

Comprehensive Prediction of Subaerial Landslide-Tsunamis via Slide Model- and Water Body Geometry-Invariant Machine Learning Techniques

David Gwilym Jenkins^{a,*}, Valentin Heller^b, Archontis Giannakidis^{a,c}

^a*Department of Physics and Mathematics, School of Science and Technology,
Nottingham Trent University, Nottingham, NG11 8NS, UK*

^b*Environmental Fluid Mechanics and Geoprocesses Research Group, Faculty of
Engineering, University of Nottingham, Nottingham, NG7 2RD, UK*

^c*Archimedes Research Unit in Artificial Intelligence, Data Science and Algorithms,
Marousi, 15125, GREECE*

Abstract

Subaerial landslide-tsunamis (SLTs) are generated by mass movements impacting water bodies. These phenomena are responsible for substantial loss of life and damage to properties making their prediction vital. There is limited literature on deriving cross-slide model and water geometry prediction equations. Herein, state-of-the-art machine learning methods were utilised for building slide model- and water body geometry-invariant models, using data sets of four laboratory studies involving a wide range of experimental conditions. Models on individual data sets are also built. Other novel contributions include the development of: (i) ensemble models for predicting maximum SLT characteristics, (ii) sequence models for predicting wave decay, and (iii) a neural network for categorising wave types. All proposed models trained on individual data sets outperformed or matched performance of existing empirical equations. The slide model- and water body geometry-invariant models held performance across different slide models and water body geometries. The developed models are successfully validated by predicting the maximum height of the 1958 Lituya Bay tsunami with 2.47% error and the maximum wave run-up of the 2007 Chehalis Lake case with 9.52% error. This study, based on a range of idealised experimental conditions and

*Corresponding author

Email address: david.jenkins@ntu.ac.uk (David Gwilym Jenkins)

several studies, is expected to shape preliminary hazard assessment of SLTs.

Keywords: Gradient boosting, Landslide-tsunamis, Machine learning, Neural networks, Sequence modelling, Wave propagation

1. Introduction

1.1. Overview

When mass movements such as rockfalls, mud slides, debris flows, or glacier calvings impact a body of water, a subaerial landslide-tsunami (SLT) is generated. These events pose a prevalent threat in coastal and mountainous areas. Notable cases in recent history include the 1958 Lituya Bay SLT with an observed wave run-up height of 524 m (Fritz et al., 2001), the 2014 Lake Askja rockslide on Iceland with a maximum run-up height of 71 m (Gylfadóttir et al., 2017), the 2014 Greenland SLT at the Eqip Sermia outlet glacier front with a maximum amplitude of 50 m (Lüthi and Vieli, 2016), and the 2018 Anak Krakatau volcano tsunami SLT in Indonesia which resulted in a maximum run-up height of 13.5 m in Java, around 50km from the slide impact zone, and led to 437 fatalities (Muhari et al., 2019; Heidarzadeh et al., 2020). Along with destruction of the built environment, recent SLTs and associated phenomena have been responsible for over 58,000 fatalities (Heller and Ruffini, 2023). Reliable prediction of these events is therefore vital.

The characteristics of SLTs depend on a wide range of features of the landslide as well as the geometry and bathymetry of the water body. Most of these features can be reproduced in prototype-specific laboratory experiments (e.g. WCHL, 1970; Davidson and Whalin, 1974; Fuchs and Boes, 2010; Fuchs et al., 2011; Lindstrøm et al., 2014; Wang et al., 2016) and numerical simulations (e.g. Løvholt et al., 2008; Abadie et al., 2012; Gabl et al., 2015; Gylfadóttir et al., 2017; Tan et al., 2018; Rauter et al., 2022; Sabeti and Heidarzadeh, 2022) resulting in satisfactory predictions. However, such approaches are resource- and time-consuming. For instance, the Kühtai reservoir laboratory study of Fuchs et al. (2011) took approximately 1 year to complete, and in Chen et al. (2020) some simulations of tsunamis generated by iceberg calving impacts took up to 96 hours.

Obtaining generic empirical equation-based preliminary SLT estimates from idealised laboratory experiments and numerical simulations is an efficient alternative approach to assist in preliminary hazard assessment (Evers et al., 2019). According to this approach, approximately 20 parameters

(slide velocity, volume, geometry, hill slope, water depth, etc.) are systematically varied under idealised conditions (e.g. granular or block slide; two-dimensional 2D flume or three-dimensional 3D basin water body geometry), and the SLT characteristics are then expressed through empirical equations as functions of these parameters (Huber and Hager, 1997; Heller and Hager, 2010; Heller and Spinneken, 2015; Zitti et al., 2016; Evers et al., 2019). Heller and Ruffini (2023) presented a review of 76 studies applying this method.

It is nevertheless unrealistic to investigate all these parameter combinations within a single study, as a systematic variation of these parameter values may require approximately 17.54 million experiments (Heller and Ruffini, 2023). As a result, most studies varied four to five (or even fewer) parameters generating empirical equations with a limited application range. Further, empirical equations based on data from experiments conducted under comparable conditions can lead to one order of magnitude (or even larger) differences in the SLT characteristic predictions (Heller and Hager, 2010; Watt et al., 2012; Evers and Hager, 2016; Bullard et al., 2019; Heller and Ruffini, 2023). As an indication, applying empirical equations from eight different 3D studies in order to predict the maximum wave amplitude of the 2007 Chehalis case (Roberts et al., 2013) yielded a prediction range between 5.9 and 256.2 m (Heller and Ruffini, 2023). The present study relies on machine learning (ML) as a method to consistently represent a combined data set over several studies that cover wide parameter ranges and experimental conditions.

1.2. Literature Review

In contrast to water waves in general (e.g. Alizadeh and Nourani, 2024; Minuzzi and Farina, 2024), ML algorithms have been employed in SLT prediction only a handful of times. Panizzo et al. (2005) pioneered this research by using the data from 288 3D laboratory experimental runs. They deployed a number of shallow artificial neural networks (ANNs), consisting of one hidden layer with five hidden units. These models gave good predictions and compared well with predictions from empirical equations derived on the same data. Meng et al. (2020) also applied ANNs, using two alternatives for the model input, namely the landslide parameters upon impact with the water body, and the landslide parameters in the initial position. The former involved 291 2D experiments on which the ML models performed stronger than the derived empirical equations. Ruffini et al. (2021) investigated how three different bathymetric types affect tsunami propagation. They deployed

1 deeper ANN models with two hidden layers and 15 hidden units to predict
 2 both the ratio of the transformed wave height and amplitude after each bathy-
 3 metric feature, outperforming empirical equation analysis. Wu et al. (2022)
 4 used physics-informed neural network models (Raissi et al. 2019), a branch
 5 of ANNs that incorporate physics knowledge, in this case, conservation of
 6 mass and momentum. To simulate the propagation process of SLTs, the 2D
 7 spatial coordinates of the fluid particles were used as inputs to the models,
 8 whereas the updated position and velocities of the particles served as the
 9 outputs. The simulations were found to have a good agreement with the
 10 physical experimental results. Finally, Feng et al. (2022) proposed a random
 11 forest-based method to describe the SLT profiles generated by block slides
 12 in 2D.

13 All studies described in this section display the great potential success of
 14 implementing ML in this field, since the ML techniques consistently outper-
 15 formed, or were on par with, empirical equations based on traditional corre-
 16 lation methods. However, these studies are limited to making predictions on
 17 individual data sets characterised by relatively narrow parameter ranges such
 18 as block slides in 3D (Panizzo et al., 2005), block slides in 2D (Feng et al.,
 19 2022; Wu et al., 2022) or granular/viscous slides in 2D (Meng et al., 2020).
 20 To date, only a small number of studies (e.g. Tang et al., 2018; Sabeti and
 21 Heidarzadeh, 2022) introduced empirical equations for predicting SLT maxi-
 22 mum wave characteristics for both block and granular materials. In addition,
 23 previous ML studies have mainly deployed ANNs. However, work undertaken
 24 by Schwartz-Ziv and Armon (2022) indicates gradient boosting (GB), based
 25 on the idea that the combination of multiple weaker predictive models can
 26 outperform any individual model (Friedman, 2001), produces more accurate
 27 predictions for tabular data. GB has not yet been applied to SLTs. Fur-
 28 thermore, nearly all ML studies concentrated on predicting maximum wave
 29 characteristics. None of them predicted wave evolution with distance from
 30 the impact zone. Another weakness of existing empirical equations (Heller
 31 and Hager, 2011; Xue et al., 2019) is that they are not able to distinguish
 32 between cnoidal- and solitary-like wave types for a given SLT. Lastly, ML
 33 models that make predictions across different water body geometries (2D,
 34 3D) and slide models (granular, block slides) have not yet been explored.

35 *1.3. Aims and Objectives*

36 The aim of this article is to introduce universal models proposing ML
 37 techniques that make SLT predictions on a wide range of idealised exper-

1 imental conditions. This aim will be achieved by including data from the
2 four SLT laboratory studies Heller and Hager (2010) (granular slides in 2D),
3 Heller and Spinneken (2013) (block slides in 2D), Huber (1980) (granular
4 slides in 3D), and Heller and Spinneken (2015) (block slides in 2D and 3D).
5 This article involves the following objectives:

- 6 • Apply GB to predict maximum SLT values.
- 7 • Build sequence models using both traditional recurrent neural network
8 (RNN) and cutting edge transformer-based (Vaswani et al., 2017) ap-
9 proaches to predict wave characteristic evolution.
- 10 • Present an ANN-based method to predict the four SLT wave types,
11 namely Stokes-like, cnoidal-like, solitary-like, and bore-like.
- 12 • Validate the predictive power of the proposed ML predictions on an
13 idealised 2D study of the 1958 Lituya Bay case and on the 2007 Chehalis
14 Lake case.

15 The remainder of this article is structured as follows: Section 2 describes the
16 relevant input features and prediction variables, as well as the ML models and
17 techniques used and the evaluation protocol. Section 3 presents, quantifies,
18 and visualises the performance of the implemented models. Section 4 gives
19 a detailed discussion of these results and finally, Section 5 outlines the main
20 conclusions.

2. Methodology

2.1. Data Set Acquisition and Description

Data from Heller and Hager (2010) (granular slides in 2D), Heller and Spinneken (2013) (block slides in 2D), Huber (1980) (granular slides in 3D), and Heller and Spinneken (2015) (block slides in 2D and 3D) was used to produce the ML models in this work. While a limited number of additional studies may suit this work, these four studies were chosen as they involve a wide range of idealised experimental conditions, investigated the desired parameters for this work, and the data was readily available.

The experiments of Heller and Hager (2010) (also included in Heller, 2007 and Heller and Hager, 2011) were conducted in a flume of length 11 m, width $b = 0.500$ m, and height 1 m. With the help of a pneumatic landslide generator, they systematically varied the slide thickness s , bulk slide volume V_s , bulk slide density ρ_s , grain diameter d_g , slide mass m_s , slide impact velocity V_s , slide impact angle α , and the still water depth h (Fig. 1a) over a total number $m_{To} = 211$ of experiments. The $m_{To} = 144$ block experiments of Heller and Spinneken (2013) were carried out in a 24.5 m long, $b = 0.600$ m wide, and 1 m high flume. The transition type at the ramp toe consisted of either no transition or a circular-shaped transition of radius 0.400 m. The varied parameters were h , V_s , the slide width b_s , and the slide front angle ϕ . Other parameters used were s , ρ_s , m_s , V_s , and α (Fig. 1b). The free water surfaces in the above two 2D studies was measured as a function of the streamwise distance x using seven wave gauges along the flume (Fig. 1c). Along with 2D experimental runs (not considered), Huber (1980) also conducted SLT experiments in a 3D water basin (6.0 m wide and 10.0 m long, with a selection $m_{To} = 18$ considered herein) using granular slides. The varied parameters were h , V_s , m_s , V_s . s was not directly available from Huber (1980). Consequently, it was estimated by computing the slide release position via Eq. (3.5) of Evers et al. (2019) and V_s . The slide thickness, available from the triangular-shaped slide in the initial position, was then reduced to s at the impact location according to the slide profile evolution shown on photographs in Huber (1980) and Savage and Hutter (1989). Heller and Spinneken (2015) investigated the characteristics of SLTs both in 2D (21 m long, 0.600 m wide, $m_{To} = 30$) and 3D (unobstructed width of 7.4 m, 20 m long, $m_{To} = 18$). The flume experiments were conducted under similar conditions to those in Heller and Spinneken (2013) but they varied ρ_s , m_s , and V_s and maintained b_s , ϕ , and α constant. For the 3D experiments,

1 measurements were taken using wave gauges at radial distance r , and wave
 2 propagation angle γ (Fig. 1d). A breakdown of the included data sets can
 3 be seen in Table 1.

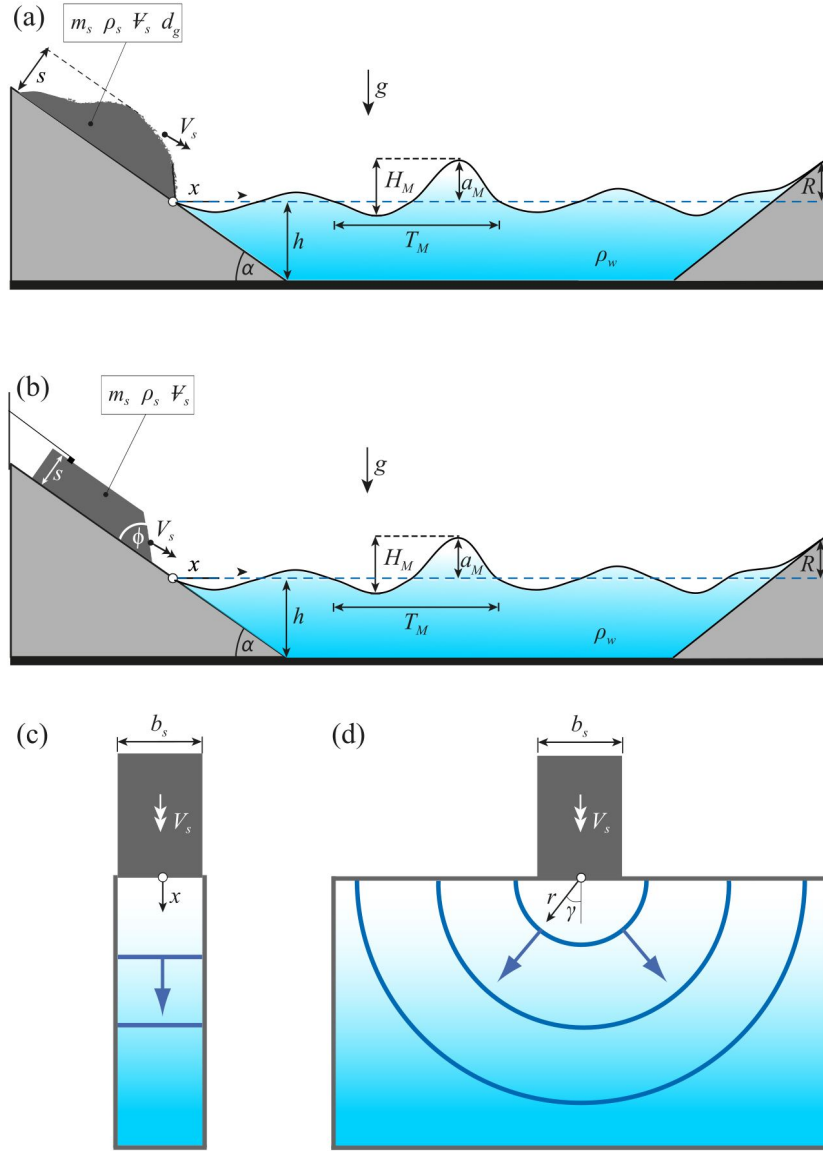


Figure 1: Definition sketches of the side views of (a) granular and (b) block slides and SLTs, and of the plane views of (c) 2D and (d) 3D water body geometries.

Table 1: A summary of the data sets used in this study.

Study	Landslide Model	Water Body Geometry	Number of Experiments $\mathbf{m_{To}}$
Heller and Hager (2010)	Granular	2D	211
Heller and Spinneken (2013)	Block	2D	144
Huber (1980)	Granular	3D	18
Heller and Spinneken (2015)	Block	2D, 3D	30 (2D), 18 (3D)

The parameter ranges in the four introduced studies are summarised in Table 2. These parameters can be transformed to the following dimensionless ones (Heller and Spinneken, 2013; Heller and Ruffini, 2023):

- Slide Froude number $F = \frac{V_s}{(gh)^{\frac{1}{2}}}$;
- Relative slide thickness $S = \frac{s}{h}$;
- Relative slide mass $M = \frac{m_s}{(\rho_w b_s h^2)}$;
- Relative grain diameter $D_g = \frac{d_g}{h}$;
- Slide impact angle α ;
- Blockage ratio $B = \frac{b_s}{b}$;
- Expression considering the slide front angle $\Phi = \sin^{\frac{1}{2}}\phi$;
- Expression considering the transition type $T_s = t_s/((h + V_s/(sb_s))/V_s)$;
- Relative distance $X = \frac{x}{h}$ in 2D;
- Relative radial distance r/h in 3D;
- Wave propagation angle γ in 3D.

Here, g is the gravitational acceleration, t_s the characteristic time of submerged landslide motion, and ρ_w is the water density.

Table 2: Experimental parameter ranges of Heller and Hager (2010), Heller and Spinneken (2013; 2015), Huber (1980), and the four studies combined. The given ranges for Huber (1980) are for the selected 3D data only, and $h = 0.120$ m was not considered to exclude tests with significant scale effects (Heller et al., 2008).

Symbol	Unit	Heller and Hager (2010)	Heller and Spinneken (2013)	Huber (1980)	Heller and Spinneken (2015)	Combined
h	(m)	0.150, 0.200, 0.300, 0.450, 0.600	0.300, 0.600	0.180, 0.360	0.240, 0.480	0.150 – 0.600
s	(m)	0.053 – 0.249	0.120	0.0321 – 0.160	0.120	0.053 – 0.249
d_g	(mm)	2, 4, 5, 8, and mixtures	—	8 – 30	—	2 – 30
V_s	(m/s)	2.060 – 8.340	0.590 – 3.560	2.000	0.940 – 3.790	0.590 – 8.340
V_s	(m ³)	0.017, 0.033, 0.067	0.037	0.0079 – 0.026	0.020, 0.038, 0.057	0.0079 – 0.067
ρ_s	(kg/m ³)	≈ 590 , ≈ 1720	≈ 1534	1890	1451, 1586, 1597	590 – 1890
m_s	(kg)	19.690 – 113.300	57.230	15.000, 30.000, 50.000	32.510, 60.140, 82.670	19.690 – 113.300
α	(°)	30, 45, 60, 90	45	30, 45, 60	45	30 – 90
b_s	(m)	0.500	0.530 – 0.590	0.500	0.580	0.500 – 0.590
ϕ	(°)	—	30, 45, 60, 90	—	45	30 – 90
x	(m)	0 – 8.90	0 – 14.00	—	0 – 16.80	0 – 16.80
r	(m)	—	—	5 – 20	0 – 16.80	0 – 16.80
Transition type	—	—	None with stopper, circular	—	Circular	None with stopper, circular
F	—	0.86 – 6.83	0.34 – 2.07	1.06 – 1.51	0.54 – 2.47	0.34 – 6.83
S	—	0.09 – 1.64	0.20 – 0.40	0.09 – 0.89	0.25, 0.50	0.09 – 1.64
M	—	0.11 – 10.02	0.27 – 1.21	0.23– 3.10	0.25 – 2.49	0.11 – 10.02
D_g	—	0.003 – 0.040	—	0.020 – 0.160	—	0.003 – 0.040
T_s	—	—	0.34 – 1.00	—	—	0.34 – 1.00
B	—	—	0.88 – 0.98	—	—	0.88 – 0.98
X	—	0 – 59	0 – 40	—	0 – 35	0 – 59
r/h	—	—	—	5 – 20	0 – 35	0 – 35
mTo	—	211	144	18	48	421

1 2.2. Model Prediction Variables

2 Important SLT characteristics are the wave amplitude a , height H , and
3 period T . a is defined as the distance from the still water surface to the wave
4 crest. H represents the sum of a and the absolute value of the corresponding
5 wave trough. Finally, T marks the time between the first and second up-
6 crossing of the water surface elevation. Both the maximum (subscript M)
7 a_M , H_M , and T_M and the evolution of the characteristic value over distance
8 ($a(x)$, $a(r, \gamma)$, $H(x)$, $H(r, \gamma)$, and $T(x)$, $T(r, \gamma)$) are relevant. T_M is defined as
9 the period of the crest in which a_M is measured (Fig. 1a). When predicting
10 these values, they are typically made dimensionless with h and g resulting in
11 the relative maximum wave amplitude a_M/h , height H_M/h , and the period
12 $T_M(g/h)^{1/2}$.

13 2.3. Case Studies

14 2.3.1. Lituya Bay Case

15 The slide model- and water body geometry-invariant ML model for max-
16 imum height prediction and the ML model for predicting the 2D granular
17 SLT type proposed in this study, are validated on the Fritz et al. (2001)
18 idealised 2D study of the 1958 Lituya Bay case. In July 1958, a magnitude
19 7.8 to 8.3 earthquake occurred along the Queen Charlotte Fault in Alaska.
20 This event caused a rockslide to impact the Lituya Bay which generated an
21 SLT (Fritz et al., 2001). The rock landslide had a bulk slide density of $\rho_s =$
22 1620 kg/m^3 , mean slide width $b_s = 823 \text{ m}$, maximum slide thickness of $s =$
23 92 m , and a slide bulk volume $V_s = 51.0 \times 10^6 \text{ m}^3$. The mean water depth
24 was $h = 122 \text{ m}$. Slingerland and Voight (1979) calculated the slide impact
25 velocity as $V_s = 92 \text{ m/s}$. The relevant dimensionless parameters are therefore
26 given as $F = 2.66$, $S = 0.75$, and $M = 6.74$. Based on the observed run-up
27 height of 524 m , Fritz et al. (2001) inversely computed the maximum wave
28 height (Fig. 1a) as $H_M = 162 \text{ m}$ with the empirical equation of Hall and
29 Watts (1953).

30 2.3.2. Chehalis Lake Case

31 The slide model- and water body geometry-invariant ML model for wave
32 height decay, proposed in this study, is validated on the 2007 Chehalis Lake
33 case (Roberts et al., 2013). Located in a SLT prone area of British Colombia,
34 Canada, Chehalis Lake is particularly vulnerable to this kind of event. With
35 the exception of the 1980 volcanic eruption of Mount St. Helens and the

1 following rock avalanche (Voight et al., 1981), the vast majority of tsunami-
2 causing landslides in the area have no clear triggers. In December 2007, one
3 such landslide with a volume $V_s = 3 \text{ Mm}^3$ and density $\rho_s = 1650 \text{ kg/m}^3$
4 impacted Chehalis Lake with a velocity $V_s = 60 \text{ m/s}$. The mean slide width
5 was $b_s = 210 \text{ m}$, the maximum slide thickness $s = 40 \text{ m}$, and the still water
6 depth was $h = 120 \text{ m}$. The relevant dimensionless parameters for this event
7 are therefore given as $F = 1.75$, $S = 0.33$, and $M = 1.64$. The maximum
8 run-up height R , which depends on H at a given distance from the slide
9 impact, is an important parameter for prediction. In this case, this value
10 was measured 800 m away from the landslide impact at $\gamma = 0$ as $R = 37.8$
11 m in a field survey (Roberts et al., 2013).

12 2.4. Machine Learning Algorithms

13 2.4.1. Fully Connected ANNs

14 Fully connected (FC) ANNs (Rosenblatt, 1958) are employed to classify
15 the four SLT wave types. FC ANNs comprise of individual neurons that are
16 connected to form layers. Each neuron takes in information, processes it, and
17 provides an output. This processing is dictated by the parameter tensor $\mathbf{W}$,
18 and the values of $\mathbf{W}$ are adapted when the model is learning. In a FC ANN,
19 the layer l has a weights matrix $\mathbf{W}^{(l-1)}$, which controls the values in that
20 layer. The q^{th} row, $\mathbf{W}_q^{(l-1)}$, of that matrix contains weights for each neuron
21 in layer l . An example of an ANN with an input layer, three hidden layers,
22 and the output layer can be seen in Fig. 2. The output value for that neuron
23 is calculated using the $\mathbf{W}_q^{(l-1)}$ weights and the output values of all neurons in
24 the previous layer by using the formula:

$$a_q^{(l)} = \sigma \left(\sum_{i=1}^{u_{l-1}} \mathbf{W}_{qi}^{(l-1)} a_i^{(l-1)} + \mathbf{b}_q \right), \quad (1)$$

25 where u_{l-1} is the number of units in the previous layer, σ denotes the ac-
26 tivation function, and $\mathbf{b}_q$ is a bias matrix. The discrepancy between the
27 predicted and the observed values is calculated for all output nodes of the
28 model during the training process of a supervised problem using a loss func-
29 tion. These error values are then propagated backwards, such that all weight
30 values are updated using an iterative gradient descent variant. The softmax
31 loss function was deployed in this study.

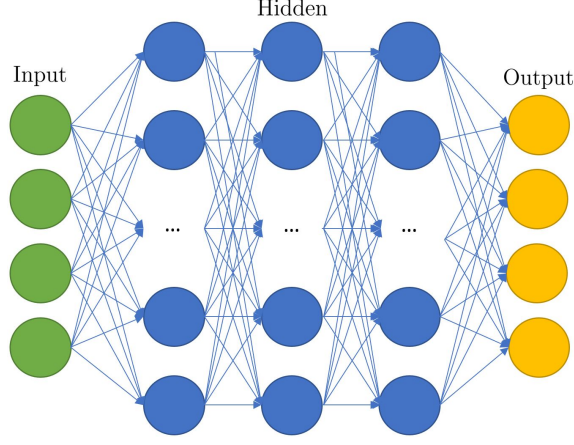


Figure 2: An example of the architecture of a fully connected ANN with three hidden layers. The input and output layers both have four neurons.

2.4.2. Sequence Models

Long short-term memory (LSTM) and gated recurrent unit (GRU) sequence models are developed to predict wave evolution characteristics. Unlike standard RNNs, such models are able to retain information from very early steps (frames) in the sequence when making predictions.

LSTM units (Hochreiter and Schmidhuber, 1997; Sak et al., 2014; Zaremba et al., 2015; Bohoran et al., 2023c) have two distinct methods of storing information from previous frames: A short-term state $\mathbf{h}_{(t)}$ and a long-term state $\mathbf{c}_{(t)}$ where subscript (t) denotes the frame. $\mathbf{c}_{(t)}$ carries desired information further along the model and is tuned independently via forget gates $\mathbf{f}_{(t)}$, and input gates, $\mathbf{i}_{(t)}$. The output gate $\mathbf{o}_{(t)}$ dictates which information should be passed to the output of the unit.

Each LSTM cell's long-term state, $\mathbf{c}_{(t)}$, and the output, $y_{(t)}$, are computed by

$$\begin{aligned}
 \mathbf{i}_{(t)} &= \sigma(\mathbf{W}_{xi}^T \mathbf{x}_{(t)} + \mathbf{W}_{hi}^T \mathbf{h}_{(t-1)} + \mathbf{b}_i), \\
 \mathbf{f}_{(t)} &= \sigma(\mathbf{W}_{xf}^T \mathbf{x}_{(t)} + \mathbf{W}_{hf}^T \mathbf{h}_{(t-1)} + \mathbf{b}_f), \\
 \mathbf{o}_{(t)} &= \sigma(\mathbf{W}_{xo}^T \mathbf{x}_{(t)} + \mathbf{W}_{ho}^T \mathbf{h}_{(t-1)} + \mathbf{b}_o), \\
 \mathbf{g}_{(t)} &= \tanh(\mathbf{W}_{xg}^T \mathbf{x}_{(t)} + \mathbf{W}_{hg}^T \mathbf{h}_{(t-1)} + \mathbf{b}_g), \\
 \mathbf{c}_{(t)} &= (\mathbf{f}_{(t)} \otimes \mathbf{c}_{(t-1)}) + (\mathbf{i}_{(t)} \otimes \mathbf{g}_{(t)}), \\
 y_{(t)} &= \mathbf{h}_{(t)} = \mathbf{o}_{(t)} \otimes \tanh(\mathbf{c}_{(t)}),
 \end{aligned} \tag{2}$$

1 where $\mathbf{x}$ is the input sequence whereas $\mathbf{W}_{\mathbf{x}\mathbf{i}}$, $\mathbf{W}_{\mathbf{x}\mathbf{f}}$, $\mathbf{W}_{\mathbf{x}\mathbf{o}}$, $\mathbf{W}_{\mathbf{x}\mathbf{g}}$, $\mathbf{W}_{\mathbf{h}\mathbf{i}}$, $\mathbf{W}_{\mathbf{h}\mathbf{f}}$,
2 $\mathbf{W}_{\mathbf{h}\mathbf{o}}$, and $\mathbf{W}_{\mathbf{h}\mathbf{g}}$ are the weights of the FC ANNs of $\mathbf{x}$ and $\mathbf{h}_{(t-1)}$, respec-
3 tively. $\mathbf{b}_{\mathbf{i}}$, $\mathbf{b}_{\mathbf{f}}$, $\mathbf{b}_{\mathbf{o}}$, and $\mathbf{b}_{\mathbf{g}}$ are the corresponding bias terms which are added
4 to increase performance on unseen examples, and $\otimes$ denotes element-wise
5 multiplication.

6 GRU units (Cho et al., 2014), on the other hand, are a simplified version of
7 LSTM units as $\mathbf{c}_{(t)}$ is dropped, allowing $\mathbf{h}_{(t)}$ to carry forward all information
8 from previous frames. A new gate $\mathbf{z}_{(t)}$ controls both the forget and input
9 gates. $\mathbf{o}_{(t)}$ is also dropped in favour of a new gate $\mathbf{r}_{(t)}$ which performs a
10 similar function but is better suited to the new unit set-up. The equations
11 are

$$\begin{aligned} \mathbf{z}_{(t)} &= \sigma(\mathbf{W}_{\mathbf{x}\mathbf{z}}^T \mathbf{x}_{(t)} + \mathbf{W}_{\mathbf{h}\mathbf{z}}^T \mathbf{h}_{(t-1)} + \mathbf{b}_{\mathbf{z}}), \\ \mathbf{r}_{(t)} &= \sigma(\mathbf{W}_{\mathbf{x}\mathbf{r}}^T \mathbf{x}_{(t)} + \mathbf{W}_{\mathbf{h}\mathbf{r}}^T \mathbf{h}_{(t-1)} + \mathbf{b}_{\mathbf{r}}), \\ \mathbf{g}_{(t)} &= \tanh(\mathbf{W}_{\mathbf{x}\mathbf{g}}^T \mathbf{x}_{(t)} + \mathbf{W}_{\mathbf{h}\mathbf{g}}^T (\mathbf{r}_{(t)} \otimes \mathbf{h}_{(t-1)}) + \mathbf{b}_{\mathbf{g}}), \\ y_{(t)} &= \mathbf{h}_{(t)} = (\mathbf{z}_{(t)} \otimes \mathbf{h}_{(t-1)}) + ((1 - \mathbf{z}_{(t)}) \otimes \mathbf{g}_{(t)}), \end{aligned} \quad (3)$$

12 where $\mathbf{W}_{\mathbf{x}\mathbf{z}}$, $\mathbf{W}_{\mathbf{x}\mathbf{r}}$, $\mathbf{W}_{\mathbf{x}\mathbf{g}}$, $\mathbf{W}_{\mathbf{h}\mathbf{z}}$, $\mathbf{W}_{\mathbf{h}\mathbf{r}}$, and $\mathbf{W}_{\mathbf{h}\mathbf{g}}$ are the weights for the FC
13 ANNs with $\mathbf{x}_{(t)}$ and $\mathbf{h}_{(t-1)}$ as inputs, respectively. Finally, $\mathbf{b}_{\mathbf{z}}$, $\mathbf{b}_{\mathbf{r}}$, and $\mathbf{b}_{\mathbf{g}}$ are
14 the corresponding bias term vectors. The $\mathbf{W}$ and $\mathbf{b}$ tensors are tuned during
15 the training process, which aims to maximise the conditional probability of
16 a target sequence given a source sequence.

17 In this study, $\mathbf{x}_{(t)}$ contains information about the SLT governing param-
18 eters and the distance from the landslide impact zone at frame t , whereas
19 the output $y_{(t)}$ contains information about the wave characteristic to be pre-
20 dicted.

21 2.4.3. Multi-head Self-attention Models

22 Transformer-based models (Vaswani et al., 2017) are developed to fore-
23 cast characteristics of wave evolution in cross-study settings. The main ad-
24 vantage of these models over LSTM or GRU models is the introduction of
25 the multi-head self-attention mechanism, a system that allows the model to
26 automatically detect and focus on the most relevant parts of an input (Lyst-
27 baek et al., 2023). This results in the ability to process the entirety of an
28 input data set in parallel, removing the need to make calculations sequen-
29 tially and thus store hidden weights in memory. To achieve this, queries Q ,
30 keys K , and values V are assigned to the input of the model which are used
31 to establish importance of each position in that input. These variables are

1 obtained by applying FC ANN layers, which are optimised during training.
 2 After these values are determined, a scaled dot-product attention mechanism
 3 is calculated by:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_K}}\right)V, \quad (4)$$

4 where d_K denotes the dimension of K . Huang et al. (2020) introduced
 5 the TabTransformer model (Bohoran et al., 2023b), designed to optimally
 6 perform on tabular inputs with a mixture of categorical and continuous vari-
 7 ables. TabTransformer models do not use positional encodings but instead
 8 unique identifiers in the categorical feature embeddings. In addition, they
 9 use layer normalisation to process numerical features and are able to take ad-
 10 vantage of transformer techniques to effectively process categorical features,
 11 both of which are concatenated and then fed to a FC ANN for the final layer
 12 of learning.

13 2.4.4. Gradient Boosting

14 GB (Friedman, 2001) is deployed to predict maximum SLT characteristic
 15 values. GB is an ensemble learning method that functions by adding new
 16 decision tree (DT) models to base learners that are trained to improve the
 17 final prediction (Bohoran et al., 2023a). It is an iterative process in which
 18 first a DT is trained on the training data. The residual errors between the
 19 predictions from this model and the true values are found and then a second
 20 DT is trained using again the original features for inputs and the residual
 21 errors as targets. This process is repeated a given number of times and a
 22 final prediction from the model is taken by simply summing the predictions
 23 from all DTs in the ensemble.

24 The problem of supervised regression can be defined as the creation of a
 25 mapping $F(\mathbf{X}_I)$ that takes in a random variable $\mathbf{X}_I$ and predicts some output
 26 y . The prediction is denoted by $\hat{y}$. The ambition is then to find the optimal
 27 mapping $F^*(\mathbf{X}_I)$, which minimises a predetermined error between $\hat{y}$ and y .
 28 This error, or loss, can be expressed as $L(y, F(\mathbf{X}_I))$ where L is the chosen
 29 loss function. $F^*(\mathbf{X}_I)$ can hence be described mathematically as

$$F^*(\mathbf{X}_I) = \arg \min_F E_{y, \mathbf{x}_I} L(y, F(\mathbf{X}_I)), \quad (5)$$

30 where E denotes the expected value. GB takes a non-parametric approach to
 31 the estimation function and applies numerical optimisation in the function

1 space. That is, it considers $F(\mathbf{X}_I)$ evaluated at each point of $\mathbf{X}_I$ to be a
 2 parameter and it seeks to minimise the expected error of the mapping at
 3 each $\mathbf{X}_I$ with respect to $F(\mathbf{X}_I)$. $F(\mathbf{X}_I)$ for a GB model can therefore be
 4 described by:

$$F(\mathbf{X}_I) = \sum_{i=1}^N \beta_i Z(\mathbf{X}_I; \zeta_i), \quad (6)$$

5 where each $Z(\mathbf{X}_I; \zeta_i)$ is an incremental function (boost) or DT which is
 6 trained using the same input features to predict the residual errors from
 7 the previous step in the algorithm. ζ_i represents the splitting variables, split
 8 locations, and terminal node means of each of the N individual function
 9 approximation methods (DTs), and β_i is a line search multiplication term,
 10 which is chosen to minimise the overall loss. $F^*(\mathbf{X}_I)$ is then defined as the
 11 model that minimises a predetermined loss function.

12 2.5. Implementation and Training

13 Selection of the input features of the ML models used in this work is done
 14 by capitalising on knowledge acquired by previous related studies, and these
 15 features differ slightly for each different application. For predicting maximum
 16 wave characteristics and wave types of 2D SLTs generated by granular slides,
 17 the input features of the models were F , S , M , and α . The corresponding
 18 input features for 2D SLTs generated by block slides were F , S , M , B , ϕ , and
 19 T_s . On the contrary, producing and evaluating ML models that can predict
 20 SLT characteristics across different slide models and water body geometries
 21 requires the merging of data sets that have both different and overlapping
 22 columns. In this study, to create and evaluate slide model- and water body
 23 geometry-invariant networks, data from all included data sets is combined
 24 by only using the three common input features F , S , M , two categorical
 25 features to indicate a granular or block slide and if the example refers to
 26 a 2D or 3D experiment, and finally γ . D_g was not used as an input as
 27 it was found to have a negligible effect on predictions in Heller and Hager
 28 (2010). α , although significant and varied in Heller and Hager (2010) and
 29 Huber (1980), was also not used as it had not been varied in the studies
 30 for block slide experiments, meaning it is not entirely independent as an
 31 input variable. Finally, when designing sequence models for wave evolution
 32 predictions, the relative distance x/h in 2D or r/h in 3D is also included to

capture spatial information when training. Then, the output of the sequence model was a vector of length equal to the number of wave gauges. The data in Huber (1980) does not provide decay values for all of the six wave gauges simultaneously. As such, these missing values were interpolated according to the decay term in Huber and Hager (1997).

The data set in this work used for invariant a and T predictions consisted of 403 experiments. Each of the 18 3D experiments of Heller and Spinneken (2015) was broken down into six distinct 2D example points by treating each γ line as its own example. This angle then became an input feature (for 2D $\gamma = 0^\circ$). All readings at 90° were excluded as they are adjacent to the side wall and propagate as edge waves (governed by different physics than the herein investigated offshore propagating waves (Longuet-Higgins, 1967; Heller and Spinneken, 2015)). Combining the remaining 108 ($= 18 \times 6$) data points with the 385 ($= 403 - 18$ 3D tests) 2D experiments gave a total of 493 examples for use. The data set in this work used for invariant H prediction was the data set described above complemented with 90 ($= 18 \times 5$) 3D experiments of Huber (1980) (again broken down into five distinct 2D example points by treating each γ in the same way) giving a total of 583 examples for use.

The available data was split at random into two disjoint sets: The training (80% of total examples) and testing sets (20%). The former was used to train and optimise each ML model, while the latter was used to evaluate the performance, as described in Section 3. All example input features were normalised before being fed to the ML models.

All ML algorithms and techniques have properties, known as hyperparameters, which heavily impact performance. While model weights (elements of $\mathbf{W}$) are the values that change when an ML system is said to learn and where the knowledge of the model is stored, hyperparameters dictate how these values change. Hyperparameters include, but are not limited to, the number of neurons in a given ANN layer, the number of layers of LSTM units in a RNN model, or the number of boosts in a GB model. To produce optimal ML models, hyperparameter tuning, the process of identifying the hyperparameters which allow the ML algorithm to minimise the given loss function, was performed. In this work, hyperparameter tuning was done using a process known as grid search. In this process, all combinations of hyperparameter values were used to produce a model on the training data (Section 2.5) and the combination giving the lowest value for the chosen loss function was chosen.

1 The hyperparameters that were considered for GB training were the learn-
2 ing rate, the loss function, the evaluation metric, the number of estimators,
3 the maximum depth, the minimum sample leaf, and the minimum sample
4 split. For optimising the design of the FC classification ANN, the hyperpa-
5 rameters that were tuned were the number of hidden layers, the number of
6 hidden units, the activation function, the optimiser, the loss function, the
7 evaluation metric, the batch size, and the number of epochs. The LSTM and
8 GRU hyperparameters were largely the same as for FC ANNs. Lastly, the
9 relevant hyperparameters for the TabTransformer models were the dimension
10 of the categorical feature encodings, the number of layers of TabTransformer
11 units (depth), the number of heads for the attention mechanism, the atten-
12 tion and feed forward dropout values (which dictate the regularisation rate
13 for the attention and feed forward layers, respectively), the loss function, the
14 optimisation algorithm, the batch size, and the number of epochs.

15 The ML training and testing was conducted on a workstation with an
16 NVIDIA TITAN RTX (24GB) GPU. Python was used to implement ML
17 algorithms in this paper¹ and all experimental run information was stored in
18 MS Excel (.xlsx) files.

19 All ANN and sequence model algorithms were built using the Keras li-
20 brary with TensorFlow back end. The wave type classification model was
21 initially defined using the *Sequential()* function, layers were then added us-
22 ing the *Dense()* class, and compiled to form the model using the *Compile()*
23 function. The model was trained by calling *fit()*. Similarly, the LSTM and
24 GRU models were created using the *Sequential()* function and then instan-
25 tiating the *LSTM()* and *GRU()* layer classes, respectively. GB models were
26 developed using the CatBoost (Dorogush et al., 2018) class *CatBoostRegres-*
27 *sor()*. The hyperparameter search was performed using the Scikit-Learn
28 class *GridSearchCV()*. The Transformer-based models were implemented in
29 PyTorch, utilising the TabTransformer module with the in-built *Backward()*
30 and *Optimizer()* functions to oversee the training process.

¹The models developed in this article can be accessed at:
https://github.com/davjenko/Ocean_Engineering_VSI.

1 2.6. Model Evaluation and Statistical Comparison

2 To assess the regressions, the coefficient of determination R^2 was used:

$$R^2 = 1 - \frac{\sum_{i=1}^{m_{Te}} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{m_{Te}} (y_i - \bar{y}_i)^2}, \quad (7)$$

3 where $\hat{y}_i$ is the predicted outcome, y_i is the true value, and m_{Te} is the number
4 of testing examples. For evolution predictions, the average was taken of the
5 values at each output frame.

6 The mean absolute error (MAE) and mean absolute percentage error
7 (MAPE) were also employed to evaluate regressions:

$$MAE = \frac{\sum_{i=1}^{m_{Te}} |\hat{y}_i - y_i|}{m_{Te}}, \quad (8)$$

$$MAPE = \frac{1}{m_{Te}} \sum_{i=1}^{m_{Te}} \left| \frac{\hat{y}_i - y_i}{\max(\epsilon, |y_i|)} \right|, \quad (9)$$

8 where $\epsilon = 2^{-52}$ is the tolerance constant to avoid complications if y_i is zero.
9 For qualitative assessment, plots of the predicted values for the various meth-
10 ods versus the experimental values were also drawn.

11 The accuracy metric

$$Accuracy = \frac{1}{m_{Te}} \sum_{i=1}^{m_{Te}} Ind(\hat{y}_i, y_i), \quad (10)$$

12 was used to evaluate the wave type classification performance, where

$$Ind(\hat{y}_i, y_i) = \begin{cases} 1 & \text{if } \hat{y}_i = y_i, \\ 0 & \text{if } \hat{y}_i \neq y_i, \end{cases} \quad (11)$$

13 is the indicator function. Confusion matrices were also employed.

14 To determine if there is a statistically significant difference (at a 95% con-
15 fidence level) between the performance of the proposed and existing methods,
16 the Wilcoxon signed-rank test was used (Wilcoxon, 1992) with the alterna-
17 tive hypothesis being that the proposed ML methods prediction errors were
18 lower than those of the existing methods. The p -value given is the probability
19 that the increase in performance is due to random chance; $p < 0.05$ is taken
20 as statistically significant.

1 3. Results

2 3.1. Final Model Hyperparameters

3 The optimal hyperparameter value ranges of the GB, FC ANN, LSTM/GRU,
4 and TabTransformer models in all single- and cross-study ML experiments
in this work, are given in Tables 3, 4, 5, and 6, respectively.

Table 3: Optimal hyperparameter value ranges of the best performing GB regressions for maximum SLT characteristic prediction in all single- and cross-study ML experiments in this work. R^2 denotes the coefficient of determination and MSLE denotes the mean square log error.

Hyperparameter	Value range
Learning rate	0.01 – 0.3
Number of estimators	10 – 75
Maximum depth	1 – 6
Minimum sample leaf	2 – 9
Minimum sample split	2 – 5
Loss function	Squared error R^2
Evaluation metric	MSLE

Table 4: Optimal hyperparameter values of the best performing FC ANN to predict SLT wave types. ReLU is the rectified linear unit activation function.

Hyperparameter	Value
Number of hidden layers	3
Number of units	50
Activation function	ReLU, softmax
Optimisation algorithm	RMSprop
Loss function	Categorical cross-entropy
Evaluation metric	Accuracy
Batch size	10
Number of epochs	200

5

Table 5: Optimal hyperparameter value ranges of the best performing LSTM and GRU models for SLT characteristic evolution prediction in all single- and cross-study ML experiments of this work. MSE denotes mean squared error and MAE denotes mean absolute error.

Hyperparameter	Value range
Number of hidden layers	3 – 4
Number of units	32 – 64
Activation function	Sigmoid function
Optimisation algorithm	RMSprop
Loss function	MSE
Evaluation metric	MAE
Batch size	10 – 20
Number of epochs	100 – 200

Table 6: Optimal hyperparameter ranges of the best performing slide model- and water body geometry-invariant TabTransformer model for SLT characteristic evolution prediction. MSE denotes mean squared error and MAE denotes mean absolute error.

Hyperparameter	Value
Dimension	32
Depth	6
Number of heads	8
Attention dropout	0.1
Feed forward dropout	0.1
Loss function	MSE
Optimisation algorithm	RMSprop
Batch size	1
Number of epochs	200

3.2. Prediction of SLT Characteristics in Individual Experimental Set-ups

Tables 7 and 8 present the maximum wave characteristic and evolution performance comparisons, respectively, of the proposed GB ML prediction methods and the empirical equations from Heller and Hager (2010) for 2D SLTs created by granular slides. The same information is given in Tables 9 and 10 for Heller and Spinneken (2013) for 2D SLTs created by block slides, Tables 11 and 12 for Huber (1980) for 3D SLTs created by granular slides, and Tables 13 and 14 for Heller and Spinneken (2015) for 3D SLTs created by block slides. Baseline refers to the predictions from the empirical equations given in the respective study.

Table 7: Goodness of fit to the relative maximum wave amplitude a_M/h , height H_M/h , and period $T_M(g/h)^{1/2}$ values of the proposed GB and empirical equations from the baseline study Heller and Hager (2010) for 2D SLTs generated by granular slides. MAPE denotes mean absolute percentage error and MAE denotes mean absolute error. The Wilcoxon signed-rank test p -values are also given. Boldface indicates best performance.

Parameter	Metric	Baseline	GB	GB-Baseline p -value
a_M/h	R^2	0.93	0.94	-
	MAPE	0.25	0.23	0.18
	MAE	0.098	0.083	0.16
H_M/h	R^2	0.87	0.96	-
	MAPE	0.16	0.10	0.0030
	MAE	0.10	0.050	0.0030
$T_M(g/h)^{1/2}$	R^2	0.57	0.73	-
	MAPE	0.26	0.15	0.0020
	MAE	2.26	1.42	0.0020

Table 8: Goodness of fit to the relative wave amplitude $a(x)/h$, height $H(x)/h$, and period $T(x)(g/h)^{1/2}$ evolution data of the proposed sequence models and equations from the baseline study Heller and Hager (2010) for 2D SLTs generated by granular slides. MAPE denotes mean absolute percentage error and MAE denotes mean absolute error. The Wilcoxon signed-rank test p -values are also given. Boldface indicates best performance.

Parameter	Metric	Baseline	LSTM	GRU	LSTM-Baseline p -value	GRU-Baseline p -value
$a(x)/h$	R^2	0.80	0.93	0.93	-	-
	$MAPE$	0.29	0.15	0.15	0.014	0.033
	MAE	0.11	0.064	0.066	0.011	0.04
$H(x)/h$	R^2	0.74	0.94	0.94	-	-
	$MAPE$	0.21	0.12	0.12	0.0010	0.0015
	MAE	0.13	0.07	0.07	0.0011	0.0012
$T(x)(g/h)^{1/2}$	R^2	-	0.75	0.76	-	-
	$MAPE$	-	0.14	0.13	-	-
	MAE	-	1.90	1.80	-	-

1 GB achieved excellent performance in predicting maximum SLT charac-
 2 teristics in individual studies with consistently high R^2 values, $R^2 = 0.94$,
 3 0.97 , and 0.98 for a_M on individual studies (Tables 7, 9, and 13, respectively).
 4 Similar values of R^2 can be observed for H_M predictions, $R^2 = 0.96$, 0.97 ,
 5 0.92 , and 0.98 (Tables 7, 9, 11, and 13, respectively). For both of these sets
 6 of predictions, $MAPE$ and MAE values were broadly similar, displaying the
 7 on average low relative and absolute error terms in predictions. GB meth-
 8 ods, while achieving notably lower values of R^2 and higher errors terms as
 9 measured by $MAPE$ and MAE compared to a_M and H_M , gave good overall
 10 predictions for $T_M(g/h)^{1/2}$. $R^2 = 0.73$, 0.84 , and 0.74 (Tables 7, 9, and 13,
 11 respectively) are strong values, partially considering the usual difficulty in
 12 predicting this value. As outlined in Tables 8 and 10, the sequence models
 13 designed to predict the evolution of $a(x)/h$, $H(x)/h$, and $T(x)(g/h)^{1/2}$ for
 14 each of the two 2D slide models gave predictions with a significant ($p <$
 15 0.05) improvement in performance over the existing empirical equations as
 16 measured by $MAPE$ and MAE , along with similarly high R^2 values. The
 17 $a(r)/h$ and $H(r)/h$ decay predictions that related to 3D block slides were
 18 comparable to those made by empirical equations ($p \geq 0.05$) despite the se-
 19 quence models achieving a higher R^2 score (Table 14). Table 12 shows that
 20 ML for prediction of 3D granular slides using the data of Huber (1980), while
 21 still outperforming the empirical equations, yields results subpar compared
 22 to the other data sets.

1 The classification FC ANN developed for predicting the wave type of a
2 given 2D granular SLT achieved a high accuracy score of 98% on the allotted
3 testing set. An exact breakdown of the performance across the four wave
4 types is illustrated in the confusion matrix in Fig. 3. A direct comparison
5 between the ML model developed in this work and the decision boundaries
6 described by Heller and Hager (2011) can be seen in Fig. 4. Importantly, the
7 ML model was able to differentiate between cnoidal- and solitary-like wave
types.

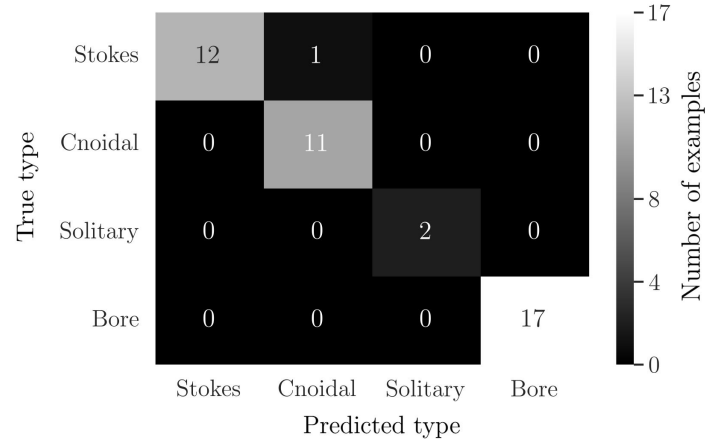


Figure 3: Confusion matrix for ML predictions of the wave type in the 2D granular SLTs of Heller and Hager (2011). The colour of each segment represents the number of examples from the testing set which fall into that category.

8

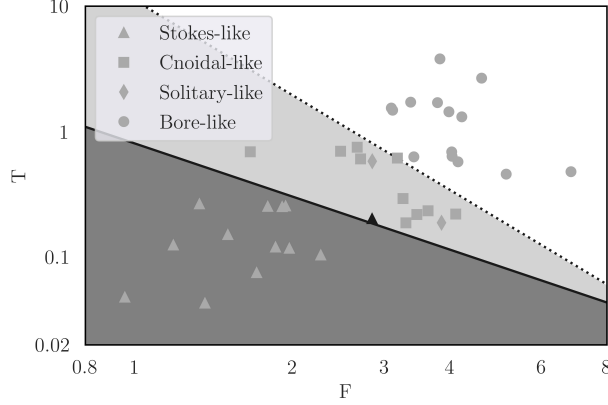


Figure 4: ML wave type prediction using the data of Heller and Hager (2011) where $T = S^{\frac{1}{3}} M \cos\left(\frac{6\alpha}{7}\right)$ is the wave type product and F is the slide Froude number. Also shown are the classification boundaries proposed by the same study: $T < \frac{4}{5}F^{-\frac{7}{5}}$ for Stokes-like, $\frac{4}{5}F^{-\frac{7}{5}} \leq T \leq 11F^{-\frac{5}{2}}$ for cnoidal- or solitary-like, and $T > 11F^{-\frac{5}{2}}$ for bore-like waves. Light grey points indicate a correct prediction, whereas black points indicate an incorrect prediction. Note that the empirical equations cannot differentiate between cnoidal- and solitary-like waves.

Table 9: Goodness of fit to the relative maximum wave amplitude a_M/h , height H_M/h , and period $T_M(g/h)^{1/2}$ values of the proposed GB and empirical equations from the baseline study Heller and Spinneken (2013) for 2D SLTs generated by block slides. The Wilcoxon signed-rank test p -values are also given. Boldface indicates best performance.

Parameter	Metric	Baseline	GB	GB-Baseline p -value
a_M/h	R^2	0.84	0.97	-
	$MAPE$	0.26	0.09	8.89×10^{-6}
	MAE	0.050	0.020	1.90×10^{-5}
H_M/h	R^2	0.92	0.97	-
	$MAPE$	0.14	0.090	0.024
	MAE	0.052	0.033	0.022
$T_M(g/h)^{1/2}$	R^2	0.29	0.84	-
	$MAPE$	0.26	0.093	1.60×10^{-4}
	MAE	1.51	0.63	1.60×10^{-4}

Table 10: Goodness of fit to the relative wave amplitude $a(x)/h$, height $H(x)/h$, and period $T(x)(g/h)^{1/2}$ evolution data of the proposed sequence models and equations given by the baseline study Heller and Spinneken (2013) for 2D SLTs generated by block slides. MAPE denotes mean absolute percentage error and MAE denotes mean absolute error. The Wilcoxon signed-rank test p -values are also given. Boldface indicates best performance.

Parameter	Metric	Baseline	LSTM	GRU	LSTM-Baseline p -value	GRU-Baseline p -value
$a(x)/h$	R^2	0.86	0.97	0.97	-	-
	$MAPE$	0.37	0.13	0.011	0.040	0.019
	MAE	0.041	0.019	0.018	0.020	0.0010
$H(x)/h$	R^2	0.92	0.97	0.96	-	-
	$MAPE$	0.22	0.12	0.10	0.020	0.00088
	MAE	0.047	0.027	0.029	0.021	0.0040
$T(x)(g/h)^{1/2}$	R^2	0.14	0.67	0.75	-	-
	$MAPE$	0.25	0.12	0.10	0.0034	0.00058
	MAE	2.22	1.058	0.90	0.0028	0.00084

Table 11: Goodness of fit to the relative maximum wave height H_M/h values of GB predictions for the data of Huber (1980) for 3D SLTs generated by granular slides. MAPE denotes mean absolute percentage error and MAE denotes mean absolute error.

Parameter	Metric	GB
H_M/h	R^2	0.92
	$MAPE$	0.23
	MAE	0.015

Table 12: Goodness of fit to the relative wave height $H(r)/h$ evolution data of the proposed sequence models and equations given by the baseline study Huber and Hager (1997) for 3D SLTs created by granular slides. MAPE denotes mean absolute percentage error and MAE denotes mean absolute error. The Wilcoxon signed-rank test p -values are also given. Boldface indicates best performance.

Parameter	Metric	Baseline	LSTM	GRU	LSTM-Baseline p -value	GRU-Baseline p -value
$H(r)/h$	R^2	0.57	0.82	0.76	-	-
	$MAPE$	0.64	0.64	0.66	0.047	0.047
	MAE	0.015	0.015	0.016	0.0031	0.01

Table 13: Goodness of fit to the relative maximum wave amplitude a_M/h , height H_M/h , and period $T_M(g/h)^{1/2}$ values of GB predictions for the data of Heller and Spinneken (2015) for 3D SLTs generated by block slides. MAPE denotes mean absolute percentage error and MAE denotes mean absolute error.

Parameter	Metric	GB
a_M/h	R^2	0.98
	$MAPE$	0.16
	MAE	0.019
H_M/h	R^2	0.98
	$MAPE$	0.15
	MAE	0.033
$T_M(g/h)^{1/2}$	R^2	0.74
	$MAPE$	1.04
	MAE	0.34

Table 14: Goodness of fit to the relative wave amplitude $a(r)/h$ and height $H(r)/h$ evolution data of the proposed sequence models and equations given by the baseline study Heller and Spinneken (2015) for 3D SLTs created by block slides. MAPE denotes mean absolute percentage error and MAE denotes mean absolute error. The Wilcoxon signed-rank test p -values are also given. Boldface indicates best performance.

Parameter	Metric	Baseline	LSTM	GRU	LSTM-Baseline p -value	GRU-Baseline p -value
$a(r)/h$	R^2	0.91	0.96	0.98	-	-
	$MAPE$	0.44	0.21	0.31	0.15	0.12
	MAE	0.018	0.013	0.011	0.16	0.061
$H(r)/h$	R^2	0.90	0.95	0.96	-	-
	$MAPE$	0.24	0.23	0.24	0.32	0.36
	MAE	0.022	0.017	0.020	0.29	0.27

1 3.3. Slide Model- and Water Body Geometry-Invariant Modelling

2 Figures 5a, b, and c show the slide model- and water body geometry-
 3 invariant GB predictions developed to output a_M/h , H_M/h , and $T_M(g/h)^{1/2}$,
 4 respectively, against the observed values. The performance of the slide model-
 5 and water body geometry-invariant predictions of the maximum values and
 6 evolution of wave characteristics are given in Tables 15 and 16, respectively.

7 The GB predictions across multiple experimental set-ups, for which no
 8 empirical equations exist, were found to perform excellently, with $R^2 = 0.96$
 9 for both a_M and H_M predictions, and $R^2 = 0.73$ on the commonly more
 10 difficult T_M predictions. All relative and absolute error terms, as measured
 11 by *MAPE* and *MAE*, also indicate strong performance of the predictions
 12 across all three prediction variables. Figures 6a, b, and c display the his-
 13 tograms of errors of these three models. All three error distributions present
 14 approximately normal, again suggesting the models have captured the over-
 15 all trend of the data. The errors for a_M and H_M in particular are grouped
 16 tightly around 0, with only a very small number of examples giving larger
 17 errors. The errors for T_M predictions are observed to be more spread and
 18 across a larger range than a_M and H_M . This is in line with observations
 19 from other evaluation metrics, but these errors still present mostly around 0.
 20 Across all models, no decrease in performance was seen compared to apply-
 21 ing the methods on the individual data sets using any of the three evaluation
 22 metrics.

23 The slide model- and water body geometry-invariant models developed
 24 to predict $a(x)/h$, $H(x)/h$, and $T(x)(g/h)^{1/2}$ evolution also showed strong
 25 performance of $R^2 = 0.93$, 0.95 , and 0.73 , respectively (Table 16). Figures
 26 7a, b, and c show the slide model- and water body geometry-invariant height
 27 predictions versus the observed values for the LSTM model at relative prop-
 28 agation distances (r/h and x/h) 3 to 5, 6 to 10, and 15 to 22, respectively.
 29 In all three sub-figures, almost all data points align closely with the diagonal
 30 line, indicating the model has captured the trend of the data well. While
 31 not entirely followed, an expected general decrease in the values of prediction
 32 is seen across the three sub-figures as the wave propagates further from the
 33 slide impact zone, but no noticeable increase in stray from the diagonal is
 34 observed, again indicating strong performance across the model.

Table 15: Goodness of fit of the proposed slide model- and water body geometry-invariant models for relative maximum wave amplitude a_M/h , height H_M/h , and period $T_M(g/h)^{1/2}$ values for any SLT analysed in this study. MAPE denotes mean absolute percentage error and MAE denotes mean absolute error.

Parameter	Metric	GB
a_M/h	R^2	0.97
	$MAPE$	0.16
	MAE	0.048
H_M/h	R^2	0.96
	$MAPE$	0.18
	MAE	0.051
$T_M(g/h)^{1/2}$	R^2	0.74
	$MAPE$	0.17
	MAE	1.46

Table 16: Evaluation of slide model- and water body geometry-invariant sequence networks on predicting the evolution of relative wave amplitude $a(x \text{ or } r)/h$, height $H(x \text{ or } r)/h$, and period $T(x \text{ or } r)(g/h)^{1/2}$ data for any SLT analysed in this study. TT denotes Tab-Transformer, MAPE denotes mean absolute percentage error, and MAE denotes mean absolute error. Boldface indicates best performance.

Parameter	Metric	LSTM	GRU	TT
$a(x \text{ or } r)/h$	R^2	0.92	0.93	0.93
	$MAPE$	0.38	0.37	0.36
	MAE	0.06	0.06	0.05
$H(x \text{ or } r)/h$	R^2	0.95	0.95	0.94
	$MAPE$	0.28	0.30	0.28
	MAE	0.05	0.05	0.06
$T(x \text{ or } r)(g/h)^{1/2}$	R^2	0.80	0.74	0.80
	$MAPE$	0.14	0.13	0.11
	MAE	1.40	1.40	1.40

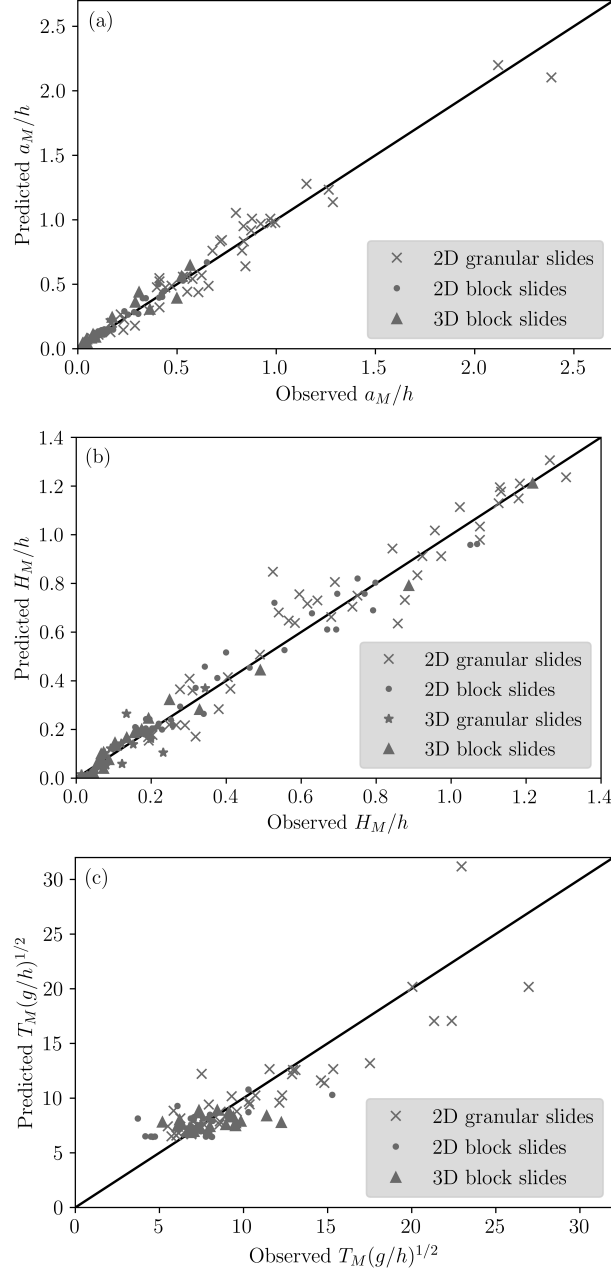


Figure 5: Comparison between the slide model- and water body geometry-invariant GB model predictions and observed values for the (a) maximum wave amplitude ($R^2 = 0.97$), (b) height ($R^2 = 0.96$), and (c) period ($R^2 = 0.74$).

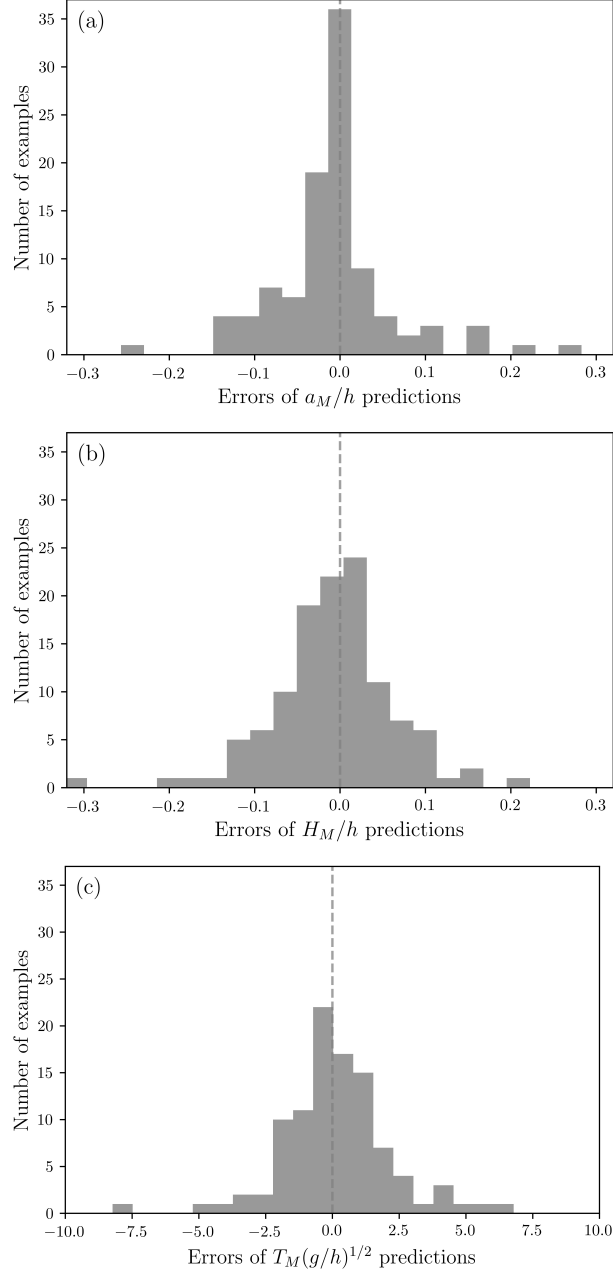


Figure 6: Histograms of errors (observed - predicted) in maximum wave (a) amplitude ($R^2 = 0.97$), (b) height ($R^2 = 0.96$), and (c) period ($R^2 = 0.74$) predictions by the slide model- and water body geometry-invariant models.

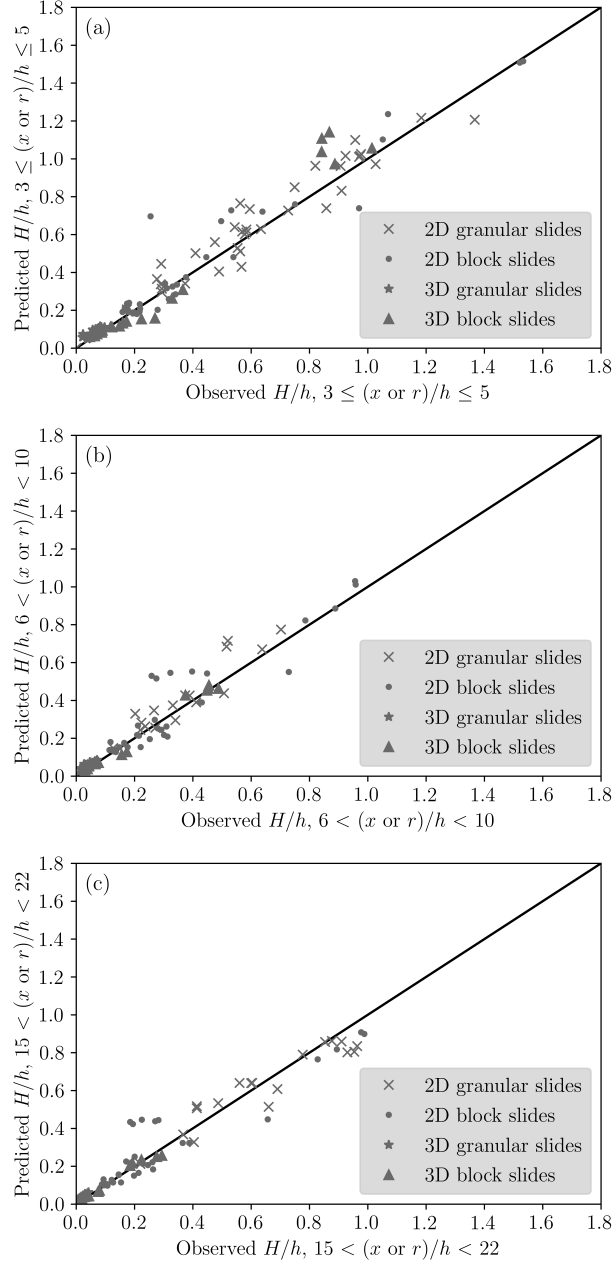


Figure 7: Comparison between the slide model- and water body geometry-invariant LSTM model predictions and observed values for the height ($R^2 = 0.95$) at relative propagation distances (r/h and x/h) (a) 3 to 5, (b) 6 to 10, and (c) 15 to 22.

1 4. Discussion

2 4.1. Main Findings

3 In this work, various ML models were implemented to predict character-
4 istics of SLTs of individual experimental set-ups. The models performed well
5 on nearly all problems based on the global evaluation metrics used and were
6 mostly superior to existing empirical predictions, as made evident by the
7 obtained Wilcoxon signed-rank test p -values. A step forward was then taken
8 with the designed ML models to make predictions across multiple data sets
9 acquired using different slide models and water body geometries, for which
10 no empirical equations exist. These models also achieved a high level of per-
11 formance, comparable to that of the individual data set models, indicating
12 great capacity to generalise.

13 ML predictions were able to significantly outperform the empirical equa-
14 tions, except for a_M/h in 2D granular slides, for which the Wilcoxon signed-
15 rank test p -values did not indicate a significant improvement (Table 7). The
16 models developed to predict $T_M(g/h)^{1/2}$ did not perform to the same stan-
17 dard as those for a_M/h and H_M/h . This decrease in performance is to be
18 expected as the corresponding empirical equations had a similar drop in qual-
19 ity. Across all experiments, the ML predictions were found to perform better
20 in the lower value ranges of predictions, while higher value ranges were some-
21 times less well predicted, particularly for T_M (Fig. 5). A similar trend for the
22 superior performance of a_M and H_M over T_M prediction is visible in Figure
23 6, in which the errors are seen to be overall further than 0 for the latter.

24 This study is the first to apply sequence models for studying wave evolu-
25 tion. These ML predictions also achieved an overall high level of performance.
26 The sequence models, however, failed to learn the $T(r)(g/h)^{1/2}$ evolution for
27 3D block experiments due to high variance in the prediction variables. Fortu-
28 nately, T plays a significantly less important role in SLT hazard assessment
29 than a or H (Evers et al., 2019). It was noted that LSTM and GRU gave
30 very similar performances in all cases.

31 The FC ANN was able to classify all four SLT wave types. The confusion
32 matrix (Fig. 3) showed that in all but one case the prediction and true
33 wave type classes match. For the only incorrect prediction in the testing set,
34 the placement of the incorrect prediction is close to the empirical decision
35 boundary that separate the predicted and true classes (Fig. 4). An advantage
36 of ML is the ability to separately predict cnoidal- and solitary-like waves, a

prediction typically not possible using empirical equations (Heller and Hager, 2011).

These ML predictions, based on experimental conditions, are aimed eventually for preliminary hazard assessments, e.g. for situations where a mass starts to creep in proximity of a water body (Baihetan Reservoir case, Chen et al., 2023; La Palma Island, Abadie et al., 2012) or in the planning and operational phases of hydropower dams (K  thai reservoir, Fuchs et al., 2011). Therefore, the slide model- and geometry-invariant model was applied to the Fritz et al. (2001) idealised 2D study of the 1958 Lituya Bay case and predicted $H_M = 166$ m. This is in excellent agreement with the prediction of $H_M = 162$ m in Fritz et al. (2001), corresponding to an error of 2.47%. Similarly, by applying the slide model- and water body geometry-invariant wave height decay model to the 2007 Chehalis Lake case (Section 2.3.2), and using the wave run-up height equation of Evers and Boes (2019), under the assumption $a = H/2$, the model inference for R was 41.4 m. This is in excellent agreement (error of 9.52%) with the field observation of $R = 37.8$ m in Roberts et al. (2013). Finally, the ANN classification model was applied on the idealised 2D study of the 1958 Lituya Bay case and predicted a Stokes-like wave type. This validation prediction of the Stokes-like wave could not be fully confirmed due to the presence of wave reflection in the corresponding wave profile in Fig. 11 of Fritz et al. (2001).

In future work, these models will be further verified by applying them to additional real-world cases. Such efficient and inexpensive preliminary predictions help to decide whether more expensive and time-consuming prototype-specific physical or numerical model tests are necessary.

4.2. Study Limitations

Even though two of the developed models were validated on an idealised 2D experiment of the 1958 Lituya Bay case and the 2007 3D Chehalis Lake case as described in Roberts et al. (2013), the parameters of which were not seen in the training data, the generalisation capability of the ML predictions can be further improved. Further, the ML predictions implemented herein are based on data from idealised laboratory conditions. As such, similar limitations as for the laboratory data apply; this involves a limited data set, idealised slide models (granular and block slides), water body geometries (2D and 3D) and bathymetries (horizontal water body bottom).

Some of the input features from each study’s data set, such as α , needed to be removed as the slide model- and water body geometry-invariant models

1 need to have precisely identical inputs to function. This results in a loss of in-
2 formation. Due to the lack of measured values of s for the Huber (1980) data
3 set, an approximation method was used (Section 2.1). While this method
4 was based on equations and methods common in literature, this does intro-
5 duce possible uncertainty into the ground truth values used. The applied ML
6 methods in this study were unable to appropriately predict period evolution
7 in 3D block SLTs, due to high variability in the values of the prediction
8 variables. However, it is common that methods which aim to predict wave
9 period characteristics are less accurate (Heller and Hager, 2010; Panizzo et
10 al., 2005).

11 Due to the nature of ML, the models of this study provided predictions of
12 tsunami characteristics without giving mathematical justification as to how
13 they did so. Future research will involve the boosting of the developed models
14 by including explainability mechanisms (Bohoran et al., 2023a), which give
15 more insight into how a final prediction is made.

16 Finally, the data concerns the landslide impact and wave propagation
17 zones only and wave run-up is not covered. However, the predicted wave
18 characteristics contain all information for typical run-up calculation equa-
19 tions (Evers et al., 2019), as demonstrated for the 2007 Chehalis Lake case
20 (Section 4.1).

21 5. Conclusions

22 Subaerial landslide-tsunamis (SLTs) are natural phenomena that have
23 the potential to cause extensive damage to both human life and property.
24 Reliable, robust, and rapid methods to predict these events are crucial for
25 hazard assessment and mitigation. A large number of empirical equations
26 on idealised laboratory studies (e.g. granular or block slide models; two-
27 dimensional 2D flume or three-dimensional 3D basin water body geometries)
28 are available for these purposes. However, to date, no study has incorporated
29 the full range of slide models and water body geometries. Machine learning
30 (ML) algorithms have the potential to develop universal models by merging
31 data sets across numerous studies with a wide range of conditions. State-of-
32 the-art ML methods have been implemented and applied herein to the data
33 sets of four laboratory studies, namely, Heller and Hager (2010) (2D granular
34 slides), Heller and Spinneken (2013) (2D block slides), Huber (1980) (3D
35 granular slides), and Heller and Spinneken (2015) (2D and 3D block slides),
36 individually and combined with a total of 421 experiments.

The analysed SLT parameters were the wave amplitude a , height H , and period T , both for maximum values and evolution over x (for 2D) and radial distance r and wave propagation angle γ (for 3D).

Gradient boosting models were designed to predict maximum SLT characteristics values. Sequence models (long short-term memory, gated recurrent unit, and transformer-based) were developed to predict the evolution of these characteristics. In addition, the wave types were classified using Artificial Neural Network (ANN). After development and tuning, all ML predictions were evaluated on unseen testing examples and compared with existing empirical equations where possible. The results obtained showed that the ML predictions, with one noted exception, work well and overall outperform or are on par with existing empirical equations. The ML predictions across multiple slide models and water body geometries maintained performance. The accuracy achieved by the ANN wave type classification was 98%. Importantly, in contrast to the empirical equations, ANN was able to discriminate between solitary- and cnoidal-like waves.

The slide model- and water body geometry-invariant model for the maximum value of H was successfully validated on the Fritz et al. (2001) idealised 2D study of the 1958 Lituya Bay case, resulting in an error of 2.47%. Similarly, the slide model- and water body geometry-invariant predictions for wave height decay were successfully validated on the 2007 Chehalis Lake case run-up height, resulting in an error of 9.52%.

Future research may involve feature tokeniser transformer-based methods (Gorishniy et al.,2023; Bohoran et al.,2023b) and physics-informed neural network models which incorporate governing equations towards developing models that may even better predict field cases.

Notation

a	= wave amplitude, m;
a	= value of a neuron in an ANN;
B	= blockage ratio;
b	= flume width, m;
$\mathbf{b}$	= bias matrix;
b_s	= slide width, m;

1	$\mathbf{c}$	= long-term state for LSTM units;
2	D_g	= relative grain diameter;
3	d_g	= grain diameter, mm;
4	d_K	= dimension of keys;
5	E	= expected value;
6	F	= slide Froude number;
7	F	= mapping;
8	$\mathbf{f}$	= term controlling the forget gate for a LSTM cell;
9	g	= gravitational acceleration, m/s ² ;
10	$\mathbf{g}$	= term carrying information about previous and current cells
11		in LSTM and gated recurrent units;
12	H	= wave height, m;
13	$\mathbf{h}$	= RNN memory state;
14	h	= water depth, m;
15	Ind	= indicator function;
16	$\mathbf{i}$	= term controlling the input gate for a LSTM cell;
17	K	= keys (transformer models);
18	L	= loss function;
19	l	= layer of neural network;
20	M	= relative slide mass;
21	m_s	= slide mass, kg;
22	m_{Te}	= number of testing examples;
23	m_{To}	= total number of examples;

1	N	= number of DTs in a GB model;
2	$\mathbf{o}$	= term controlling the output gate for a LSTM cell;
3	p	= value from Wilcoxon signed-rank test;
4	Q	= queries (transformer models);
5	R	= run-up height;
6	R^2	= coefficient of determination;
7	$\mathbf{r}$	= gate controller in a GRU model's cell;
8	r	= radial distance, m;
9	S	= relative slide thickness;
10	s	= slide thickness, m;
11	T	= wave period, s;
12	T	= wave type product;
13	T_s	= expression considering the transition type;
14	t	= frame of a RNN;
15	t_s	= characteristic time of submerged landslide motion, s;
16	V	= values (transformer models);
17	V_s	= slide impact velocity, m/s;
18	V_s	= bulk slide volume, m ³ ;
19	$\mathbf{W}$	= neural network parameter tensor;
20	X	= relative streamwise distance;
21	$\mathbf{X}_I$	= input features;
22	$\mathbf{x}$	= input sequence for sequence model;
23	x	= streamwise distance, m;

1	y	= target outputs used for supervised ML model training;
2	$\hat{y}$	= prediction made by a ML model for a given example;
3	$\mathbf{z}$	= gate controller in a gated recurrent unit model's cell;
4	Z	= regression DT;
5	α	= slide impact angle, °;
6	β	= line search multiplication term;
7	γ	= wave propagation angle, °;
8	ϵ	= tolerance constant;
9	ζ	= decision tree splitting hyperparameter;
10	ρ_s	= bulk slide density, kg/m ³ ;
11	ρ_w	= water density, kg/m ³ ;
12	σ	= activation function;
13	Φ	= expression considering the slide front angle;
14	ϕ	= slide front angle, °.

15

16 *Subscript*

17	$\mathbf{f}$	= related to forget gate;
18	$\mathbf{g}$	= related to gate value;
19	g	= grain;
20	$\mathbf{h}$	= related to memory state;
21	i	= index to denote the value at a specific training example;
22	$\mathbf{i}$	= related to input gate;
23	$\mathbf{I}$	= input;

1	K	= keys;
2	M	= maximum;
3	$\mathbf{o}$	= related to output gate;
4	q	= row of weight matrix;
5	$\mathbf{r}$	= related to GRU gate;
6	s	= slide;
7	T_e	= testing;
8	T_o	= total;
9	(t)	= step in the sequence of frame;
10	u_{l-1}	= number of units in previous ANN layer;
11	w	= water;
12	$\mathbf{X}_I$	= input features;
13	$\mathbf{x}$	= input sequence for sequence model;
14	y	= output targets;
15	$\mathbf{z}$	= related to control gate.
16		
17	<i>Superscript</i>	
18	l	= layer of neural network;
19	N	= number of DTs in a GB model;
20	T	= transpose of a matrix;
21	$*$	= optimal.

1 Abbreviations

2	ANN	= artificial neural network;
3	DT	= decision tree;
4	FC	= fully connected;
5	GB	= gradient boosting;
6	GPU	= graphics processing unit;
7	GRU	= gated recurrent unit;
8	LSTM	= long short-term memory;
9	MAE	= mean absolute error;
10	MAPE	= mean absolute percentage error;
11	ML	= machine learning;
12	MSE	= mean squared error;
13	MSLE	= mean squared log error;
14	ReLU	= rectified linear activation function;
15	RMSprop	= root mean squared propagation;
16	RNN	= recurrent neural network;
17	SLT	= subaerial landslide-tsunami;
18	TT	= TabTransformer;
19	2D	= two-dimensional;
20	3D	= three-dimensional.

1 Acknowledgements

2 The authors would like to thank Dr Andreas Huber for providing ex-
3 perimental data of granular slides in a 3D water body geometry. The re-
4 maining laboratory data of VH were measured in projects supported by the
5 Swiss National Science Foundation (Grant No. 200020-103480/1), an Im-
6 perial College London Research Fellowship (cohort 2011), and the Natural
7 Environment Research Council (Grant No. NE/K000578/1). The position
8 of DGJ is funded by the Department of Physics and Mathematics, School of
9 Science and Technology, Nottingham Trent University.

10 References

- 11 Abadie, S.M., Harris, J.C., Grilli, S.T., Fabre, R., 2012. Numerical modeling
12 of tsunami waves generated by the flank collapse of the Cumbre Vieja
13 Volcano (La Palma, Canary Islands): Tsunami source and near field
14 effects. *Journal of Geophysical Research*, 117, C05030.
- 15 Alizadeh, M.J., Nourani, V., 2024. Multivariate GRU and LSTM models for
16 wave forecasting and hindcasting in the southern Caspian Sea. *Ocean*
17 *Engineering*, 298, 117193.
- 18 Bullard, G.K., Mulligan, R.P., Carreira, A., Take, W.A., 2019. Experimental
19 analysis of tsunamis generated by the impact of landslides with high
20 mobility. *Coastal Engineering*, 152, 103538.
- 21 Bohoran, T.A., Kampaktsis, P.N., McLaughlin, L., Leb, J., Moustakidis, S.,
22 McCann, G.P., Giannakidis, A., 2023a. Embracing uncertainty flexibil-
23 ity: Harnessing a supervised tree kernel to empower ensemble modelling
24 for 2D echocardiography-based prediction of right ventricular volume.
25 *Proceedings of the 16th International Conference of Machine Vision*
26 *(ICMV 2023)*, Yerevan, Armenia.
- 27 Bohoran, T.A., Kampaktsis, P.N., McLaughlin, L., Leb, J., Moustakidis,
28 S., McCann, G.P., Giannakidis, A., 2023b. Right Ventricular Volume
29 Prediction by Feature Tokenizer Transformer-Based Regression of 2D
30 Echocardiography Small-Scale Tabular Data. In: Bernard, O., Clarysse,
31 P., Duchateau, N., Ohayon, J., Viallon, M. (eds) *Functional Imaging and*
32 *Modeling of the Heart (FIMH 2023)*, June 19-22, Lyon, France. *Lecture*
33 *Notes in Computer Science*, 13958, 292–300.
- 34 Bohoran, T.A., Parke, K.S., Graham-Brown, M.P.M., Meisuria, M., Singh,
35 A., Wormleighton, J., Adlam D., Gopalan, D., Davies, M.J., Williams,

- 1 B., Brown, M., McCann, G.P., Giannakidis A., 2023c. Resource efficient
2 aortic distensibility calculation by end to end spatiotemporal learning of
3 aortic lumen from multicentre multivendor multidisease CMR images.
4 Scientific Reports, 13, 21794.
- 5 Chen, F., Heller, V., Briganti, R., 2020. Numerical modelling of tsunamis
6 generated by iceberg calving validated with large-scale laboratory ex-
7 periments. Advances in Water Resources, 142, 103647.
- 8 Chen, S., Shi, A., Xu, W., Yan, L., Wang, H., Tian, L., Xie, W.C., 2023. Nu-
9 merical investigation of landslide-induced waves: a case study of Wangji-
10 ashan landslide in Baihetan reservoir, China. Bull. Eng. Geol. Environ.
11 82, 110.
- 12 Cho, K., Merriënboer, B., Bahdanau, D., Bengio, Y., 2014. On the properties
13 of neural machine translation: Encoder–decoder approaches. Comput-
14 ing Research Repository, arXiv:1409.1259.
- 15 Davidson, D.D., Whalin, R.W., 1974. Potential landslide-generated water
16 waves, Libby Dam and Lake Koocanusa, Montana. U.S. Army Engi-
17 neering Waterways Experiment Station. Corps of Engineers, Vicksburg.
18 Technical Report, H-74-15.
- 19 Dorogush, A.V., Ershov, V., Gulin, A., 2018. CatBoost: gradient boost-
20 ing with categorical features support. Computing Research Repository,
21 arXiv:1810.11363.
- 22 Evers, F.M., Boes, R.M., 2019. Impulse wave runup on steep to vertical
23 slopes. Journal of Marine Science and Engineering, 7(1), 8.
- 24 Evers, F.M., Hager, W.H., 2016. Spatial impulse waves: Wave height decay
25 experiments at laboratory scale. Landslides, 13, 1395–1403.
- 26 Evers, F.M., Heller, V., Fuchs, H., Hager, W.H., Boes, R.M., 2019. Landslide
27 generated impulse waves in reservoirs – Basics and Computation, 2nd
28 edition. VAW Mitteilung, 254, ETH Zurich, Zurich.
- 29 Feng, X., Cheng, L., Dong, Q., Qi, X., Xiong, C., 2022. Numerical study of
30 hydraulic characteristics of impulse waves generated by subaerial land-
31 slides. AIP Advances 12, 121518.
- 32 Friedman, J.H., 2001. Greedy function approximation: A gradient boosting
33 machine. The Annals of Statistics, 29(5), 1189-1232.
- 34 Fritz, H.M., Hager, W.H., Minor, H.-E., 2001. Lituya Bay case: Rockslide
35 impact and wave run-up. Science of Tsunami Hazards, 19(1), 3–22.
- 36 Fuchs, H., Boes, R., 2010. Berechnung felsrutschinduzierter Impuls-
37 wellen im Vierwaldstättersee. Wasser Energie Luft, 102(3), 215–221 (in German).

- 1 Fuchs, H., Pfister, M., Boes, R., Perzlmaier, S., Reindl, R., 2011. Im-
2 pulswellen infolge Lawineneinstoss in den Speicher Kühtai. Wasser-
3 wirtschaft, 101(1–2), 54–60 (in German).
- 4 Gabl, R., Seibl, J., Gems, B., Aufleger, M., 2015. 3–D–numerical approach to
5 simulate the overtopping volume caused by an impulse wave comparable
6 to avalanche impact in a reservoir. Natural Hazards and Earth System
7 Sciences, 15(12), 2617–2630.
- 8 Gorishniy, R., Rubachev, I., Khrulkov, V., Babenko, A., 2023. Revisiting
9 deep learning models for tabular data. Computing Research Repository,
10 arXiv:2106.11959.
- 11 Gylfadóttir, S.S., Kim, J., Helgason, J.K., Brynjólfsson, S., Höskuldsson,
12 A., Jóhannesson, T., Harbitz, C.B., Løvholt, F., 2017. The 2014 Lake
13 Askja rockslide induced tsunami: Optimization of numerical tsunami
14 model using observed data. Journal of Geophysical Research: Oceans,
15 122(5), 4110–4122.
- 16 Hall, J.V., Watts, G.M., 1953. Laboratory investigation of the vertical rise of
17 solitary waves on impermeable slopes. Technical Memo Report No. 33,
18 United States Army Corps of Engineers, Beach Erosion Board: Wash-
19 ington, D.C., USA.
- 20 Muhari, A., Heidarzadeh, M., Susmoro, H., Nugroho, H.D., Kriswati, E., Su-
21 partoyo, Wijanarto, A.B., Imamura, F., Arikawa, T., 2019. The Decem-
22 ber 2018 Anak Krakatau Volcano tsunami as inferred from post-tsunami
23 field surveys and spectral analysis. Pure and Applied Geophysics, 176,
24 5219–5233.
- 25 Heidarzadeh, M., Putra, P.S., Nugroho, S.H., 2020. Field survey of tsunami
26 heights and runups following the 22 December 2018 Anak Krakatau Vol-
27 cano Tsunami, Indonesia. Pure and Applied Geophysics, 177, 4577–4595.
- 28 Heller, V., 2007. Landslide generated impulse waves – Prediction of near
29 field characteristics. Ph.D. Thesis 17,531, ETH Zurich, Zurich.
- 30 Heller, V., Hager, W.H., Minor, H.-E., 2008. Scale effects in subaerial land-
31 slide generated impulse waves. Experiments in Fluids, 44(5), 691–703.
- 32 Heller, V., Hager, W.H., 2010. Impulse product parameter in landslide gen-
33 erated impulse waves. Journal of Waterway, Port, Coastal, and Ocean
34 Engineering, 136(3), 145–155.
- 35 Heller, V., Hager, W.H., 2011. Wave types of landslide generated impulse
36 waves. Ocean Engineering, 38(4), 630–640.
- 37 Heller, V., Ruffini, G., 2023. A critical review about generic subaerial land-
38 slide–tsunami experiments and options for a needed step change. Earth-

1 Science Reviews, 242, 104459.

2 Heller, V., Spinneken, J., 2013. Improved landslide–tsunami prediction: Ef-
3 fects of block model parameters and slide model. *Journal of Geophysical*
4 *Research: Oceans*, 118(3), 1489–1507.

5 Heller, V., Spinneken, J., 2015. On the effect of the water body geometry
6 on landslide–tsunamis: Physical insight from laboratory tests and 2D to
7 3D wave parameter transformation. *Coastal Engineering*, 104, 113–134.

8 Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. *Neural*
9 *Computation*, 9(8), 1735–1780.

10 Huang, X., Khetan, A., Cvitkovic, M., Karnin, K.S., 2020. TabTransformer:
11 Tabular data modeling using contextual embeddings. *Computing Re-*
12 *search Repository*, arXiv:2012.06678.

13 Huber, A., 1980. Schwallwellen in Seen als Folge von Felsstürzen. Ph.D.
14 Thesis 6,627, ETH Zurich, Zurich (in German).

15 Huber, A., Hager, W.H., 1997. Forecasting impulse waves in reservoirs.
16 *Proceedings 19th Congrès des Grandes Barrages*, Florence, International
17 *Snow Science Workshop Grenoble*, Paris, 993–1005.

18 Lindstrøm, E.K., Pedersen, G.K., Jensen, A., Glimsdal, S., 2014. Experi-
19 ments on slide generated waves in a 1:500 scale fjord model. *Coastal*
20 *Engineering*, 92, 12–23.

21 Longuet-Higgins, M., 1967. On the trapping of wave energy round islands.
22 *Journal of Fluid Mechanics*, 29(4), 781–821.

23 Løvholt, F., Pedersen, G., Gisler, G., 2008. Oceanic propagation of a poten-
24 tial tsunami from the La Palma Island. *Journal of Geophysical Research*,
25 113, C09026.

26 Lüthi, M.P., Vieli, A., 2016. Multi–method observation and analysis of a
27 tsunami caused by glacier calving. *The Cryosphere* 10(3), 995–1002.

28 Lystbaek, M., Giannakidis, A., Beliatas, M.J., Olsen, M., 2023. Removing
29 unwanted text from architectural images with multi-scale deformable
30 attention-based machine learning. *Proceedings of the IEEE Interna-*
31 *tional Conference on Imaging Systems and Techniques*, Copenhagen,
32 Denmark, 1–5.

33 Meng, Z., Hu, Y., Ancey, C., 2020. Using a data driven approach to predict
34 waves generated by gravity driven mass flows. *Water*, 12, 600.

35 Minuzzi, F.C., Farina, L., 2024. Artificial neural networks ensemble method-
36 ology to predict significant wave height. *Ocean Engineering*, 300, 117479.

37 Panizzo, A., De Girolamo, P., Petaccia, A., 2005. Forecasting impulse waves
38 generated by subaerial landslides. *Journal of Geophysical Research*, 110,

1 C12.

2 Raissi, M., Perdikaris, P., Karniadakis, G.E., 2019. Physics-informed neural
3 networks: A deep learning framework for solving forward and inverse
4 problems involving nonlinear partial differential equations. *Journal of*
5 *Computational Physics*, 378, 686-707.

6 Rauter, M., Viroulet, S., Gylfadóttir, S.S., Fellin, W., Løvholt, F., 2022.
7 Granular porous landslide tsunami modelling – The 2014 Lake Askja
8 flank collapse. *Nature Communications*, 13(678), 1–13.

9 Roberts, N.J., McKillop, R.J., Lawrence, M.S., Psutka, J.F., Clague, J.J.,
10 Brideau, M.-A., Ward, B.C., 2013. Impacts of the 2007 Landslide-
11 Generated Tsunami in Chehalis Lake, Canada. *Landslide Science and*
12 *Practice 6* (C. Margottini, P. Canuti, K. Sassa (editors)), 133–140.

13 Rosenblatt, F., 1958. The perceptron: A probabilistic model for information
14 storage and organization in the brain. *Psychological Review*, 65(6),
15 386–408.

16 Ruffini, G., Heller, V., Briganti, R., 2021. Numerical characterisation and
17 efficient prediction of landslide–tsunami propagation over a wide range
18 of idealised bathymetries. *Coastal Engineering*, 167, 103854.

19 Sabeti, R., Heidarzadeh, M., 2022. Numerical simulations of water waves
20 generated by subaerial granular and solid-block landslides: Validation,
21 comparison, and predictive equations. *Ocean Engineering*, 266(3), 112853.

22 Sak, H., Senior, A.W., Beaufays, F., 2014. Long short-term memory re-
23 current neural network architectures for large scale acoustic modeling.
24 *Interspeech conference*, 338–342.

25 Savage, S.B., Hutter, K., 1989. The motion of a finite mass of granular
26 material down a rough incline. *Journal of Fluid Mechanics*, 199, 177-
27 215.

28 Schwartz-Ziv, R., Armon, A., 2022. Tabular data: Deep learning is not all
29 you need. *Information Fusion*, 81, 84-90.

30 Slingerland, R.L., Voight, B., 1979. Occurrences, properties, and predic-
31 tive models of landslide-generated water waves. In: Voight, B. (Ed.),
32 *Rockslides and Avalanches 2*. Elsevier, Amsterdamn, 317–394.

33 Tan, H., Ruffini, G., Chen, S., Heller, V., 2018. A numerical landslide–tsunami
34 hazard assessment technique applied on hypothetical scenarios at Es
35 Vedrà, offshore Ibiza. *Journal of Marine Science and Engineering*, 6(4),
36 582–601.

37 Tang, G., Lu, L., Teng, Y., Zhang, Z., Xie, Z., 2018. Impulse waves generated
38 by subaerial landslides of combined block mass and granular material.

1 Coastal Engineering, 141(11), 68–85.

2 Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N.,
3 Kaiser, L., Polosukhin, I., 2017. Attention is all you need. Computing
4 Research Repository, arXiv:1706.03762.

5 Voight, B., Glicken, H., Janda, R.J., Douglass, P.M., 1981. Catastrophic
6 rockslide avalanche of May 18. In (Lipman PW, Millineaux DL eds)
7 The 1980 Eruptions of Mount St Helens, Washington, Professional Paper
8 1250. U.S. Geological Survey, Reston, 347–378.

9 Wang, W., Chen, G., Yin, K., Wang, Y., Zhou, S., Lui, Y., 2016. Modeling
10 of landslide generated impulsive waves considering complex topography
11 in reservoir area. Environmental Earth Sciences, 75, 1–15.

12 Watt, S.F.L., Talling, P.J., Vardy, M.E., Heller, V., Hühnerbach, V., Urlaub,
13 M., Sarkar, S., Masson, D.G., Henstock, T.J., Minshull, T.A., Paulatto,
14 M., Le Friant, A., Lebas, E., Berndt, C., Crutchley, G.J., Karstens,
15 J., Stinton, A.J., Maeno, F., 2012. Combinations of volcanic-flank and
16 seafloor-sediment failure offshore Montserrat, and their implications for
17 tsunami generation. Earth and Planetary Science Letters, 319–320,
18 228–240.

19 WCHL, 1970. Hydraulic model studies – Wave action generated by slides
20 into Mica Reservoir – British Columbia. Western Canada Hydraulic
21 Laboratories, Vancouver report.

22 Wilcoxon, F., 1992. Individual comparisons by ranking methods. In break-
23 throughs in statistics: Methodology and distribution. New York, NY:
24 Springer New York, 196–202.

25 Wu, Y., Shao, K., Piccialli, F., Mei, G., 2022. Numerical modeling of the
26 propagation process of landslide surge using physics-informed deep learn-
27 ing. Advanced Modeling and Simulation in Engineering Sciences, 9, 14.

28 Xue, H., Ma, Q., Diao, M., Jiang, L., 2019. Propagation characteristics
29 of subaerial landslide-generated impulse waves. Environmental Fluid
30 Mechanics, 19(1), 203–230.

31 Zaremba, W., Sutskever, I., Vinyals, O., 2015. Recurrent neural network
32 regularization. Computing Research Repository, arXiv:1409.2329.

33 Zitti, G., Ancey, C., Postacchini, M., Brocchini, M., 2016. Impulse waves
34 generated by snow avalanches: Momentum and energy transfer to a
35 water body. Journal of Geophysical Research: Earth Surface, 121(12),
36 2399–2423.