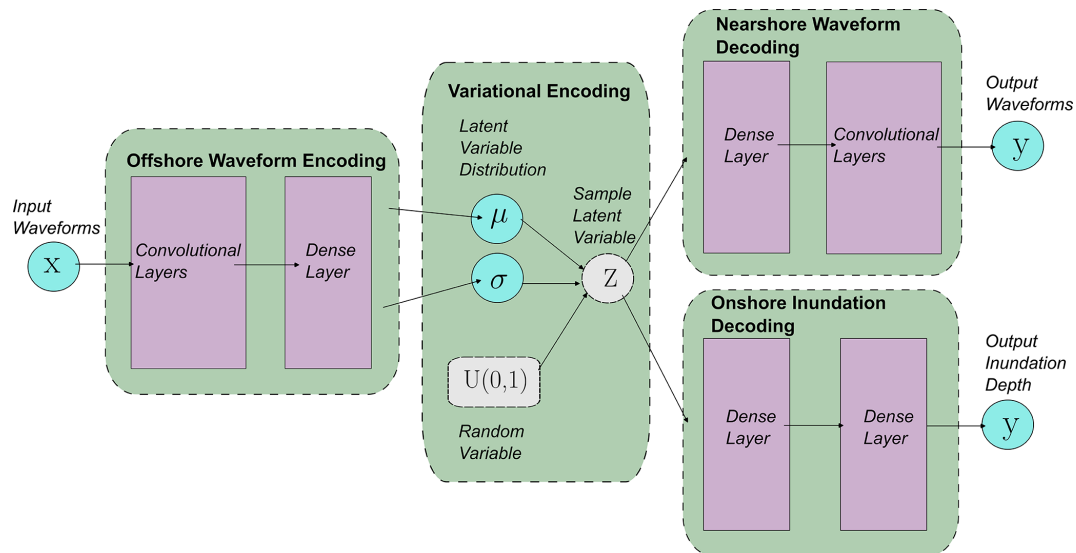
Figure 9. Schematic of a variational encoder–decoder.

of a Gaussian distribution using a reparameterisation method (Kingma and Welling, 2014).

The decoding layers take the sample from the latent space and reconstruct the outputs. For nearshore VED, to predict the nearshore time series of the water level, we reorder the encoder architecture in reverse using four transposed convolutional layers. For the onshore VED, to predict the max inundation depths, we use a dense layer that maps the sample from the latent space to the outputs, i.e. the max inundation depth directly. The model hyperparameters, including the number of latent variables, a learning rate of 0.0005, and no weight decay, were chosen through rough experimentation to achieve optimal results. Table 3 summarises the size of input, output, and latent variables used in the two surrogate models. Tables S1 and S2 also provide more information on the model layers and operations used in the two surrogate models.

The loss function used to train the models consists of two components: the prediction loss and the Kullback–Leibler (KL) divergence loss. The prediction loss measures the difference between the constructed data and the output data in terms of the mean squared error. The KL divergence loss (Kingma and Welling, 2014; Liu et al., 2021) encourages the latent space distribution to be close to a unit Gaussian distribution, acting as a regularisation term. This additional regularisation also helps prevent overfitting and improve generalisation. The components of the loss function in Eq. (3) can also be weighted to improve the overall model prediction.

$$\text{Total loss} = \underbrace{\frac{1}{2} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}_{\text{MSE (prediction loss)}} + \underbrace{\text{KL}(q(z|x)||p(z))}_{\text{KL divergence loss}} \qquad (3)$$

Given the relatively small size of our training dataset, we employ k-fold cross-validation with five folds to fully use the available dataset for training and testing. The dataset is randomly partitioned into five equal-sized folds. Each fold is used once as the test set, while the remaining four folds are used for model training. This process is repeated five times, with each fold serving as the test set exactly once. The characteristics and parameter ranges for each test fold, as when all the test folds are combined together, mirror those of the overall training dataset as found in Table 2 and Fig. 6. This helps with assessing the model's generalisation ability and sensitivity to overfit with such limited training information (Mulia et al., 2020). We conducted the training in a single batch. This approach differs from conventional training settings, where mini-batch training is adopted to handle large training datasets. The single-batch approach allows the model parameters to be optimised using the entire fold's dataset in each training iteration, effectively leveraging all available information to update its parameters. This strategy of data splitting, cross-validation, and VED model architecture facilitates better convergence and helps prevent model overfitting during training or data leakage during evaluation.

Applying an ensemble approach to the prediction, we use the variational encoding as in Fig. 9 to generate 100 sample predictions for each test event of the fold. Further, when generating prediction for the historic event in Sect. 2.1.3, we use all five models trained as part of the 5-fold cross-validation to generate 5×100 predictions, thus using the overall dataset for model training and prediction for these test events.

Table 3. Hyperparameters for nearshore and onshore surrogates in different regions.

	Region	Number of latent variables	Input dimensions	Output dimensions	Training size
Nearshore surrogate	Rikuzentakata	25	1×1024	1×1024	490
	Ishinomaki	150	3×1024	1×1024	425
	Sendai	100	3×1024	1×1024	465
Onshore surrogate	Rikuzentakata	10	1×1024	6648	490
	Ishinomaki	50	3×1024	54 671	425
	Sendai	30	3×1024	129 941	465

3 Metrics for evaluation

To assess the general fit between the numerical simulation and ML-predicted values, the mean squared error (MSE), the coefficient of determination (R^2), and the goodness of fit (G) are used. MSE quantifies the average squared difference between simulated and predicted values, with lower values indicating better model performance. R^2 measures the proportion of variance in the dependent variable (simulations) that is predictable from the independent variable (predictions), providing information on the accuracy of the model based on the correlation or dependence between simulations and prediction, with high values close to 1 indicating a good result. G is used as a cost function to measure the disagreement between observed and predicted values, considering both the magnitude and direction of the deviations, with a lower value close to 0 being better. Lastly, we used the L2Norm to estimate the magnitude of the event, using the vector representing the time series at the test gauge or the maximum inundation depths for a region.

In the case of the nearshore surrogate, where accurately predicting the peaks and their timing is crucial, relative peak deviation (RPD) and maximum peak delay (MPD) are used. RPD measures the relative difference between simulated and predicted peak values, providing insight into peak accuracy. MPD evaluates the time delay between highest simulated and predicted peaks, crucial for assessing the timing accuracy of peak. For the onshore surrogates, where accurately predicting inundation extents and depth is the focus, the accuracy score quantifies the proportion of correctly predicted inundation locations relative to the total number of locations. This metric provides a clear indication of the model's predictive performance in accurately mapping inundation for depths above the threshold of 10 cm or a select depth class (say 0.1 to 1 m). These metrics collectively offer a comprehensive evaluation of the ML surrogate's prediction.

4 Results

4.1 General fit

As a first step, we evaluate the sensitivity of the learning or parameter optimisation to the training data as part of the 5-fold cross-validation study. The mean squared error values are summarised in Table 5 for each of the folds. For the nearshore surrogate (predicts the time series), it ranges between 0.042 and 0.197, and for the onshore surrogate (predicts the max inundation depth map), it ranges between 0.169 and 1.129 at the three test sites. For the nearshore surrogate, the relatively consistent MSE across folds suggests that the model is robust, with no significant overfitting and an overall good fit to the data. However, for the onshore surrogate, there is noticeable sensitivity to the training data, as evidenced by slightly higher MSE values for certain locations, such as Rikuzentakata (fold 2) and Ishinomaki (fold 0).

This increased sensitivity could be attributed to two factors, namely the training size and the complexity of the output. The smaller size of the training set may lead to higher variance in model performance, as the model has less data to learn from, making it more sensitive to the specific events included in each training fold. The onshore surrogate is tasked with predicting inundation, which is inherently more complex and variable compared to nearshore waveforms. This complexity can further lead to greater variation in model performance across different folds, especially with limited training data. In summary, the observed variance in MSE across folds is not random but is influenced by the size and complexity of the training data. For the nearshore surrogate, the model demonstrates stable performance, while the onshore surrogate shows some sensitivity, which is expected given the challenging nature of the inundation predictions.

We also record the maximum elevation values from the surrogate for each test event in Sect. 2.1.2 and compare them with the simulation values of the GeoClaw tsunami model described in Sect. 2.1.1 using the scatter plots of Figs. 10 and 11. The predictions are in good agreement, well correlated with the simulation values, and lay close to the true line. In the case of the nearshore surrogates, we notice larger uncertainty for Sendai compared to the other test sites, while
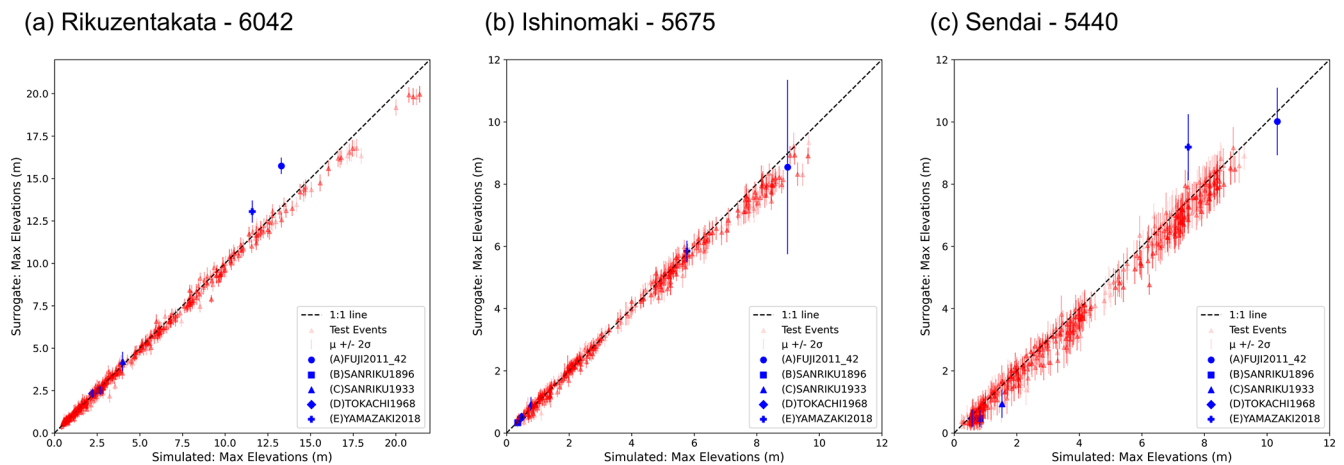
Figure 10. Prediction of maximum water levels from the nearshore surrogate for test and historic events.
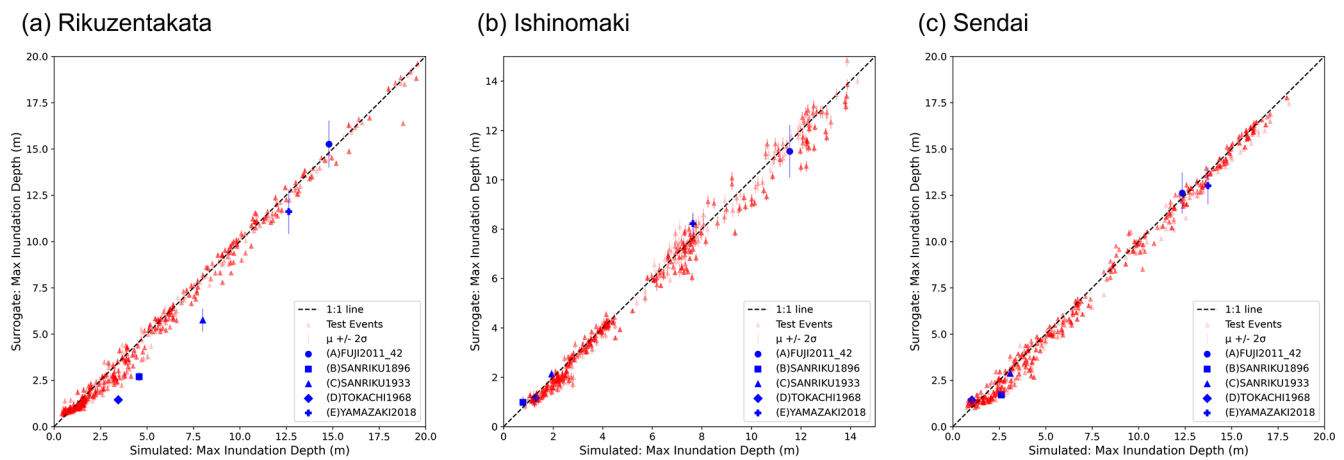
Figure 11. Prediction of maximum inundation depth from the onshore surrogate for test and historic events.

for the onshore surrogates Ishinomaki has more uncertainty. The plot also marks the prediction for the five historic test events, discussed further in Sect. 4.4. The predictions of the maximum elevation for the waveforms and the maxima of the maximum inundation depth show a good fit with little underestimation or overestimation (Figs. 10 and 11) and highlight the generalisation of the model across unseen data.

4.2 Nearshore approximation

The nearshore surrogate predicts waveforms of 4 h duration; the time series predictions for gauge 6042 at Rikuzentakata, and gauge 5675 at Ishinomaki, gauge 5440 at Sendai for five random events from the fold 0 test set are shown in Fig. 12. These figures show the mean prediction along with the 2 standard deviation uncertainty bands for all three gauges. This uncertainty represents the epistemic uncertainty of the ML model from a variational latent space. They show good agreement with the simulation, but the simulation values do not always lie within the uncertainty band of the prediction,

as seen in event 319 at Rikuzentakata for the full time period. Examples such as Type B (event 14) that show an excellent match and a narrow uncertainty range can be linked to training events sharing the same fault location but different slip distributions for the test event. Furthermore, the uncertainty range provides a useful indicator of the stability and sensitivity of the selection of model parameters and latent variables. We also evaluate the performance of the mean prediction using the evaluation metrics for the time series in Table 4. There is a relative peak deviation of up to 0.2 and a maximum peak delay of up to 2 h when the magnitude of the highest peak is not captured accurately.

4.3 Onshore approximation

With the onshore surrogate, we predict the maximum inundation depth at the selected grid location for the three test sites. The prediction maps at Rikuzentakata for five events from the fold 0 test set are shown in Fig. 14. We plot the true GeoClaw simulation values and the mean predictions of
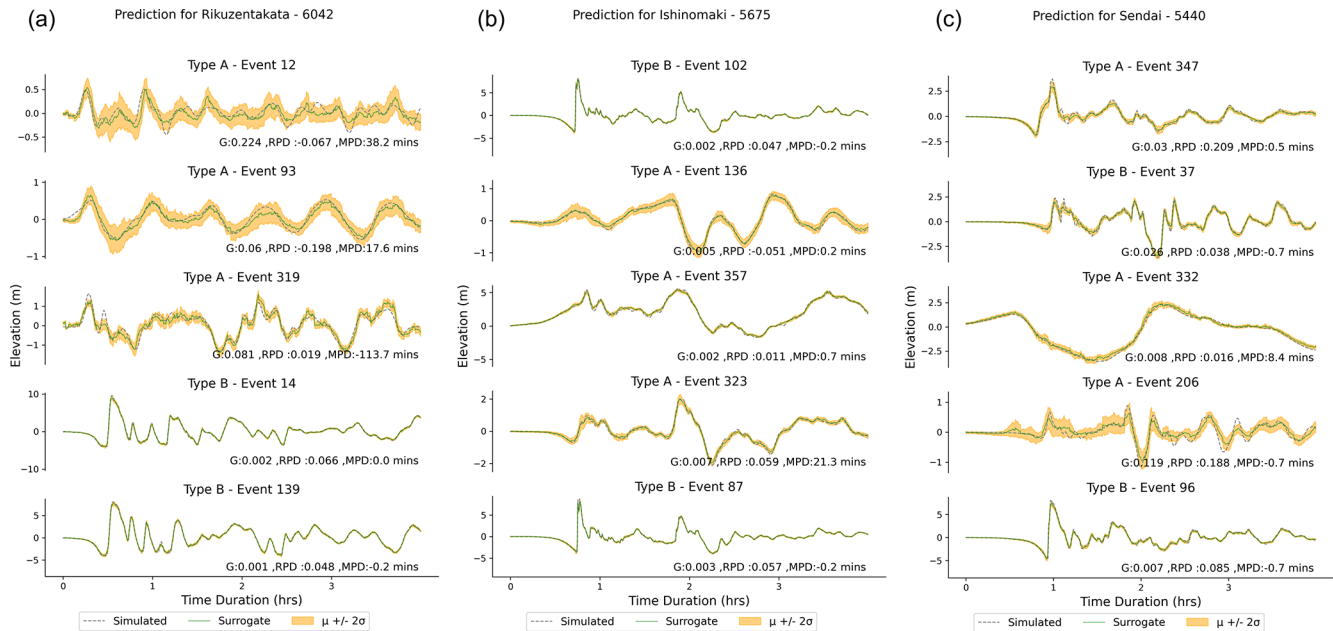
Figure 12. Prediction examples from the nearshore surrogate for the DOE test events.

Table 4. Summary of evaluation metrics and formulas.

Metric	Formula	Focus
Mean squared error (MSE)	$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$	Mean error
Coefficient of determination (R^2)	$1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}$	Correlation or dependence
Goodness of fit (G)	$1 - \frac{2\sum_{i=1}^{n}(y_i \cdot \hat{y}_i)}{\sum_{i=1}^{n}y_i^2 + \sum_{i=1}^{n}\hat{y}_i^2}$	General performance
Relative peak deviation (RPD)	$\frac{y_{\text{peak}} - \hat{y}_{\text{peak}}}{y_{\text{peak}}}$	Delay in peak
Maximum peak delay (MPD)	$t_{y_{\text{peak,max}}} - t_{\hat{y}_{\text{peak,max}}}$	Peak of the waveform
Accuracy score (A)	$\frac{\text{True predictions}}{\text{Number of locations}}$	Flooded area/mapping
L2Norm	$\sqrt{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$	Magnitude of event

the surrogate in columns 1 and 2. The misfit between simulation with the mean ML prediction and ±2 standard deviation values are plotted in columns 3–5. The uncertainty in the on-shore surrogate leads to variability in both the mapped inundation area and the water depth at each grid location for the predictions. This uncertainty depends on the ML parameters and the latent variables. The predictions show good agreement with the simulations, though there is a tendency to have some artefact flooding, i.e. to predict inundation at grids disconnected from the main flood extent which is an unphysical characteristic of the surrogate. Similar maps for Ishinomaki and Sendai are made available in Tables S1 and S2.

We also evaluate the performance of the mean prediction ±2 standard deviation using the evaluation metrics for the time series in Table 4. The accuracy score (A) focuses on the identification of the flooded area with depth greater than 10 cm or not, while the goodness of fit (G) provides an assessment of the correct prediction of the maximum inundation depth for the location with depth greater than 10 cm. The general tendency is that misfit reduces when using mean minus 2 standard deviation values, highlighting some overestimation in the mean prediction. We plot the distribution of the performance metrics G and A for two example events of Type A and Type B for Rikuzentakata using the ensemble of the predictions to show how the ensemble captures predictions close to the desired simulation results in Figs. S9 and S10. Across the different test events and ensemble predictions, 93.397 % have an accuracy score greater than 0.9 and G values less than 0.1. The spread of the G and A values for different depth classes is plotted in Fig. 16.

Table 5. Results of the *k*-fold cross-validation using the mean squared error loss metric for the two types of surrogate models tested at the three test sites.

	Region	MSE				
		Fold 0	Fold 1	Fold 2	Fold 3	Fold 4
Nearshore surrogate	Rikuzentakata	0.098	0.057	0.042	0.122	0.111
	Ishinomaki	0.042	0.065	0.061	0.071	0.053
	Sendai	0.197	0.087	0.067	0.168	0.059
Onshore surrogate	Rikuzentakata	0.179	0.406	1.129	0.262	0.169
	Ishinomaki	0.886	0.349	0.327	0.282	0.373
	Sendai	0.35	0.263	0.377	0.261	0.576

4.4 Generalisation ability: predicting historical test events

Generalisation of the ML surrogates ensures that they would perform well on unseen data by capturing the underlying characteristics of the data rather than memorising and overfitting the training data. To test the generalisation, we examined the prediction for the historical test events described in Sect. 2.1.3. These results are the predictions using all five VED models trained on different data subsets in the cross-validation exercise, providing a wider ensemble with 500 predictions compared to the 100 predictions from a single VED model in the previous section.

Figure 13 plots the time series prediction for these events for the three regions, similar to the fold test events in Fig. 12. Compared to the results of the fold test events previously described, the peak values and their timing are well captured across the test sites and events. As observed earlier, the G values tend to be higher than desirable for the smaller events (B, C, and D). The surrogate is unable to predict the high-frequency characteristics (event B and C) but is able to capture it in the uncertainty bands of the waveform ensemble. When comparing the mean prediction with the simulation, the G values range 0.054–0.225 for Rikuzentakata, 0.071–0.442 for Ishinomaki, and 0.07–0.411 for Sendai. Similar to the prediction from the test fold, underestimation in some of the wave peaks results in an MPD of up to 107 and 65 min in Rikuzentakata and Ishinomaki, respectively, and 193 min in Sendai.

Figure 15 plots the inundation map for Rikuzentakata, similar to the fold test events shown in Fig. 14. For the large-magnitude 2011 Tohoku events (test events A and E), the accuracy metric for the flood mapping is high: 0.9 using the mean prediction for event A and 0.93 for event E. However, there is significant underprediction in flood depth and misfit of up to 5 m. For the smaller-magnitude events related to the Sanriku events (test events B and C), there is overestimation in the flood mapping, with many locations having predictions of small depths, resulting in a lower accuracy metric of 0.83 for event C and 0.65 for event B. For the 1933 Tokachi-oki event (test event D), accuracy in flood mapping is high with

a value of 0.96, but flood depths are underestimated with a low G value of 0.623.

For Ishinomaki, the prediction performs well in characterising the small events in terms of the depth and area flooded. There is an underestimation in flood depths and mapping area for event A and an overestimation in flood area for event E. For Sendai, the accuracy in the flood mapping footprint is high but significantly underestimated in the prediction of the flood depth for the large-magnitude 2011 Tohoku events (test events A and E). Refer to Figs. S5–S8 for the flood prediction and evaluation metrics.

There is high accuracy in the mapping of the flood area, but there is a tendency to misrepresent the flood depths in the prediction by the onshore model for the test events. We examine this misfit by plotting the accuracy score (A) for flood mapping and goodness of fit (G) for the different depth classes of 0.1–1, 1–2, 2–5, 5–10, and 10–15 m for test events in Sect. 2.1.3 individually and events in Sect. 2.1.2 lumped together, at the three test sites as shown in Fig. 16.

To provide an overview of the surrogate model's behaviour, we conducted a comparative analysis between the mean prediction of the test events from the Design of Experiments (DOE) and historical events. In Fig. 17a, we plot the prediction performance, assessed by G, with the magnitude of the event represented by the L2Norm. Our analysis reveals a degradation in G values for events with L2Norm values below 10. This trend can be attributed to the optimisation process driven by the use of MSE as the loss function, which tends to prioritise accuracy for larger values. Furthermore, in Fig. 17b, we compare the coefficient of determination between predictions and simulation against G as a metric for prediction performance. This visualisation effectively identifies areas where the model performs exceptionally well, but also highlights events where low G value can have a poor COD value due to poor prediction for a portion of the time series, as seen for test event A for Rikuzentakata, or overprediction as seen for the maximum inundation depths for event B in Rikuzentakata.
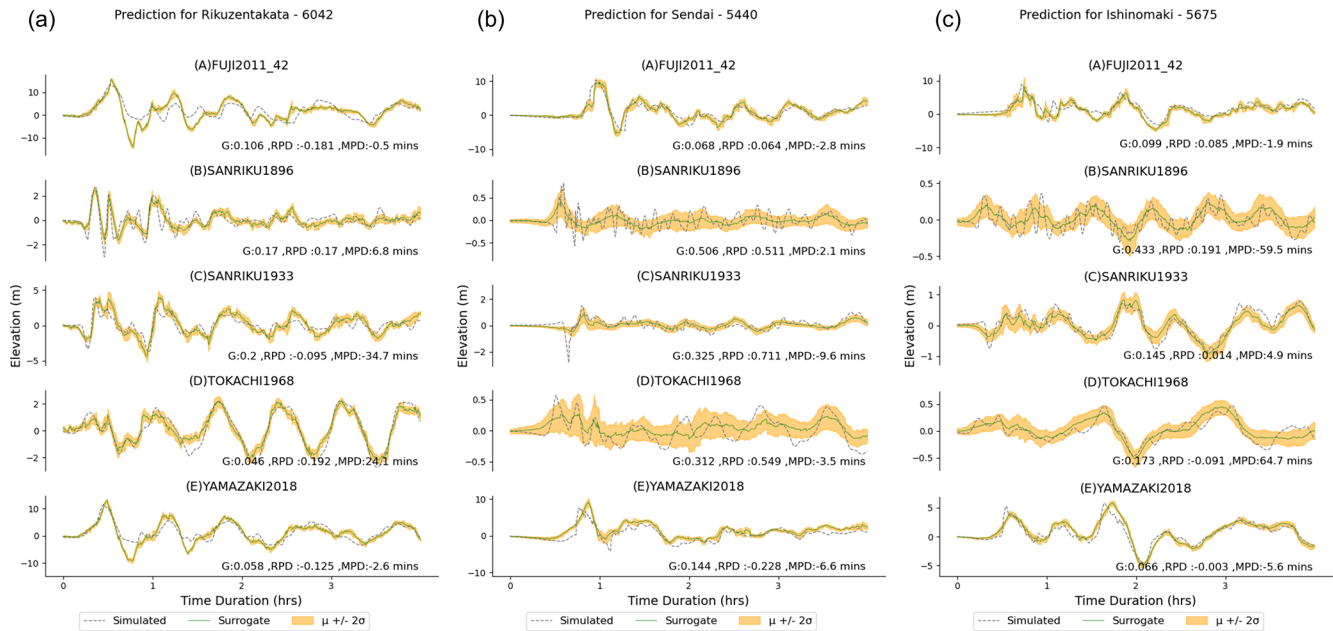
Figure 13. Predictions from the nearshore surrogate for the historical events.

5   Discussion and conclusions

A key challenge in using neural network surrogates for probabilistic tsunami hazard assessment (PTHA) is their limited validation in generalising beyond the specific event types used during training. For instance, their performance on events from different regions, involving multiple sources and mechanisms, has not been sufficiently validated in previous studies. By testing ML surrogates on a diverse set of events, this research broadens their applicability and makes a significant contribution to validating the generalisation capabilities of ML surrogates for PTHA. Furthermore, training datasets often require thousands of events from each source region to fully capture the inherent variability. In many cases, previous studies using ML have prioritised accurate predictions for larger-magnitude events, which are crucial for early-warning systems. However, for a comprehensive PTHA, it is equally important that surrogates can predict both large- and small-magnitude events with high accuracy. Recognising this, we designed two specialised ML surrogates to address the distinct challenges of approximating nearshore and onshore tsunami hazards that help offset the related computational costs, while overcoming the limitations of imbalanced and limited training datasets. The nearshore surrogate predicts the time history of tsunami waves at the shore, while the onshore surrogate predicts the inundation depths across vast locations overland. The hybrid ensemble approach introduced here leverages model and parameter sensitivity to enhance prediction accuracy, marking a step forward in integrating uncertainty quantification into ML-based PTHA. This expanded capability, which has been underexplored in

prior research, is essential for extending tsunami hazard and risk assessment.

The data preparation, model architecture, and training procedures developed here represent key advancements in overcoming the limitations discussed. We train and test our surrogates using a synthetic tsunami dataset modelled using ruptures with both heterogenous and uniform slip. Additionally, the surrogate was also tested for historical events modelled with different rupture models and source locations to assess their effectiveness and investigate the influence of such uncertainties in the initial displacement on surrogate prediction.

To enhance prediction information, we implemented a hybrid ensemble approach that uses multiple variational models. The effectiveness of surrogate models depends on the tuning of latent variables, which vary in the type of prediction parameter, size of output, and complexity of the region. Our analysis suggests that more latent variables are necessary for the nearshore surrogate compared to the onshore surrogate; i.e. more information needs to be compressed for time series prediction than inundation. Also, the complexity of the region has more influence than the size of the prediction locations. Cross-validation testing confirms the feasibility of using the surrogate in hazard analyses.

The surrogate models exhibit excellent performance in predicting time series data, effectively characterising waveforms for different locations. Despite the peculiar source characteristics and input time series compared to the training dataset, the models accurately capture the temporal dynamics of the events. However, high-frequency characteristics are not captured for some test events, which can be due to the choice of shallow or low number of convolution layers
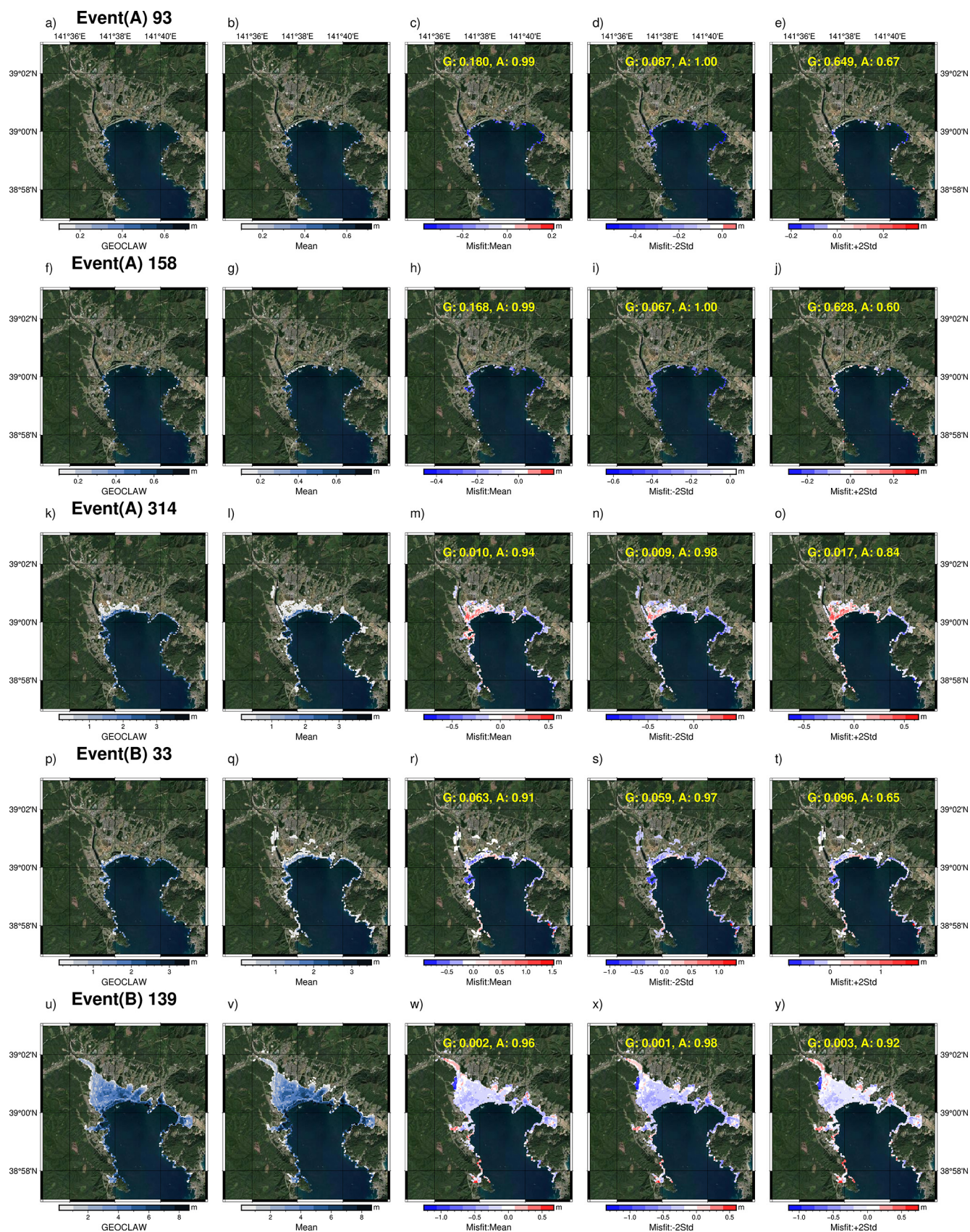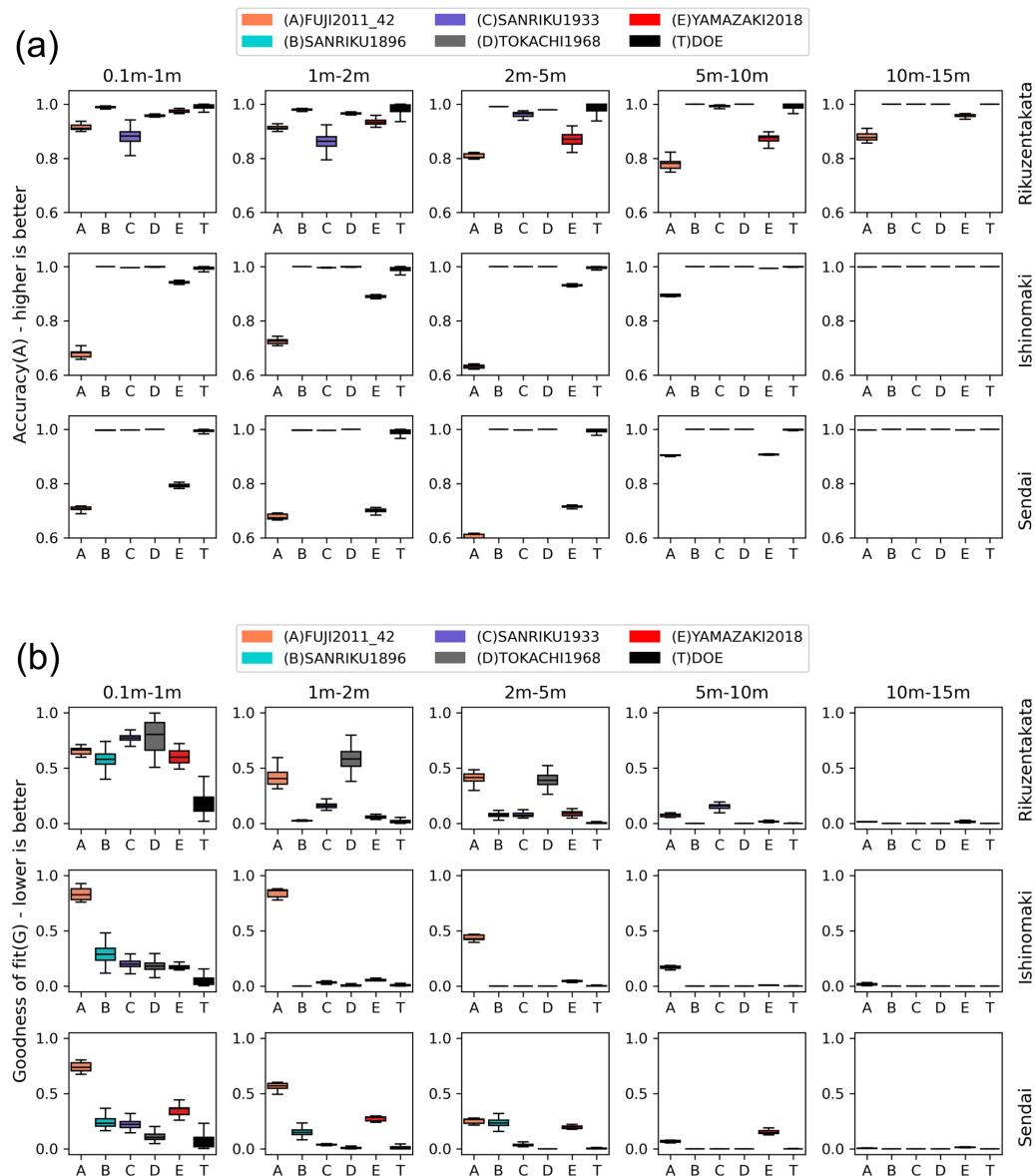
Figure 14. Prediction examples from the onshore surrogate at Rikuzentakata for the DOE test events (Basemap from Esri World Imagery).
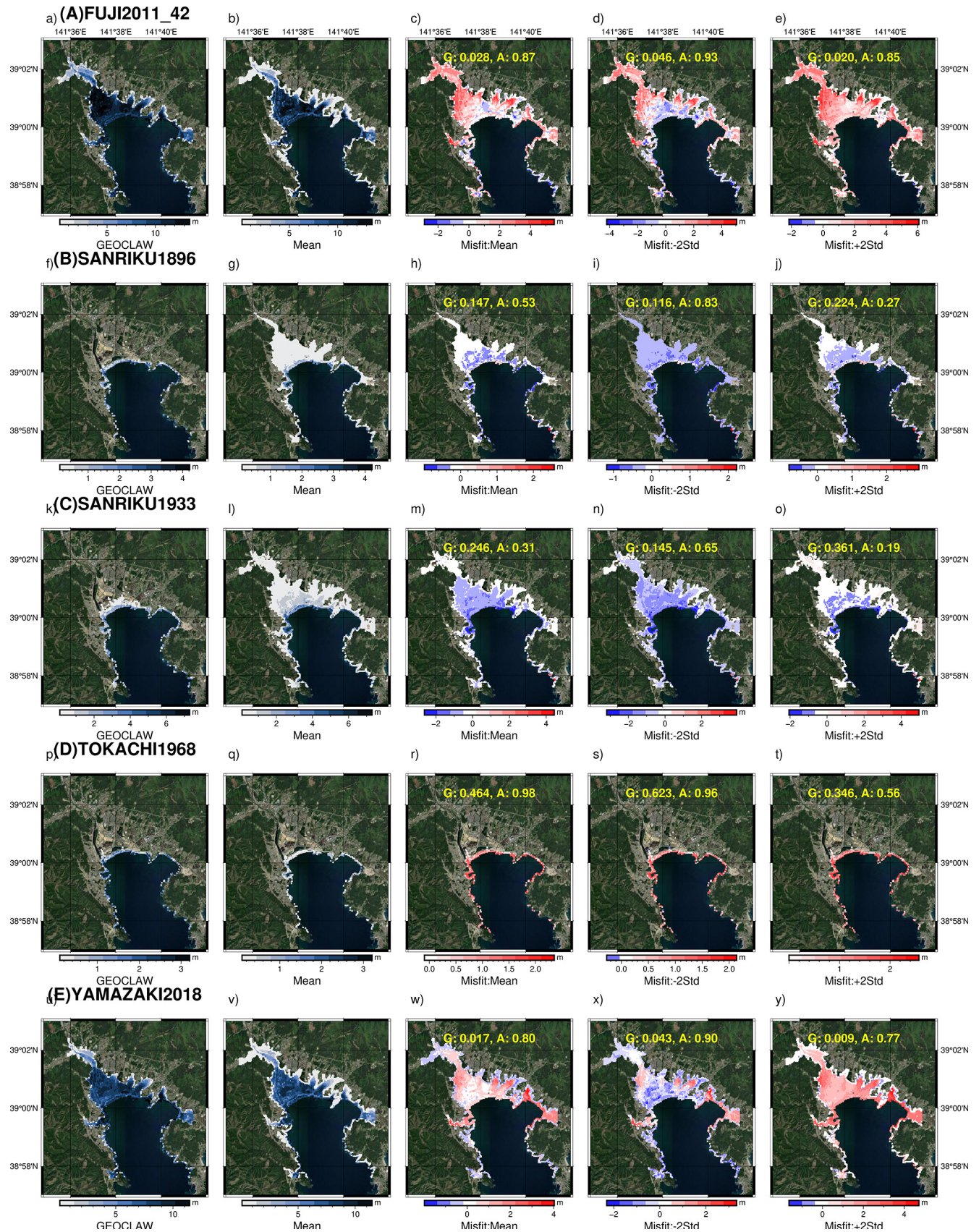
Figure 15. Predictions from the onshore surrogate at Rikuzentakata for the historical events (Basemap from Esri World Imagery).
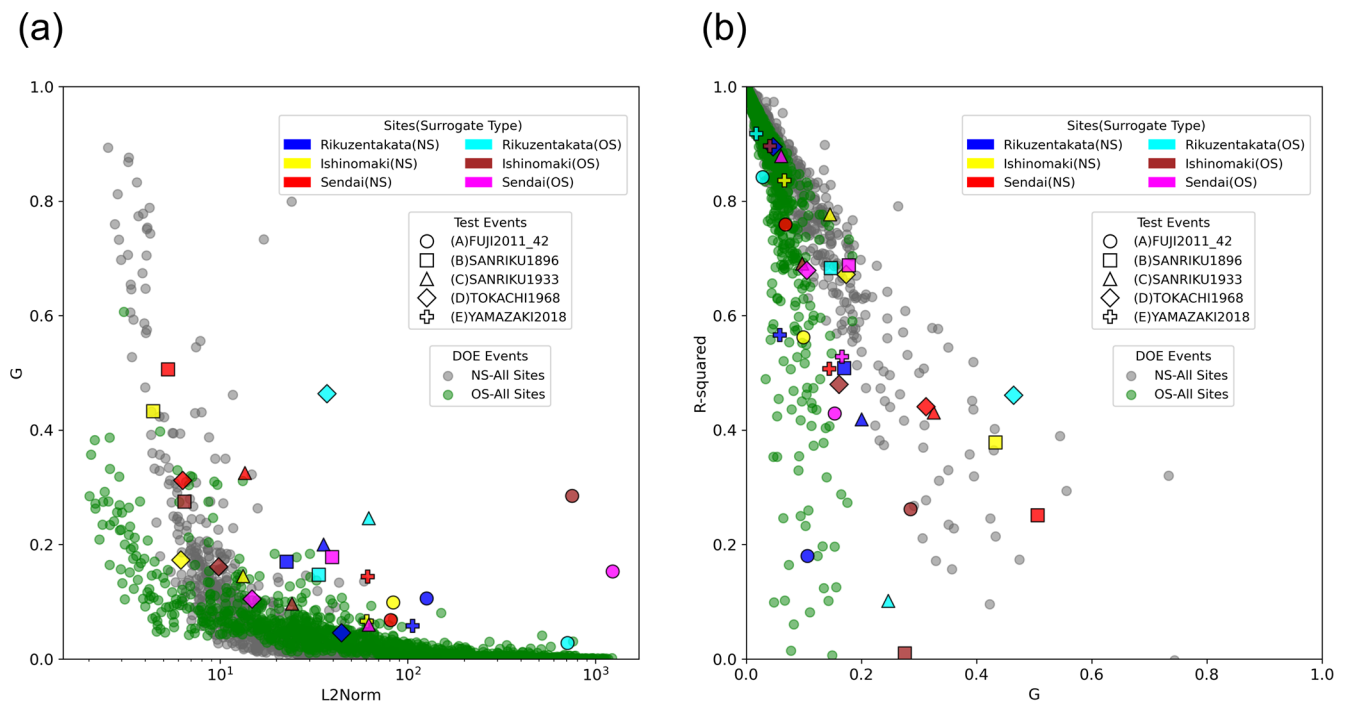
Figure 16. Assessing the variability of prediction performance across different depth classes, events, and test sites: **(a)** accuracy and **(b)** goodness of fit.

and limited information in the training dataset. There are also certain portions with misrepresented phases, which could be due to not including the local deformation explicitly in the learning framework but instead as a pre-processing step.

Inundation prediction, however, presents greater challenges due to the large number of locations and asymmetry in flood occurrence across the prediction locations, resulting in artefacts or disconnected flooding points and underprediction in depth in the generalisation testing. Unlike time series prediction with fixed time steps, inundation prediction requires mapping to the large number of predicted locations (ranging from 6648 to 129 941) tested in the study. Addressing these challenges requires further optimisation and refinement of

the model architecture, which will be investigated in future studies.

While artefact flooding can be removed with simple post-processing routines, there is significant potential for further improvement in the onshore decoder. We set the grid resolution in our inundation simulation at 50 m due to our computational constraints; using a higher resolution of around 5–10 m would mean significantly more points. Implementing multi-head decoders for prediction will provide more efficient and accurate predictions. Other architectural improvements and training methods when using imbalanced datasets, such as weighted sampling, upsampling of data, and curriculum learning, should also be investigated but are beyond the

(a)

(b)

Figure 17. Assessing the variability of performance given by G with respect to: **(a)** the magnitude of the event as indicated by the L2Norm and **(b)** the coefficient of determination R^2.

scope of this article. A customised loss function balancing predictions across small and large values by weighting could further enhance model performance and address discrepancies observed in inundation prediction.

Our surrogate model demonstrates stable performance, predicting both small and large events. We observe good results for events outside the training parameter space. This highlights the versatility of surrogate models in the prediction of events with different magnitudes and characteristics. Although our training data did not consider outer rise mechanisms, the model generalises well as also seen with the 1933 Sanriku Tsunami. This underscores the promise of further improving the method by incorporating a broader range of training data and using better source modelling techniques, as done by Liu et al. (2021) and Núñez et al. (2022). Our analyses show that the nearshore model outperforms the onshore model in terms of generalisation for the historical events.

The training information from our DOE is constrained by both the quality and quantity of events. Recognising this limitation is crucial for guiding future improvements in the experimental design and training dataset along with advances in the model architecture and training. First, the geographic focus is limited to the Tohoku subduction source region and modelled with a simple scheme, restricting the diversity of the training data and impacting the model's ability to generalise to other regions or varied tsunami scenarios, such as the historic test events (B, C, and D). Second, there is an

event imbalance for inundation, particularly for the onshore surrogate, where more events of large inundation are needed to provide sufficient training scenarios at locations far from the coast, which are rarely inundated in the training dataset. Finally, the generalisation test on the onshore surrogate highlights varying prediction accuracies across different test sites, at Rikuzentakata, Sendai, and Ishinomaki. These variations reflect the complexities and limitations of the DOE, where certain test sites with more complex inundation patterns are not well represented in the training data, leading to less accurate predictions.

To improve our understanding of ML models in tsunami prediction, further efforts are needed in acquiring more extensive training and testing datasets, conducting benchmarking and comparison studies with other surrogates or test regions, and making them available and open source. In summary, our study demonstrates the potential of surrogate models in accurately predicting tsunami hazard variables while also highlighting areas for further refinement to improve model performance and reliability in hazard assessment tasks.

Understanding the impact of the size of the training dataset and the variability is important and should be examined. Topography and coastal morphology, the number of model inputs, the type of prediction, and the tsunami sources considered (among other factors) all contribute to the size of training data and compute resources needed, as well as the resulting surrogate prediction performance. The runtime informa-

Table 6. Runtime information for tsunami numerical simulations using GeoClaw.

Total cell updates	Device type	Parallelisation	Time taken per event (h)	No. of events	Total compute time (h)
0.531×10^{11}	CPU	10 CPU threads	3.45	564	1945

Table 7. Runtime information for machine learning training for the different region-specific surrogates.

	Region	Max epoch	Time taken per training (s)	No. of folds	Total training time (min)
Nearshore surrogate	Rikuzentakata	3000	54	5	4.5
	Ishinomaki	3000	72	5	6
	Sendai	4000	73	5	6
Onshore surrogate	Rikuzentakata	20 000	156	5	13
	Ishinomaki	20 000	167	5	14
	Sendai	20 000	177	5	15
				Total	58.5 (1 h)

tion for the tsunami numerical simulation using an Intel Xeon Silver 4216 CPU device with 2.1 GHz and 313 GB of RAM is provided in Table 6, and information for the ML training from this work is provided in Table 7 using an NVIDIA A100 80 GB GPU device.

The computation time for training an ML surrogate is in minutes; this results in a remarkable efficiency gain of approximately 2000 times during the model inference when applying the ML surrogate, as compared to running the full simulation. We should also consider that multiple training runs are needed for fine-tuning the model architecture, hyperparameters, and other choices at hand for the ML model which were not quantified here. Standardising the model architecture and procedures along with the use of pre-trained models for transfer learning can help minimise these training-related costs.

The nearshore approximation approach is useful as a dynamic hazard proxy at the coast where considering the arrival time of a tsunami and its peaks is important for studying applications such as evacuation planning. However, direct application may be limited by the need to train the model for each location. This may hinder the application at scale, but the option to predict the inundation maps helps to efficiently assess large regions. Our approach can help extend offshore tsunami hazard information (Davies et al., 2018; Basili et al., 2021), available at deep offshore points, to the much-needed onshore hazard and risk assessments, with a relatively limited number of simulations and associated computational costs. The learning from this study can easily be adapted for other intensity measures or parameters like velocity or momentum flux, but also finds use in early-warning settings where direct tsunami sensors are not yet available. Similar computational challenges also exist for other coastal flood hazards like storm surges, riverine coastal flooding, tropical cyclone

rainfall, and compound flood drivers, and the use of such surrogates can also be investigated in these cases using different sets of input parameters.

Code availability. The tsunami simulations use GeoClaw; PyTorch was used as the ML framework in Python, and the other analysis was conducted using bespoke scripts in Python. Maps were prepared using the PyGMT package in Python and QGIS. All scripts and codes used in the project and those for the preparation of the paper are available at https://github.com/naveenragur/tsunami-surrogates.git (last access: 7 May 2025; https://doi.org/10.5281/zenodo.15337936, Ragu Ramalingam, 2025).

Data availability. The Japan Cabinet Office data from the modelling study by special Investigation Committee on Subduction-type Earthquakes around the Japan Trench and Chishima Trench (Central Disaster Prevention Council) are publicly available at https://www.geospatial.jp/ckan/organization/naikakufu-002 (Japan Cabinet Office, 2016). The COP-DEM data are available through OpenTopography from https://portal.opentopography.org/raster?opentopoID=OTSDEM.032021.4326.3 (last access: 5 May 2025; https://doi.org/10.5069/G9028PQB, European Space Agency, 2024). The observed water level, astronomical tide level, and tidal level deviation data observed from 2011 off the Pacific coast of Tohoku earthquake tsunami are publicly available from NOWPHAS at https://nowphas.mlit.go.jp/pastdata (NOWPHAS, 2011). The 2011 earthquake tsunami information for off the Pacific coast of Tohoku – field survey results used for the validation of the GeoClaw model – is publicly available from the Coastal Engineering Committee of the Japan Society of Civil Engineers website at https://www.coastal.jp/tsunami2011 (CEC, 2012). The fault parameters used to model the initial displacement for the historic events are publicly available at the links cited in the

references. Large input files for GeoClaw and their post-processed outputs used to train and test the surrogate models are available at https://doi.org/10.5281/zenodo.10817116 (Ragu Ramalingam, 2024).

Author contributions. NRR, KJ, MP, and MLVM conceived and designed the research. NRR conceived the experiments, carried out the simulation, developed the ML model, and performed the training. All authors analysed and reviewed the results and wrote the manuscript.

References

Abrahams, L. S., Krenz, L., Dunham, E. M., Gabriel, A.-A., and Saito, T.: Comparison of methods for coupled earthquake and tsunami modelling, Geophys. J. Int., 234, 404–426, https://doi.org/10.1093/gji/ggad053, 2023.

Aki, K. and Richards, P. G.: Quantitative Seismology, University Science Books, 2nd edn., ISBN 0935702962, 2002.

Basili, R., Brizuela, B., Herrero, A., Iqbal, S., Lorito, S., Maesano, F. E., Murphy, S., Perfetti, P., Romano, F., Scala, A., Selva, J., Taroni, M., Tiberti, M. M., Thio, H. K., Tonini, R., Volpe, M., Glimsdal, S., Harbitz, C. B., Løvholt, F., Baptista, M. A., Carrilho, F., Matias, L. M., Omira, R., Babeyko, A., Hoechner, A., Gürbüz, M., Pekcan, O., Yalçıner, A., Canals, M., Lastras, G., Agalos, A., Papadopoulos, G., Triantafyllou, I., Benchekroun, S., Agrebi Jaouadi, H., Ben Abdallah, S., Bouallegue, A., Hamdi, H., Oueslati, F., Amato, A., Armigliato, A., Behrens, J., Davies, G., Di Bucci, D., Dolce, M., Geist, E., Gonzalez Vida, J. M., González, M., Macías Sánchez, J., Meletti, C., Ozer Sozdinler, C., Pagani, M., Parsons, T., Polet, J., Power, W., Sørensen,

M., and Zaytsev, A.: The Making of the NEAM Tsunami Hazard Model 2018 (NEAMTHM18), Front. Earth Sci., 8, 616594, https://doi.org/10.3389/feart.2020.616594, 2021.

Behrens, J. and Dias, F.: New computational methods in tsunami science, Philos. T. Roy. Soc. A, 373, 20140382, https://doi.org/10.1098/rsta.2014.0382, 2015.

Behrens, J., Løvholt, F., Jalayer, F., Lorito, S., Salgado-Gálvez, M. A., Sørensen, M., Abadie, S., Aguirre-Ayerbe, I., Aniel-Quiroga, I., Babeyko, A., Baiguera, M., Basili, R., Belliazzi, S., Grezio, A., Johnson, K., Murphy, S., Paris, R., Rafliana, I., De Risi, R., Rossetto, T., Selva, J., Taroni, M., Del Zoppo, M., Armigliato, A., Bureš, V., Cech, P., Cecioni, C., Christodoulides, P., Davies, G., Dias, F., Bayraktar, H. B., González, M., Gritsevich, M., Guillas, S., Harbitz, C. B., Kânoğlu, U., Macías, J., Papadopoulos, G. A., Polet, J., Romano, F., Salamon, A., Scala, A., Stepinac, M., Tappin, D. R., Thio, H. K., Tonini, R., Triantafyllou, I., Ulrich, T., Varini, E., Volpe, M., and Vyhmeister, E.: Probabilistic Tsunami Hazard and Risk Analysis: A Review of Research Gaps, Front. Earth Sci., 9, 628772, https://doi.org/10.3389/feart.2021.628772, 2021.

Clawpack Development Team: Clawpack Version 5.7.1, Zenodo [code], https://doi.org/10.5281/zenodo.4025432, 2020.

Cesario, E., Giampà, S., Baglione, E., Cordrie, L., Selva, J., and Talia, D.: Forecasting Tsunami Waves Using Regression Trees, in: 2023 International Conference on Information and Communication Technologies for Disaster Management (ICT-DM), 13–15 September 2023, Cosenza, Italy, IEEE, ISBN 9798350319514, 1–7, https://doi.org/10.1109/ICT-DM58371.2023.10286955, 2023.

CEC: 2011 Tohoku Earthquake Tsunami field survey results, CEC (Coastal Engineering Committee, Japan Society of Civil Engineers) [data set], https://www.coastal.jp/tsunami2011 (last access: 5 May 2025), 2012.

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

Davies, G., Weber, R., Wilson, K., and Cummins, P.: From offshore to onshore probabilistic tsunami hazard assessment via efficient Monte Carlo sampling, Geophys. J. Int., 230, 1630–1651, https://doi.org/10.1093/gji/ggac140, 2022.

de Baar, J. H. S. and Roberts, S. G.: Multifidelity Sparse-Grid-Based Uncertainty Quantification for the Hokkaido Nansei-oki Tsunami, Pure Appl. Geophys., 174, 3107–3121, https://doi.org/10.1007/s00024-017-1606-y, 2017.

de la Asunción, M.: Prediction of Tsunami Alert Levels Using Deep Learning, Earth and Space Science, 11, e2023EA003385, https://doi.org/10.1029/2023EA003385, 2024.

European Space Agency: Copernicus Global Digital Elevation Model, OpenTopography [data set], https://doi.org/10.5069/G9028PQB, 2024.

Fauzi, A. and Mizutani, N.: Machine Learning Algorithms for Real-time Tsunami Inundation Forecasting: A Case Study in Nankai Region, Pure Appl. Geophys., 177, 1437–1450, https://doi.org/10.1007/s00024-019-02364-4, 2019.

Folch, A., Abril, C., Afanasiev, M., Amati, G., Bader, M., Badia, R. M., Bayraktar, H. B., Barsotti, S., Basili, R., Bernardi, F., Boehm, C., Brizuela, B., Brogi, F., Cabrera, E., Casarotti, E.,

Castro, M. J., Cerminara, M., Cirella, A., Cheptsov, A., Cone-jero, J., Costa, A., de la Asunción, M., de la Puente, J., Djuric, M., Dorozhinskii, R., Espinosa, G., Esposti-Ongaro, T., Farnós, J., Favretto-Cristini, N., Fichtner, A., Fournier, A., Gabriel, A.-A., Gallard, J.-M., Gibbons, S. J., Glimsdal, S., González-Vida, J. M., Gracia, J., Gregorio, R., Gutierrez, N., Halldorsson, B., Hamitou, O., Houzeaux, G., Jaure, S., Kessar, M., Krenz, L., Krischer, L., Laforet, S., Lanucara, P., Li, B., Loren-zino, M. C., Lorito, S., Løvholt, F., Macedonio, G., Macías, J., Marín, G., Martínez Montesinos, B., Mingari, L., Moguilny, G., Montellier, V., Monterrubio-Velasco, M., Moulard, G. E., Na-gaso, M., Nazaria, M., Niethammer, C., Pardini, F., Pienkowska, M., Pizzimenti, L., Poiata, N., Rannabauer, L., Rojas, O., Ro-driguez, J. E., Romano, F., Rudyy, O., Ruggiero, V., Samfass, P., Sánchez-Linares, C., Sanchez, S., Sandri, L., Scala, A., Schaef-fer, N., Schuchart, J., Selva, J., Sergeant, A., Stallone, A., Ta-roni, M., Thrastarson, S., Titos, M., Tonelllo, N., Tonini, R., Ulrich, T., Vilotte, J.-P., Vöge, M., Volpe, M., Aniko Wirp, S., and Wössner, U.: The EU Center of Excellence for Exascale in Solid Earth (ChEESE): Implementation, results, and roadmap for the second phase, Future Gener. Comp. Sy., 146, 47–61, https://doi.org/10.1016/j.future.2023.04.006, 2023.

Fujii, Y., Satake, K., Sakai, S., Shinohara, M., and Kanazawa, T.: Tsunami source of the 2011 off the Pacific coast of Tohoku Earthquake, Earth Planets Space, 63, 815–820, https://doi.org/10.5047/eps.2011.06.010, 2011.

Fukutani, Y., Moriguchi, S., Terada, K., and Otake, Y.: Time-Dependent Probabilistic Tsunami Inundation Assessment Us-ing Mode Decomposition to Assess Uncertainty for an Earth-quake Scenario, J. Geophys. Res.-Oceans, 126, e2021JC017250, https://doi.org/10.1029/2021JC017250, 2021.

Fukutani, Y., Yasuda, T., and Yamanaka, R.: Efficient probabilis-tic prediction of tsunami inundation considering random tsunami sources and the failure probability of seawalls, Stoch. Env. Res. Risk A., 37, 2053–2068, https://doi.org/10.1007/s00477-023-02379-3, 2023.

Gailler, A., Hébert, H., Schindelé, F., and Reymond, D.: Coastal Amplification Laws for the French Tsunami Warning Cen-ter: Numerical Modeling and Fast Estimate of Tsunami Wave Heights Along the French Riviera, Pure Appl. Geophys., 175, 1429–1444, https://doi.org/10.1007/s00024-017-1713-9, 2018.

Geist, E. L. and Parsons, T.: Probabilistic Analysis of Tsunami Haz-ards, Nat, Hazards, 37, 277–314, https://doi.org/10.1007/s11069-005-4646-z, 2006.

Gibbons, S. J., Lorito, S., Macías, J., Løvholt, F., Selva, J., Volpe, M., Sánchez-Linares, C., Babeyko, A., Brizuela, B., Cirella, A., Castro, M. J., de la Asunción, M., Lanucara, P., Glims-dal, S., Lorenzino, M. C., Nazaria, M., Pizzimenti, L., Romano, F., Scala, A., Tonini, R., Manuel González Vida, J., and Vöge, M.: Probabilistic Tsunami Hazard Analysis: High Performance Computing for Massive Scale Inundation Simulations, Front. Earth Sci., 8, 591549, https://doi.org/10.3389/feart.2020.591549, 2020.

Gibbons, S. J., Lorito, S., de la Asunción, M., Volpe, M., Selva, J., Macías, J., Sánchez-Linares, C., Brizuela, B., Vöge, M., Tonini, R., Lanucara, P., Glimsdal, S., Romano, F., Meyer, J. C., and Løvholt, F.: The Sensitivity of Tsunami Impact to Earthquake Source Parameters and Manning Friction in High-

Resolution Inundation Simulations, Front. Earth Sci., 9, 757618, https://doi.org/10.3389/feart.2021.757618, 2022.

Gica, E., Spillane, M. C., Titov, V. V., Chamberlin, C. D., and New-man, J. C.: Development of the forecast propagation database for NOAA's Short-term Inundation Forecast for Tsunamis (SIFT), NOAA Technical Memorandum OAR PMEL-139, NOAA, https://repository.library.noaa.gov/view/noaa/11079 (last access: 4 May 2025), 2008.

Giles, D., Gopinathan, D., Guillas, S., and Dias, F.: Faster Than Real Time Tsunami Warning with Associated Hazard Uncertainties, Front. Earth Sci., 8, 597865, https://doi.org/10.3389/feart.2020.597865, 2021.

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

Goda, K. and De Risi, R.: Probabilistic Tsunami Loss Estimation Methodology: Stochastic Earthquake Sce-nario Approach, Earthq. Spectra, 33, 1301–1323, https://doi.org/10.1193/012617eqs019m, 2017.

Goda, K., Mai, P. M., Yasuda, T., and Mori, N.: Sensitivity of tsunami wave profiles and inundation simulations to earthquake slip and fault geometry for the 2011 Tohoku earthquake, Earth Planets Space, 66, 105, https://doi.org/10.1186/1880-5981-66-105, 2014.

Gopinathan, D., Heidarzadeh, M., and Guillas, S.: Probabilis-tic quantification of tsunami current hazard using statisti-cal emulation, P. Roy. Soc. A-Math. Phy., 477, 20210180, https://doi.org/10.1098/rspa.2021.0180, 2021.

Grezio, A., Babeyko, A., Baptista, M. A., Behrens, J., Costa, A., Davies, G., Geist, E. L., Glimsdal, S., González, F. I., Griffin, J., Harbitz, C. B., LeVeque, R. J., Lorito, S., Løvholt, F., Omira, R., Mueller, C., Paris, R., Parsons, T., Polet, J., Power, W., Selva, J., Sørensen, M. B., and Thio, H. K.: Probabilistic Tsunami Hazard Analysis: Multiple Sources and Global Applications: Probabilis-tic Tsunami Hazard Analysis, Rev. Geophys., 55, 1158–1198, https://doi.org/10.1002/2017RG000579, 2017.

Grzan, D. P., Rundle, J. B., Wilson, J. M., Song, T., Ward, S. N., and Donnellan, A.: Tsunami Squares: Earthquake driven inundation mapping and validation by comparison to the Regional Ocean Modeling System, Progress in Disaster Science, 12, 100191, https://doi.org/10.1016/j.pdisas.2021.100191, 2021.

Gusman, A. R., Tanioka, Y., MacInnes, B. T., and Tsushima, H.: A methodology for near-field tsunami inundation forecasting: Ap-plication to the 2011 Tohoku tsunami, J. Geophys. Res.-Sol. Ea., 119, 8186–8206, https://doi.org/10.1002/2014JB010958, 2014.

Hayes, G. P., Moore, G. L., Portner, D. E., Hearne, M., Flamme, H., Furtney, M., and Smoczyk, G. M.: Slab2, a comprehen-sive subduction zone geometry model, Science, 362, 58–61, https://doi.org/10.1126/science.aat4723, 2018.

Imamura, F., Boret, S. P., Suppasri, A., and Muhari, A.: Recent oc-currences of serious tsunami damage and the future challenges of tsunami disaster risk reduction, Progress in Disaster Science, 1, 100009, https://doi.org/10.1016/j.pdisas.2019.100009, 2019.

Japan Cabinet Office: Modelling study data, Japan Cabinet Office, Central Disaster Prevention Council, Special Investigation Com-mittee on Subduction-type Earthquakes around the Japan Trench

and Chishima Trench [data set], https://www.geospatial.jp/ckan/organization/naikakufu-002 (last access: 5 May 2025), 2016.

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

Kingma, D. P. and Welling, M.: Auto-Encoding Variational Bayes, arXiv [preprint], https://doi.org/10.48550/arXiv.1312.6114, 9 January 2014.

Kotani, T., Tozato, K., Takase, S., Moriguchi, S., Terada, K., Fukutani, Y., Otake, Y., Nojima, K., Sakuraba, M., and Choe, Y.: Probabilistic tsunami hazard assessment with simulation-based response surfaces, Coast. Eng., 160, 103719, https://doi.org/10.1016/j.coastaleng.2020.103719, 2020.

Lee, J.-W., Irish, J. L., and Weiss, R.: Real-Time Prediction of Alongshore Near-Field Tsunami Runup Distribution From Heterogeneous Earthquake Slip Distribution, J. Geophys. Res.-Oceans, 128, e2022JC018873, https://doi.org/10.1029/2022JC018873, 2023.

LeVeque, R. J., George, D. L., and Berger, M. J.: Tsunami modelling with adaptively refined finite volume methods, Acta Numer., 20, 211–289, https://doi.org/10.1017/S0962492911000043, 2011.

Liu, C. M., Rim, D., Baraldi, R., and LeVeque, R. J.: Comparison of Machine Learning Approaches for Tsunami Forecasting from Sparse Observations, Pure Appl. Geophys., 178, 5129–5153, https://doi.org/10.1007/s00024-021-02841-9, 2021.

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

Løvholt, F., Griffin, J., and Salgado-Gálvez, M.: Tsunami Hazard and Risk Assessment on the Global Scale, in: Encyclopedia of Complexity and Systems Science, edited by: Meyers, R. A., Springer, Berlin, Heidelberg, ISBN 978-3-642-27737-5, 1–34, https://doi.org/10.1007/978-3-642-27737-5_642-1, 2016.

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

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

Marras, S. and Mandli, K. T.: Modeling and Simulation of Tsunami Impact: A Short Review of Recent Advances and Future Challenges, Geosciences, 11, 5, https://doi.org/10.3390/geosciences11010005, 2021.

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

Mori, N., Goda, K., and Cox, D.: Recent Process in Probabilistic Tsunami Hazard Analysis (PTHA) for Mega Thrust Subduction Earthquakes, in: The 2011 Japan Earthquake and Tsunami: Reconstruction and Restoration: Insights and Assessment after 5 Years, edited by: Santiago-Fandiño, V., Sato, S., Maki, N., and Iuchi, K., Advances in Natural and Technological Hazards Research, Springer International Publishing, Cham, ISBN 978-3-319-58691-5, 469–485, https://doi.org/10.1007/978-3-319-58691-5_27, 2018.

Mori, N., Satake, K., Cox, D., Goda, K., Catalan, P. A., Ho, T.-C., Imamura, F., Tomiczek, T., Lynett, P., Miyashita, T., Muhari, A., Titov, V., and Wilson, R.: Giant tsunami monitoring, early warning and hazard assessment, Nature Reviews Earth & Environment, 3, 557–572, https://doi.org/10.1038/s43017-022-00327-3, 2022.

Mulia, I. E., Gusman, A. R., and Satake, K.: Alternative to non-linear model for simulating tsunami inundation in real-time, Geophys. J. Int., 214, 2002–2013, https://doi.org/10.1093/gji/ggy238, 2018.

Mulia, I. E., Gusman, A. R., and Satake, K.: Applying a Deep Learning Algorithm to Tsunami Inundation Database of Megathrust Earthquakes, J. Geophys. Res.-Sol. Ea., 125, e2020JB019690, https://doi.org/10.1029/2020JB019690, 2020.

Mulia, I. E., Ueda, N., Miyoshi, T., Gusman, A. R., and Satake, K.: Machine learning-based tsunami inundation prediction derived from offshore observations, Nat. Commun., 13, 5489, https://doi.org/10.1038/s41467-022-33253-5, 2022.

NOWPHAS: Observed water level, astronomical tide level, and tidal level deviation data from the 2011 off the Pacific coast of Tohoku Earthquake Tsunami, NOWPHAS (Nationwide Ocean Wave information network for Ports and HArbourS) [data set], https://nowphas.mlit.go.jp/pastdata (last access: 5 May 2025), 2011.

Núñez, J., Catalán, P. A., Valle, C., Zamora, N., and Valderrama, A.: Discriminating the occurrence of inundation in tsunami early warning with one-dimensional convolutional neural networks, Sci. Rep., 12, 10321, https://doi.org/10.1038/s41598-022-13788-9, 2022.

Oishi, Y., Imamura, F., and Sugawara, D.: Near-field tsunami inundation forecast using the parallel TUNAMI-N2 model: Application to the 2011 Tohoku-Oki earthquake combined with source inversions, Geophys. Res. Lett., 42, 1083–1091, https://doi.org/10.1002/2014GL062577, 2015.

Okada, Y.: Surface deformation due to shear and tensile faults in a half-space, B. Seismol. Soc. Am., 75, 1135–1154, https://doi.org/10.1785/BSSA0750041135, 1985.

Okal, E. A., Kirby, S. H., and Kalligeris, N.: The Showa Sanriku earthquake of 1933 March 2: a global seismological reassessment, Geophys. J. Int., 206, 1492–1514, https://doi.org/10.1093/gji/ggw206, 2016.

Ragu Ramalingam, N.: Geoclaw inputs and processed outputs used in building tsunami ML surrogates for nearshore and onshore approximation, Zenodo [data set], https://doi.org/10.5281/zenodo.10817116, 2024.

Ragu Ramalingam, N.: naveenragur/tsunami-surrogates: nhess-2024-72, Zenodo [data set], https://doi.org/10.5281/zenodo.15337936, 2025.

Riko, N., Masayuki, K., and Yamanaka, Y.: Comparative study on the source processes of recurrent large earthquakes in Sanriku-oki region: the 1968 Tokachi-oki earthquake and the 1994

Nat. Hazards Earth Syst. Sci., 25, 1655–1679, 2025

https://doi.org/10.5194/nhess-25-1655-2025

Sanriku-oki earthquake, Zisin (J. Seismol. Soc. Jpn.), 54, 267–280, https://doi.org/10.4294/zisin1948.54.2_267, 2001.

Rim, D., Baraldi, R., Liu, C. M., LeVeque, R. J., and Terada, K.: Tsunami Early Warning From Global Navigation Satellite System Data Using Convolutional Neural Networks, Geophys. Res. Lett., 49, e2022GL099511, https://doi.org/10.1029/2022GL099511, 2022.

Röbke, B. R., Leijnse, T., Winter, G., van Ormondt, M., van Nieuwkoop, J., and de Graaff, R.: Rapid Assessment of Tsunami Offshore Propagation and Inundation with D-FLOW Flexible Mesh and SFINCS for the 2011 Tōhoku Tsunami in Japan, Journal of Marine Science and Engineering, 9, 453, https://doi.org/10.3390/jmse9050453, 2021.

Rodríguez, J. F., Macías, J., Castro, M. J., de la Asunción, M., and Sánchez-Linares, C.: Use of Neural Networks for Tsunami Maximum Height and Arrival Time Predictions, GeoHazards, 3, 323–344, https://doi.org/10.3390/geohazards3020017, 2022.

Salmanidou, D. M., Beck, J., Pazak, P., and Guillas, S.: Probabilistic, high-resolution tsunami predictions in northern Cascadia by exploiting sequential design for efficient emulation, Nat. Hazards Earth Syst. Sci., 21, 3789–3807, https://doi.org/10.5194/nhess-21-3789-2021, 2021.

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

Satake, K., Fujii, Y., and Yamaki, S.: Different depths of near-trench slips of the 1896 Sanriku and 2011 Tohoku earthquakes, Geoscience Letters, 4, 33, https://doi.org/10.1186/s40562-017-0099-y, 2017.

Seo, J.: Solving real-world optimization tasks using physics-informed neural computing, Sci. Rep., 14, 202, https://doi.org/10.1038/s41598-023-49977-3, 2024.

Shi, F., Kirby, J. T., Harris, J. C., Geiman, J. D., and Grilli, S. T.: A high-order adaptive time-stepping TVD solver for Boussinesq modeling of breaking waves and coastal inundation, Ocean Model., 43-44, 36–51, https://doi.org/10.1016/j.ocemod.2011.12.004, 2012.

Son, S., Lynett, P. J., and Kim, D.-H.: Nested and multi-physics modeling of tsunami evolution from generation to inundation, Ocean Model., 38, 96–113, https://doi.org/10.1016/j.ocemod.2011.02.007, 2011.

Strasser, F. O., Arango, M. C., and Bommer, J. J.: Scaling of the Source Dimensions of Interface and Intraslab Subduction-zone Earthquakes with Moment Magnitude, Seismol. Res. Lett., 81, 941–950, https://doi.org/10.1785/gssrl.81.6.941, 2010.

Tozato, K., Takase, S., Moriguchi, S., Terada, K., Otake, Y., Fukutani, Y., Nojima, K., Sakuraba, M., and Yokosu, H.: Rapid tsunami force prediction by mode-decomposition-based surrogate modeling, Nat. Hazards Earth Syst. Sci., 22, 1267–1285, https://doi.org/10.5194/nhess-22-1267-2022, 2022.

Wang, Y., Imai, K., Miyashita, T., Ariyoshi, K., Takahashi, N., and Satake, K.: Coastal tsunami prediction in Tohoku region, Japan, based on S-net observations using artificial neural network, Earth Planets Space, 75, 154, https://doi.org/10.1186/s40623-023-01912-6, 2023.

Williamson, A. L., Rim, D., Adams, L. M., LeVeque, R. J., Melgar, D., and González, F. I.: A Source Clustering Approach for Efficient Inundation Modeling and Regional Scale Probabilistic Tsunami Hazard Assessment, Front. Earth Sci., 8, 591663, https://doi.org/10.3389/feart.2020.591663, 2020.

Yamazaki, Y., Cheung, K. F., and Lay, T.: A Self-Consistent Fault Slip Model for the 2011 Tohoku Earthquake and Tsunami, J. Geophys. Res.-Sol. Ea., 123, 1435–1458, https://doi.org/10.1002/2017JB014749, 2018.