A HYBRID ML-PHYSICAL MODELLING APPROACH FOR EFFICIENT PROBABILISTIC TSUNAMI HAZARD AND RISK ASSESSMENT

Naveen Ragu Ramalingam[1], Anirudh Rao[2], Kendra Johnson[2], Marco Pagani[2,3] and Mario Martina[1]

ABSTRACT: Probabilistic tsunami hazard and risk assessment (PTHA/PTRA) are vital tools for understanding tsunami risk and planning measures to mitigate impacts. At large-scales their use and scope are currently limited by the computational costs of numerically intensive simulations which are not always feasible without large computational resources like HPCs and may require reductions in resolution, number of scenarios modelled or use of simpler approximation schemes. To conduct the PTHA/PTRA for large proportions of a coast, we need therefore to develop concepts and algorithms for reducing the number of events simulated and more rapidly approximating the needed simulation results. This case study for a coastal region of Tohoku, Japan utilizes a limited number of tsunami simulations from submarine earthquakes along the subduction interface to generate a wave propagation and inundation database at different depths and fits these simulation results to a machine learning (ML) based variational autoencoder model to predict the intensity measure (water depth, velocity, etc.) of the tsunami at the location of interest. Such a hybrid ML-physical model can be further extended to compute the inundation for probabilistic tsunami hazard and risk onshore.

Keywords: Tsunami hazard, machine learning, autoencoders

INTRODUCTION

The computational demands of probabilistic tsunami modelling currently limit their widespread application for onshore tsunami hazard and risk assessment. The ability of autoencoders in dimensional reduction, feature representation and transformation can be exploited for approximating the non-linear processes in the shallow water regions nearshore and the inundation processes onshore. In this contribution we describe how a variational autoencoder (VAE) type neural network model is trained on the simulated time series of tsunami amplitudes at points of different depths offshore and the inundation maps onshore for a location along the coastal Tohoku region in Japan. Further, the model is tested against the simulation of different realistic historical events to evaluate the efficacy and requirements of this novel approach.

GENERATION OF SYNTHETIC DATA

Tsunami Model and Test Locations

For the simulation of the synthetic events, their wave amplitude time series and onshore flood footprint, a tsunami model based on GeoClaw Version 5.7.1, (Clawpack Development Team, 2020) was developed. The model covers the Pacific Side of the Honshu Island with a nested domain approach of 1350 m-850 m-100 m grids as in Figure 1. Model development and calibration are based on local information related to topo-bathymetry, tsunami defences, 2011 Tohoku historical event's gauge and survey data.

The virtual gauges are setup at different depths of 5m, 25m, 50m and 100m as shown in Figure 2. Each simulation is run for a 6-hour duration from onset of the tsunami. Tides and waves components are ignored and the initial water level condition is zero mean sea level.

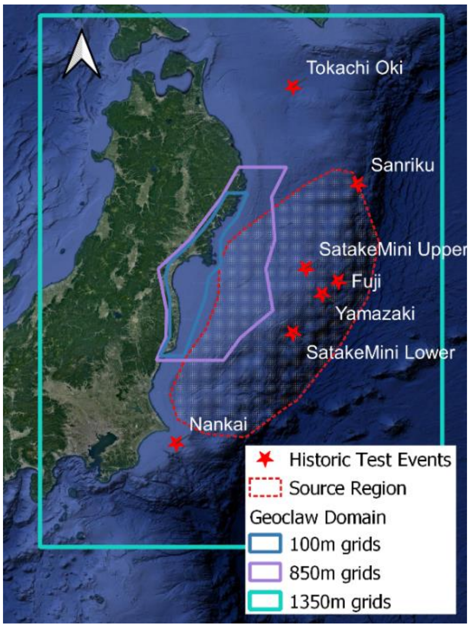

Fig. 1 Tsunami model coverage, location of source region and historical test events

[1] University School for Advanced Studies Pavia, Piazza della Vittoria 15, 27100 Pavia, ITALY (naveen.raguramalingam@iusspavia.it)
[2] GEM Foundation, Via Ferrata 1, 27100, Pavia, ITALY
[3] Institute of Catastrophe Risk Management, Nanyang Technological University, SINGAPORE

Three locations with different coastal configuration are identified for the case 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. Due to the limited scope of this article only the results related to Rikuzentakata, with offshore gauge 5832 and nearshore gauge 6042 are discussed.

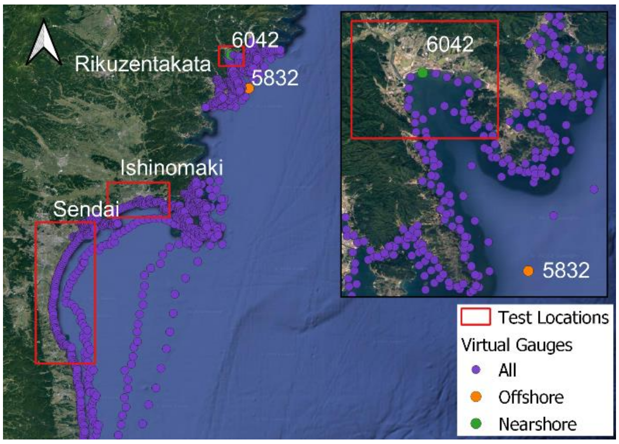

Fig. 2 Virtual gauges and the test locations in Japan (Rikuzentakata in the inset)

Design of Experiments

To train the VAE neural network model, 383 hypothetical tsunamigenic earthquake events are simulated, with M_w 7.5-9.0 and uniformly distributed over the Tohoku subduction interface earthquake source.

Information on depth, slip, strike and dip (see Table 1) are taken from the Slab2 model of the Japan trench (Hayes et al., 2018), and rake is always 90 degrees. The rupture planes are constrained to depths shallower than 16 km for M_w 9.0 events, assumed to be the lower seismogenic limit of the subduction interface. The deformation is modelled assuming homogenous slip for the rupture using Okada solution (Okada 1985) with the value of rupture length, width and slip scaled based on the magnitude of the event (Strasser et al. 2010).

Table 1 EQ event parameters

Range	Mw	Lat	Long	Depth	Dip	Strike
Min	7.5	35.73	141.15	10.2	5.54	187.20
Max	9	39.48	143.90	45.7	17	225.78

Test Events

To evaluate the performance of the VAE model on a generalized dataset, eight heterogenous slip and one homogenous slip events are used for testing:

1. 2011 Tohoku EQ – (Fujii et al. 2011), (Yamazaki et al. 2018), and four variations of (Satake et al. 2013) by varying the sub-fault location and slip distribution.
2. Events outside the Tohoku source region – 1968 Tokachi-Oki EQ (Nagai et al. 2001), 1896 Sanriku EQ (Satake et al. 2017) and a synthetic homogenous slip rupture event outside the training source region (Nankai 2022).

METHODS

Data Preprocessing

For a given gauge location and tsunami event, the wave amplitude time series is checked for a threshold (0.1 m at deep gauges, 0.5 m at shallow gauges). From the time instance when threshold is crossed, an observation window of 240 mins is selected for calculating a uniformly sampled wave amplitude time series with 1024 data points. For the inundation footprint, all flooded grid points from the overall simulation set for the region are selected and converted into a single 1D array. For each event, the 1D array element values represent the depth at these grid points. This wave propagation and inundation dataset in the learning framework is divided into 80:20 split of training and test events.

Variational Autoencoder Model (VAE)

This work extends an already existing ML approach (Liu et al. 2021) used for tsunami forecasting. Our objective is to approximate the tsunami wave amplitudes and event inundation footprints. The VAE model implementation from (Liu et al. 2021) was taken as the base model and further adapted and calibrated to this use case (refer to Figure 3 and Table 2). The current framework tries to limit the need for a large number of simulations and prescribes some additional steps to ease and improve the training of the model, like pretraining, curriculum learning, and moving mask learning which will be tested in future work.

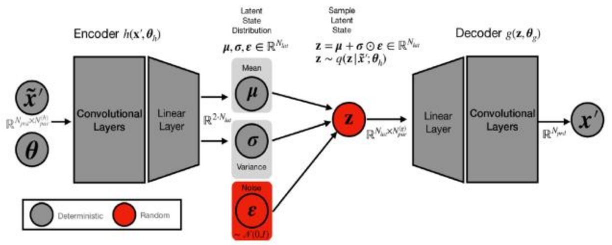

Fig. 3 Schematic of a Variational Autoencoder (Liu et al. 2021)

Table 2 CNN channel parameters of the VAE Model

	Layers	1	2	3	4	5
Encoder	In	Inputs	64	64	128	128
Network	Out	64	64	128	128	256
Decoder	In	256	128	128	64	64
Network	Out	128	128	64	64	Outputs

The main changes to the learning framework and design of VAE by Liu et al 2021 are:

1. Observation window - 240 mins.
2. Number of CNN layers - 5.
3. Number of z-latent variables - 50.

RESULTS AND DISCUSSION

For the test events, maximum amplitude and the timing of local peaks is well captured by the model, as shown in Figure 5 for the test event (#123) and Tohoku event(Yamazaki et al. 2018) at the test nearshore gauge (6042). Further, the model is able to extrapolate beyond the training information, predicting water levels higher than those in the training events, as shown in the results for the full Tohoku test event set in Figure 4.

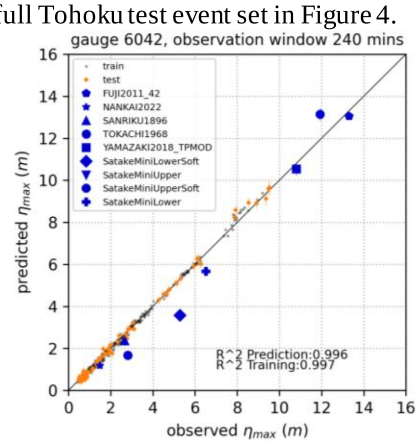

Fig. 4 Scatter plot of the observed vs predicated tsunami max amplitude at the nearshore gauge (6042)

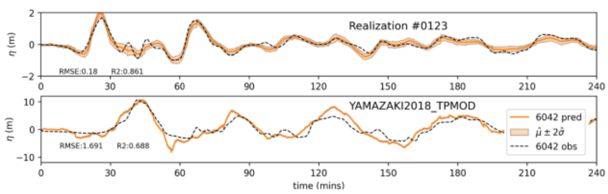

Fig. 5 Time series realization of a test event (#123) and historic Tohoku event (Yamazaki et al. 2018)

Figure 6 shows results for inundation prediction at Rikuzentakata for a low and high inundation test event. There is a minor depth error of 0.1-0.2 m reported by the model at grid cells not inundated, locations seen in the absolute error maps.

While the nearshore approximation approach is useful as hazard proxy at the coast, its application is currently limited by the need to train the model for each location and calibrate the VAE model architecture, hindering the application at a large scale, but the possibility to predict the inundation maps provides a more transferrable scheme for a larger region.

Such an approach can help extend offshore tsunami hazard information available at deep offshore points to much needed onshore hazard and risk, with a relatively limited number of simulations and associated computational costs.

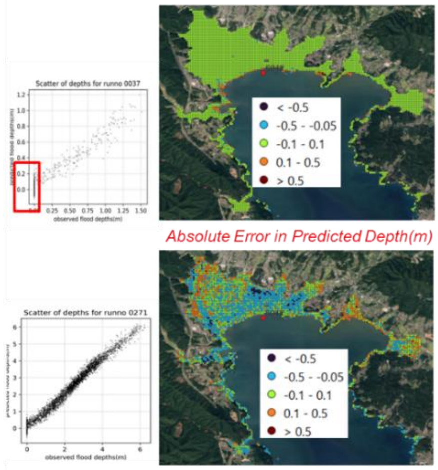

Fig. 6 Scatter plot of the observed vs predicated tsunami inundation depths of a low (#37) and high (#271) magnitude event and the absolute error map for the floodable grid locations.

REFERENCES

Clawpack Development Team (2020),Clawpack v5.7.1, http://www.clawpack.org,doi:10.5281/zenodo.4025432

Fujii, Y., K. Satake, S. Sakai,M. Shinohara, and T. Kanazawa (2011) Tsunami source of the 2011 off the Pacific coast of Tohoku EQ, Earth Planets Space, 63, 815–820.

Liu, C. M., Rim, D., Baraldi, R., & LeVeque, R. J. (2021). Comparison of machine learning approaches for tsunami forecasting from sparse observations. Pure and Applied Geophysics, 178(12), 5129-5153.

Nagai, R., M. Kikuchi, and Y. Yamanaka. 2001. Comparative Study on the Source Processes of Reccurent Large Earthquakes in Sariku-oki Region: 1968 Tokachi-oki Earthquake and 1994 Sanriku-oki Earthquake. (Journal of the Seism. Society of Japan) 54:267-280.

Okada, Y. (1985). Surface deformation due to shear and tensile faults in a half-space. Bulletin of the seismological society of America, 75(4), 1135-1154.

Satake, K., Fujii, Y., & Yamaki, S. (2017). Different depths of near-trench slips of the 1896 Sanriku and 2011 Tohoku Earthquakes. Geoscience Letters, 4(1), 1-10.

Satake, K., Fujii, Y., Harada, T. and Namegaya Y. (2013),Time and Space Distribution of Coseismic Slip of the 2011 Tohoku EQ as Inferred from Tsunami Waveform Data, Bulletin of the seismological society of America, 103(2B) 1473–1492.

Strasser, F. O., Arango, M. C., & Bommer, J. J. (2010). Scaling of the source dimensions of interface and intraslab subduction-zone earthquakes with moment magnitude. Seismol Research Letters, 81(6), 941-950.

Hayes, G., 2018, Slab2 - A Comprehensive Subduction Zone Geometry Model: U.S. Geological Survey data release.

Yamazaki, Y., Cheung, K. F., & Lay, T. (2018). A self-consistent fault slip model for the 2011 Tohoku earthquake and tsunami. Journal of Geophysical Research: Solid Earth, 123(2), 1435-1458.