Surrogate modelling of dike wave overtopping hydrodynamics using an adapted deep learning Vision Transformer

Schrama, Wouter P.; van Bergeijk, Vera M.; Mares-Nasarre, Patricia; den Bieman, Joost P.; van Gent, Marcel R.A.; Aguilar-López, Juan P.

DOI
10.1016/j.coastaleng.2025.104874

Publication date
2025

Document Version
Final published version

Published in
Coastal Engineering

Citation (APA)
Schrama, W. P., van Bergeijk, V. M., Mares-Nasarre, P., den Bieman, J. P., van Gent, M. R. A., & Aguilar-López, J. P. (2025). Surrogate modelling of dike wave overtopping hydrodynamics using an adapted deep learning Vision Transformer. *Coastal Engineering*, *203*, Article 104874. https://doi.org/10.1016/j.coastaleng.2025.104874

Coastal Engineering 203 (2026) 104874

Contents lists available at ScienceDirect

Coastal Engineering

journal homepage: www.elsevier.com/locate/coastaleng

Surrogate modelling of dike wave overtopping hydrodynamics using an adapted deep learning Vision Transformer

Wouter P. Schrama [a], Vera M. van Bergeijk [b], Patricia Mares-Nasarre [a], Joost P. den Bieman [b], Marcel R.A. van Gent [a,b], Juan P. Aguilar-López [a,*]

[a] Delft University of Technology, Department of Hydraulic Engineering, Stevinweg 1, Delft, 2628 CN, The Netherlands
[b] Deltares, Department of Coastal Structures and Waves, Boussinesqweg 1, Delft, 2629 HV, The Netherlands

ARTICLE INFO

Keywords:
Overtopping flow
Surrogate modelling
Deep learning
Vision Transformer Image to Image
Hydrodynamic modelling
Dike safety assessment
Wave overtopping simulations

ABSTRACT

Sea level rise can compromise the safety of coastal flood defences, as wave overtopping events are becoming more frequent and severe. This increasing threat emphasizes the need for accurate assessment of wave overtopping hydrodynamics over dikes, which is essential for evaluating flood safety. The currently available methods do not combine computational efficiency, detailed results and general applicability, which limits their use in modelling wave overtopping and the resulting dike erosion. To address these limitations, this study introduces the Wave Overtopping Surrogate Model (WOSM), a novel method for rapidly generating high-quality two-dimensional simulations of wave overtopping over the dike crest and landward slope. The foundation of the WOSM is the Vision Transformer Image to Image (ViTI2I), a new deep learning model that combines an adapted Vision Transformer with a convolutional decoder for next-frame prediction. Trained on CFD wave overtopping simulations, the WOSM accurately reproduces the overtopping hydrodynamics such as flow velocities, water depths, overtopping duration and vertical velocity profiles, including both spatial and temporal variations. The scope of the training data limits the applicability of the WOSM and its ability to consistently capture complex phenomena such as flow separation and reattachment, both of which could be improved by enriching the dataset. Its low computational demand makes it suitable for exploring additional applications, such as probabilistic design or simulating wave overtopping with evolving dike profiles for erosion assessment. Additionally, this study serves as a proof of concept that the WOSM framework could benefit other fields encountering comparable modelling constraints.

1. Introduction

A common form of coastal flood protection is the grass-covered earthen dike (Sharp et al., 2013). These coastal defence structures are expected to face increasingly extreme storm conditions due to climate change (IPCC, 2019). One of the consequences is sea level rise, which in turn can lead to more severe wave overtopping (Koosheh et al., 2024). Wave overtopping occurs when a wave runs up the seaward slope of a dike and exceeds the crest height, resulting in a flow over the crest and inner slope of the dike. This flow can cause erosion of the inner slope, which can ultimately result in dike failure (Van Bergeijk et al., 2022). The external erosion caused by wave overtopping has been identified as one of the main causes of grass-covered dike breaches (Danka and Zhang, 2015; Van Bergeijk et al., 2021). Therefore, to better consider the effects of climate change in the design of coastal dikes, it becomes increasingly important to accurately estimate dike safety against wave overtopping.

Current methods to estimate dike safety rely on empirical formulas or numerical models, such as commonly used computational fluid dynamics (CFD) software. Earlier research led to formulas that describe water depth and wave velocity along the dike caused by wave overtopping (Schüttrumpf, 2001; Schüttrumpf and Van Gent, 2004; Schüttrumpf and Oumeraci, 2005). However, these formulas do not account for complex geometries, such as eroded dike profiles or varying slope angles. Other formulas are proposed by Van Gent (2002) and Van Bergeijk et al. (2019) that include changes of the slope angle, but these are still simplified representations of flow velocity and water depth and provide estimations only of maximum values along the dike profile, rather than a detailed insight into wave hydrodynamics in space and time. Numerical models that solve the Reynolds-averaged Navier–Stokes equations (RANS) offer detailed simulations of wave overtopping and are suitable for a wide range of dike profiles (Van Bergeijk, 2022). Bomers et al. (2018) developed a CFD model that simulates

* Corresponding author.
E-mail address: j.p.aguilarlopez@tudelft.nl (J.P. Aguilar-López).

https://doi.org/10.1016/j.coastaleng.2025.104874
Received 13 June 2025; Received in revised form 4 September 2025; Accepted 4 September 2025
Available online 12 September 2025

an individual overtopping wave over the crest and the upper part of the inner slope of a dike, and Van Bergeijk et al. (2020) simulated wave overtopping over the dike from crest to the inner toe. These models offer detailed simulations of the full evolution of an overtopping wave, but preparing them—defining the mesh and boundary conditions—is labour-intensive and running a simulation can be time-consuming (Suzuki et al., 2017). Simpler models exist that use the Non-Linear Shallow Water equations (NLSW). Although these models require substantially lower computational demand compared to RANS models, they provide less detailed results when vertical flow components become significant, as their depth-averaged approach limits resolution in the vertical direction (Van Bergeijk, 2022). Other models that describe wave-structure interaction include CSHORE (Johnson et al., 2012), which focuses on describing erosion and returns depth-averaged hydrodynamic variables, and SBEACH (Larson and Kraus, 1989), which calculates erosion caused by storm waves and water levels but does not provide output for hydrodynamic variables. While these models are computationally efficient and effective for erosion prediction, they do not simulate overtopping wave hydrodynamics in the spatiotemporal domain.

The currently available modelling approaches impose limitations on dike safety assessments, as none can produce detailed spatiotemporal simulations while maintaining low computational demand. This results in the use of simplified scenarios to estimate erosion depth caused by wave overtopping during storm events, as current methods do not account for progressive erosion. This means that the dike profile remains unchanged throughout the dike safety assessment, neglecting the erosion process evolution over time. However, erosion occurs gradually as overtopping waves begin to create small erosion holes during storm conditions (Aguilar-López et al., 2018b; Van Bergeijk, 2022). These irregularities can influence subsequent waves and increase their turbulence, which can enhance their erosion potential (Aguilar-López et al., 2018b). Existing formulas do not account for these effects, and numerical models would require remeshing of the domain after every simulated wave, with the risk of becoming unstable for complex geometries.

Additionally, these limitations make it difficult to perform accurate sensitivity analyses on how uncertainty or variations in input variables—such as inflow boundary conditions or design choices—influence output like maximum load or erosion depth. For improved coastal dike design, current methods must also become feasible for probabilistic design, which requires testing a system under many scenarios based on probability distributions of the input variables. Fitted formulas may not be valid for this range of input variables, and numerical models are too computationally expensive to complete the required number of simulations. While several models exist for simulating wave overtopping, none provide detailed results within affordable computational times. This leaves a gap in methods that can explore the incorporation of erosion during wave overtopping simulations or enable probabilistic design.

One promising solution is the use of surrogate modelling, which focuses on the development of a computationally cheap model that emulates a more complex one (Forrester et al., 2008; Razavi et al., 2012). This offers powerful possibilities, especially with the current rise of increasingly advanced deep learning models. Deep learning techniques can learn and recognize complex, non-linear patterns, like those within fluid simulations (Pant et al., 2021). For example, Aguilar-López et al. (2018a,b) used the output of numerical models to train an artificial neural network (ANN) and build an emulator for the failure probability estimation of flood defences, Buscombe et al. (2020) used a deep learning model to estimate wave height and period from images of waves, and physical model test results are used to train deep learning and machine learning models that predict wave overtopping discharges over various coastal structures (Van Gent et al., 2007; Den Bieman et al., 2020, 2021).

While these aim to predict a singular numerical value, other researches apply deep learning models that can predict a two-dimensional grid (i.e. an image). One specific example is the next-frame prediction model, which involves the prediction of a next frame based on one or more previous-step input frames. Convolutional neural network (CNN)-based models are trained on numerical model output, using preceding conditions to predict two-dimensional instantaneous velocity-fluctuation fields (Guastoni et al., 2021), waves and hydrodynamics in the nearshore zone (Wei and Davison, 2022), and the future boundary of natural lakes (Yin et al., 2023). Although significant research has been conducted using similar models, there appears to be limited research focused on generating full simulations, i.e. two-dimensional grids evolving over time. A model with such capabilities could help overcome the limitations of currently available methods and is becoming increasingly feasible due to the continuous advancements in deep learning architectures, such as the state-of-the-art Vision Transformer model (Dosovitskiy et al., 2020).

This study focuses on the first step in addressing the broader challenges in modelling wave overtopping. It proposes a new method to create detailed spatiotemporal simulations of wave overtopping at significantly lower computational demand. The objective is to develop this model using deep learning algorithms and to validate its performance by comparing the simulated hydrodynamics. Although the motivation behind this model includes potential applications in erosion modelling with evolving dike profiles and probabilistic design, these are beyond the scope of the current work. This study is limited to simulating the hydrodynamics of an overtopping wave over a fixed dike profile. It demonstrates the feasibility of a simulation approach that produces detailed spatiotemporal results while requiring less computational effort. The key contribution is the Wave Overtopping Surrogate Model (WOSM), a surrogate model for the fast and accurate simulation of wave overtopping. The WOSM is based on a novel deep learning next-frame prediction model, called the Vision Transformer Image to Image (ViTI2I). The ViTI2I is developed by adapting the Vision Transformer (ViT), whose backbone is a transformer architecture that has proven superior in capturing both temporal and spatial patterns (Bertasius et al., 2021; Yan et al., 2021; Ye and Bilodeau, 2022). The ViTI2I is trained on a wave overtopping simulation dataset previously produced with a CFD model (Van Bergeijk et al., 2022). Various configurations of the WOSM are tested and evaluated to identify the best-performing setup. The optimal WOSM is validated against the original CFD simulations by comparing the modelled hydrodynamics.

The paper is structured as follows. Section 2 provides background information on deep learning terminology, relevant for the developed ViTI2I. In Section 3, the dataset and ViTI2I are presented and the implementation of the ViTI2I for generating simulations is described, which completes the WOSM. In Section 4, the performance of different WOSM setups is evaluated and the best-performing WOSM is validated. The WOSM performance, capabilities and limitations are discussed in Section 5. Finally, conclusions and recommendations are provided in Section 6.

2. Deep learning background

The key concept of deep learning models is to identify and replicate patterns found in data (Goodfellow et al., 2016). During training, the model adjusts its trainable parameters using a training dataset to produce predictions that resemble the original data. Once trained, it offers a faster, data-driven alternative to the traditional model, as it does not rely on a numerical solving scheme. This makes deep learning well-suited for efficiently approximating and reproducing spatiotemporal processes.

A common approach to spatiotemporal tasks is the Convolutional Long Short-Term Memory (ConvLSTM) model (Shi et al., 2015). This model uses the LSTM to distinguish and remember relevant temporal patterns, and applies convolutional layers to identify spatial patterns.

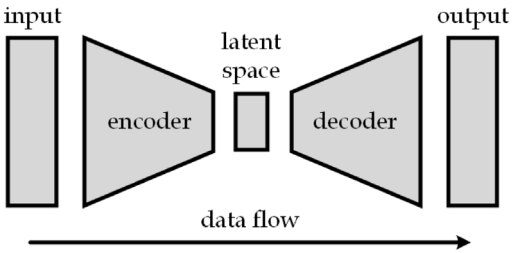

Fig. 1. A simple overview on data flow through an encoder–decoder mechanism. The raw input data is processed by the encoder, leading to the latent space. The latent space holds an abstract representation of the observed patterns in the input data. The decoder converts this into the predicted output.

Convolutional layers slide various kernels—small numerical matrices—over the data and project the identified patterns onto a new grid, that contains an abstract representation of all the observed patterns (Prince, 2023). The reversed use of such convolutional layers is called deconvolution, where kernels are used to translate complex patterns back into the desired format and resolution, such as an image (Prince, 2023). Combining convolution and deconvolution leads to an encoder–decoder mechanism (see Fig. 1). The encoder extracts patterns from the input data and converts it into the latent space, which is the abstract collection of all the observed patterns. The decoder interprets the latent space to generate a prediction. Although widely used, ConvLSTM models are limited by their inherent focus on local patterns and struggle with capturing patterns that are distant in space and time (Masafu and Williams, 2024; Wang et al., 2024).

To address this, Vaswani et al. (2017) introduced the transformer architecture, which applies the multi-head attention mechanism to capture patterns, regardless of the distance between them. Dosovitskiy et al. (2020) adapted this mechanism to spatial data with the Vision Transformer (ViT), which splits two-dimensional data into smaller patches and embeds each patch into a vector, with a size equal to the number of embedding dimensions. These patch embeddings represent local information and are processed using multi-head attention, which compares every pair of patches to determine whether one contains information that the other requires, thereby enabling patterns to be identified globally across the input. The ViT does not inherently understand the spatial position of a patch within the input, but learns this by adding a spatial embedding to each patch that encodes the location of each patch and specifies their relative position. After extracting both local and global patterns, the ViT uses a multilayer perceptron (MLP) to classify the output. The MLP consists of an input layer, one or more hidden layers, and an output layer, with nonlinear transformations enabling the network to learn complex relationships and map inputs to desired outputs. For a complete explanation of the ViT, see Dosovitskiy et al. (2020).

Training a ViT typically requires large datasets and considerable computational resources compared to convolutional-based models (Han et al., 2022). These requirements may limit its applicability in the hydraulic engineering domain, which can include scarce datasets. Large input sizes can increase the number of patches, which leads to more computations in the multi-head attention mechanism. This can constrain its use to smaller spatial domains. Nevertheless, the improved understanding of long-range patterns of the ViT can become crucial for fluid simulations when changes in fluid flow in one area start to affect the fluid in another region. For instance, water that begins to accelerate on the slope can have an accelerating effect on water downstream. These types of distant patterns may be best recognized and replicated by a transformer-based architecture.

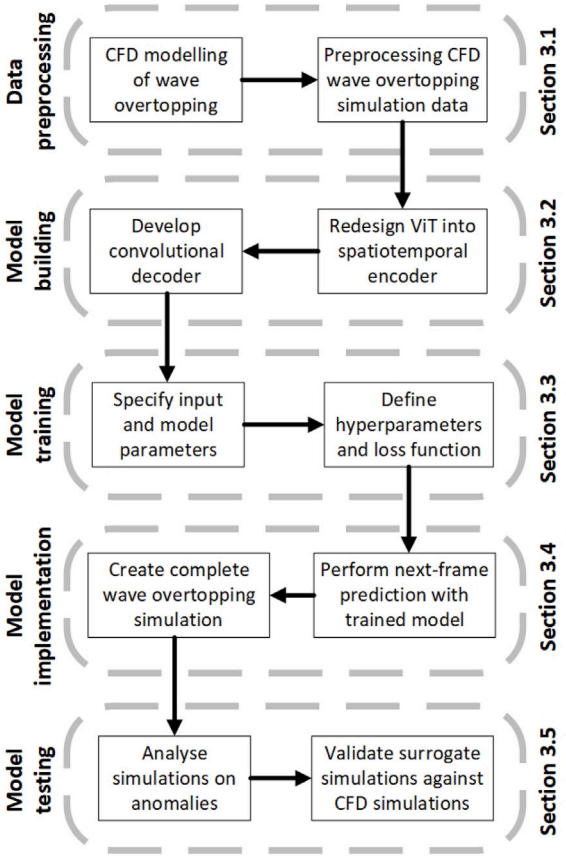

Fig. 2. Flowchart of the methodology.

3. Methodology

This section presents the building, training, testing and validation of the WOSM. The steps and main components are explained in the order presented in the flowchart in Fig. 2. The training dataset is described and preprocessed to suit the training of a deep learning model (Section 3.1). Next, the existing deep learning architectures ViT and CNN are modified and combined to be able to process the available spatiotemporal data and predict the hydrodynamics of an overtopping wave, resulting in the ViTI2I (Section 3.2). Section 3.3 describes the training of the ViTI2I, which leads to a model capable of predicting one future time step. In Section 3.4, the ViTI2I is applied to generate simulations that consist of multiple time steps, which completes the WOSM framework. Finally, Section 3.5 presents the evaluation and validation of the WOSM.

3.1. Data preprocessing

3.1.1. CFD model dataset

The dataset used for this research was already produced with the CFD software OpenFOAM by Van Bergeijk et al. (2022). The numerical model solves the two-phase Reynolds-averaged Navier–Stokes equations using the finite volume method:

$$\nabla \cdot \bar{u} = 0 \tag{1}$$

$$\frac{\partial \rho \bar{u}}{\partial t} + \rho(\bar{u} \cdot \nabla)\bar{u} = \nabla(\mu \nabla \cdot \bar{u}) - \nabla p + F \tag{2}$$

where $\bar{u}$ [m/s] denotes the velocity vector, ρ [kg/m^3] the fluid density, μ [kg/(m s)] the dynamic viscosity, p [Pa] the pressure, and F the external forces. It models water and air with the volume of fluid method

and turbulence with the $k-\omega$ SST turbulence model. This turbulence model is a hybrid model, switching between the $k-\omega$ model in the near-wall region and the $k-\epsilon$ model in the free-stream region. The computational domain represents a two-dimensional side view of the dike and includes the crest, the landward slope, and a small part beyond the toe (see upper panel in Fig. 3). The mesh resolution is 1 cm x 1 cm. The dike is modelled as impermeable, and roughness is calibrated using a Nikuradse roughness height of 8 mm. The model is built to simulate individual overtopping waves that have already exceeded the crest height and are flowing over the crest, thus it does not include incoming waves or wave run-up. The boundary conditions at the crest depend on the overtopping volume V [m^3/m], and are defined in terms of flow velocity u_0 [m/s] and layer thickness h_0 [m], following the formulations by Van Der Meer et al. (2010) and Hughes (2011).

$$u_0(t) = 5.00\,V^{0.34}\left(1-\frac{t}{T_0}\right) \tag{3}$$

$$h_0(t) = 0.133\,V^{0.5}\left(1-\frac{t}{T_0}\right). \tag{4}$$

where t [s] is time and T_0 [s] is the overtopping period, expressed as:

$$T_0 = 4.4\,V^{0.3} \tag{5}$$

The model has been validated against two field tests performed in The Netherlands and showed good performance across various velocity components (Van Bergeijk et al., 2020). For a more detailed description of the OpenFOAM setup and governing equations, see Van Bergeijk et al. (2020).

The production of the dataset used for this research is fully described in Van Bergeijk et al. (2022). It consists of 40 individual simulations of a single overtopping wave, each showing 10 seconds of wave overtopping with time steps of 0.01 s. All simulations feature a dike with a horizontal crest, a straight inner slope without a berm, and an inner toe. Each dike geometry has a crest width of 3 metres, a slope length of 6 metres, and 1 metre area behind the toe (see upper panel in Fig. 3). The 40 simulations cover a range of overtopping volumes V and slope steepnesses s [-], where each individual simulation shows a single wave with a combination of V and s (see Table 1). Some combinations of V and s lead to flow separation, which means the flow detaches from the dike bed at the location of transition from crest to slope and reattaches further along the slope (see subplot B in Fig. 3).

3.1.2. Modifying CFD output

The WOSM focuses solely on the prediction of water depth and flow velocity. Other variables that are produced by the CFD model, like pressures, stresses and turbulence-related fields, are not included in the framework, but are discussed in Section 5.2. The CFD dataset includes values for water fraction and flow velocity for every cell in the mesh, with cell size 1 cm x 1 cm (Van Bergeijk et al., 2022). These values are interpolated in a two-dimensional grid with cells of 1 cm x 1 cm, from which three types of data are generated as input for the ViTI2I: a binary dike mask, water velocity field and binary water mask. The preprocessed data is described below and presented in Fig. 3.

- **Binary dike mask (BDM).** Unlike CFD models, data-driven deep learning models do not inherently detect boundary conditions such as walls, inflow or open boundaries. Therefore, the dike geometry is added to the dataset as a binary mask to provide information on the dike profile. The BDM is created by giving every cell located in the dike a value of 1 and every cell located in the air a value of 0. Since the BDM values are already between 0 and 1, no normalization is needed.
- **Water velocity field.** The CFD dataset originally contains the velocity field for both air and water, but it is modified to show only the velocity of the water, represented by its magnitude U_{mag} and direction U_{dir}. This approach follows the assumption of Van Bergeijk et al. (2020) that cells with a calculated water fraction exceeding 60% are assumed to be filled with water, whereas cells with less than 60% are treated as empty. The data is normalized for training of the ViTI2I by applying the Max-Abs Scaling technique. This method preserves the sign of the original data and ensures all initial zero values remain zero. This maintains the physical meaning of the original data, as the sign represents the direction of the velocity and a zero velocity indicates that a cell is filled with air. As a result, values for U_{mag} fall within the range [0, 1] and these of U_{dir} within [0, 1]. The preprocessed data shows entrapped air bubbles, which are caused by the modified grid and the CFD model using a water fraction boundary of 0.8, resulting in an inflow of a water–air mixture (see subplot A in Fig. 3).

  $$\text{Max-Abs Scaling: } X' = \frac{X}{\max(|X|)} \tag{6}$$

- **Binary water mask (BWM).** The training dataset is completed by adding the BWM, which represents a simplified version of the water fraction. Cells that contain water have a value of 1 and cells that contain air have a value of 0. The BWM follows the same transformation where all initial water fraction values above 0.6 are changed to 1 and those below to 0. This leads to the same entrapped air bubbles as for the water velocity field. Without the BWM, the velocity fields represent two aspects: (1) whether water is present and (2) if water is present, the velocity value. Including the BWM in the dataset ensures that the model can focus on learning and predicting the presence and velocity of water in cells separately. No normalization is required for the BWM, as its values are already within the range of 0 to 1.

Table 1
Overview of the wave overtopping simulations in the training, validation and test datasets.

Slope steepness [–]	Training set Volume [m^3]	Validation set Volume [m^3]	Test set Volume [m^3]
1:1	0.25, 0.5, 1.5, 2.5	–	–
1:1.2	1.5	–	–
1:1.5	–	–	1.5
1:1.7	0.35, 0.75, 1.0, 1.5, 1.662, 1.75, 2.5, 3.5, 4.0	1.5, 2.0, 3.0	0.25, 0.5, 1.25
1:2	1.5, 2.5	–	–
1:2.2	1.0	1.25	1.5
1:2.5	1.5, 2.25	–	2.0
1:2.7	1.5	–	–
1:3	1.5, 2.0, 2.5, 3.0	4.0	–
1:3.5	–	1.5	–
1:4	1.5, 2.5	–	–
1:5	1.5, 2.5	–	–

The preprocessed dataset is divided into a training dataset (28 simulations), a validation dataset (6 simulations), and a test dataset (6 simulations) (see Table 1). This commonly used split ensures that a sufficient portion of the data is available for model training, while retaining enough data for tuning the model during validation and evaluating its performance on unseen data during testing. To create the widest range of training data, the simulations in the training dataset include overtopping volumes ranging from the smallest to the highest volume (0.25 - 4 m^3/m) and slope steepnesses ranging from the gentlest to the steepest slope (1:5 - 1:1).

3.2. Model building

A next-frame prediction model is developed that can process the two-dimensional preprocessed data, capable of predicting a future frame based on a sequence of previous frames. The new model, called the Vision Transformer Image to Image (ViTI2I), is an encoder–decoder

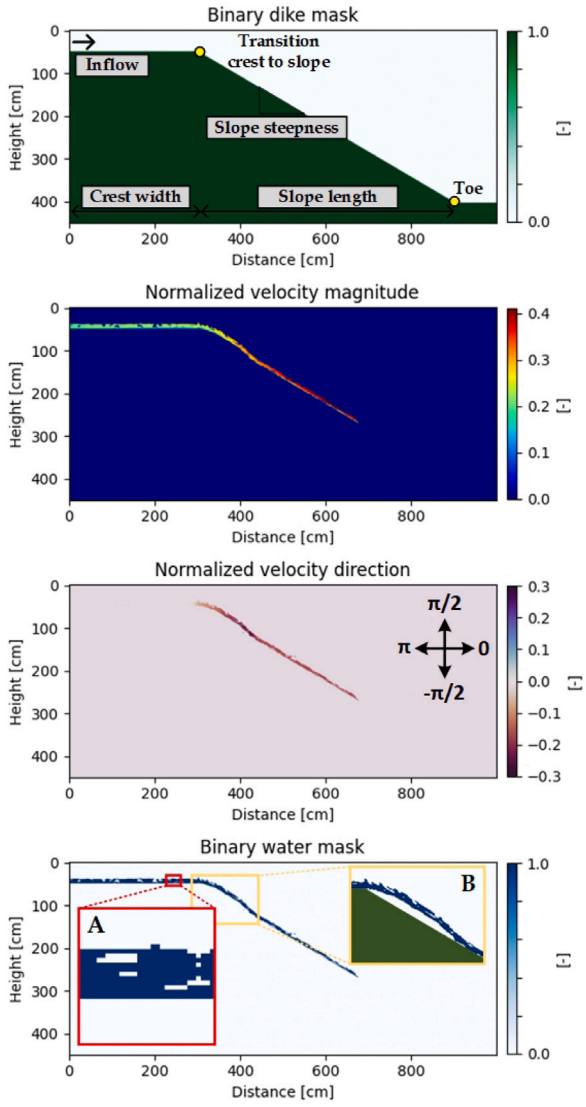

Fig. 3. A snapshot at t = 1.35 s of the preprocessed simulation of an overtopping wave with an overtopping volume of 1.25 m^3 over a dike with slope steepness 1:1.7. The location of the inflow boundary and definitions of dike geometry characteristics are given in the binary dike mask. No water is visible on the dike crest in the normalized velocity direction plot, because water flowing rightward has a velocity direction value close to zero. Consequently, although water is present, it is not visible in this specific plot. Subplot A shows entrapped air bubbles in the binary water mask. Subplot B shows flow separation, with the flow detaching at the transition point and reattaching around x = 420.

mechanism designed to learn spatiotemporal patterns. An overview of the data flow through the ViTI2I is given in Fig. 4. The ViTI2I takes as input data a sequence of frames. A frame is composed of all the two-dimensional grids that complete an individual time step (i.e. BDM, water velocity field and BWM). Each frame is divided into patches through patch embedding, and each patch gets assigned a spatial and temporal embedding. These spatial and temporal embeddings provide the ViTI2I knowledge on the location of a patch in space and time. In the Transformer Encoder, each patch is compared with all other patches and information is transferred globally over the grid, resulting in the transformed patches. These now contain global and local information, due to the patch embedding and the attention mechanism. This knowledge is projected into an abstract representation of future hydrodynamics by an MLP, leading to the future patch embeddings. The convolutional decoder is capable of understanding this abstract information and converting it into detailed predictions for the BWM and the velocity fields. The BDM is omitted from the prediction because the dike profile is assumed to remain static across consecutive time steps in these simulations.

For readers interested in the development of the ViTI2I and architecture characteristics, a detailed description is given in Sections 3.2.1 and 3.2.2. Background information and deep learning terminology is provided in Section 2.

3.2.1. Encoder mechanism

The encoder is based on the original ViT architecture, which applies patch embedding and the multi-head attention mechanism to capture local and global patterns within a single frame. This architecture is modified to process a sequence of frames, recognizing patterns over space and time and converting them into the latent space.

As the ViTI2I requires a sequence of frames, the shape of the ViT input data is modified by adding an extra dimension, which represents the time component. Every time step can hold multiple grids, which allows for the representation of several variables per time step. As a result, the shape of the input data is (time steps, variables, grid height, grid width). The input data is divided into patches, which are embedded and to which two positional embeddings are added, one spatial embedding and one temporal embedding. This is an adjustment compared to the ViT, which only adds the spatial embedding since it does not involve a time component. The transformer encoder remains unchanged and applies the multi-head attention mechanism among all pairs of patches, enriching the patches with global context. Contrary to the ViT, the ViTI2I applies attention over both space and time due to the extra dimension of the input data and the additional temporal embedding. The original ViT includes an MLP for classification, but this is replaced by a new MLP that projects the information in the transformed patches of the final input time step into patch embeddings belonging to the future time step. Therefore, the input size and output size of this MLP are equal to the number of embedding dimensions, as every patch embedding is passed through this MLP. The embedding dimension has been set to 120 in the ViTI2I, as it is large enough to capture local patterns whilst keeping the model size computationally manageable. The core of the MLP consists of two linear layers with a GELU activation function in between and the hidden layer contains 300 neurons. The future patch embeddings can be regarded as the latent space of the ViTI2I.

3.2.2. Decoder mechanism

The decoder mechanism is a convolutional decoder consisting of three deconvolution layers. These layers transform the future patch embeddings into the original size and resolution of the patches. The first layer has a kernel size and stride equal to the patch size and zero padding, which means the kernel projects every patch embedding into a separate region of the output space without overlap. This ensures that the future patch embeddings are upsampled from the compacted latent space back into the original size and resolution, by spreading each individual patch over its original area. The second and third layer both apply a kernel size of 3, a stride of 1 and a padding of 1. This preserves the size and resolution of the data as produced by the first layer, while allowing the network to refine the predicted frames. The number of kernels determines the number of output grids. To gradually reduce the 120 embedded dimensions to three predicted grids, the model uses a staged approach: the first deconvolutional layer applies 60 kernels, the second uses 30, and the final layer applies 3 kernels. This results in three two-dimensional output grids produced by the ViTI2I: the BWM and the two velocity fields.

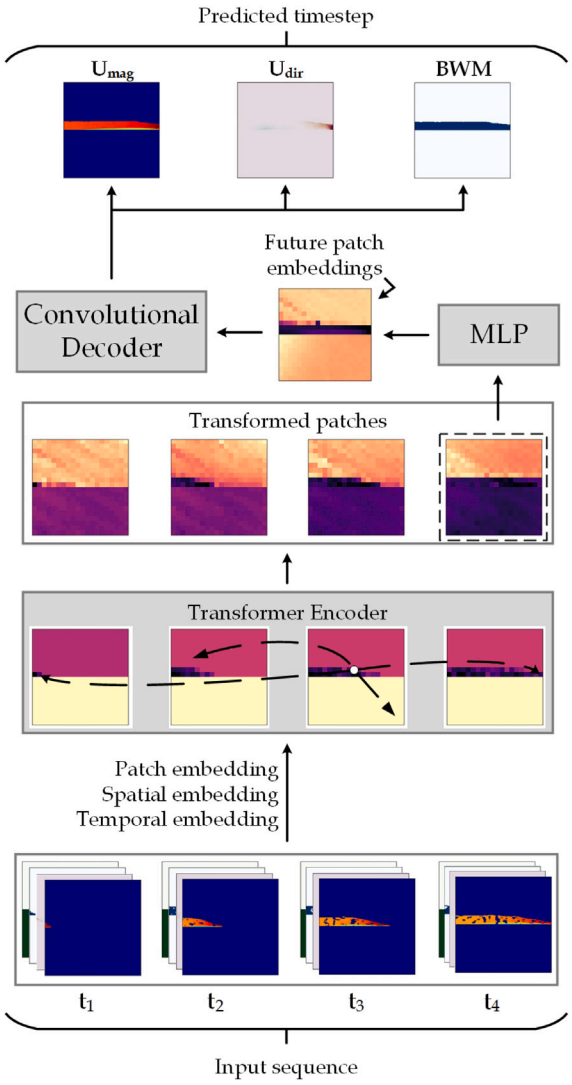

Fig. 4. Each time step in the input sequence consists of multiple grids, completing an input sequence of four frames. The input sequence is embedded into patches, to which a spatial and temporal embedding is added. The transformer encoder extracts global patterns, by sharing information across each pair of patches. The transformed patches of the final time step are projected into a future time step by the MLP, after which the convolutional decoder converts them into a detailed next-frame prediction.

3.3. Model training

3.3.1. Input data specifics

The ViTI2I is designed to predict the next frame for the BWM and the two velocity grids, using a sequence consisting of the BDM, the BWM and the two velocity grids as input. Training samples are generated to train the model, each providing an input sequence along with the corresponding target prediction. This means the model extracts and combines information across all channels and time steps, to simultaneously predict the next time step for velocity magnitude, velocity direction and the BWM. As such, the prediction for the three grids is trained jointly and not separately.

The size of the input data is constrained by computational limitations, as large input leads to increased memory usage and computational load when processed by the ViTI2I. The input size depends on both the number of time steps and the dimensions of the input grids. Using more time steps provides information on temporal behaviour, and an input sequence length of 4 time steps is used to balance the size of the data and the information it contains. The ViTI2I is trained on fixed-size input grids, as it relies on dividing the input grid into a fixed number of patches and learning the corresponding spatial and temporal embeddings. The total number of patches determines the number of computations done in the multi-head attention mechanism. The size of these patches is a variable and set to 10 × 10 cells. This is based on the findings of Vaswani et al. (2017) that smaller patches can improve model performance without greatly increasing model size, although they may require more computation due to the higher number of attention operations. A 10 × 10 size is considered small enough to detect and predict local patterns while still capturing a meaningful spatial domain, as it allows for input grids of 200 × 200 cells. This results in 400 patches per time step, which keeps the computational load manageable.

Table 2
Training hyperparameters.

Hyperparameter	Value
Epochs	200
Batch size	32
Learning rate	0.001
Optimizer	Adam

The input sequence is completed by defining a time step size between consecutive frames. Changing the step size affects how the training data is sampled, as larger step sizes result in input and predicted frames further apart in time. This also means that the model is trained to predict across that specific time interval, so the chosen step size directly determines the time step interval of the later simulations. Different step sizes are tested based on the concept that deep learning models are generally good at interpolating the patterns observed in the training data, but struggle at making predictions outside the boundaries of these patterns (Forrester et al., 2008). Larger step sizes are expected to allow more significant changes between frames, potentially offering a wider range of learnable patterns. The amount of frame-to-frame difference is quantified using Shannon entropy H [-] and the Structural Similarity Index Measure (SSIM).

Shannon entropy measures the randomness in a grid (Shannon, 1948). Within this study, low entropy indicates a grid with fewer unique values, which implies equal velocity values within the water body. The Shannon entropy of a single grid is computed as

$$H(X) = -\sum_{i=1}^{n} P(x_i) \log P(x_i) \tag{7}$$

where $P(x_i)$ is the probability of a particular value x_i occurring in that grid, based on the distribution of all cell values within that grid. The change in entropy between two consecutive frames is given by

$$\text{Entropy change} = H_{t+\Delta t} - H_t \tag{8}$$

The SSIM is a metric to quantify the difference in structure, brightness and contrast between two images. For this research, the SSIM compares overall velocity patterns, such as acceleration, deceleration and average velocity. High values for SSIM indicate large similarities, whereas low SSIM values reveal larger differences. In Section 4.1.1, entropy and SSIM are calculated for different step sizes to identify a step size that offers sufficient learnable patterns. The influence of this step size on the generated simulations is also discussed.

3.3.2. Training hyperparameters and loss function

The hyperparameters used for training are presented in Table 2. Additionally, early stopping occurs when the validation loss has not decreased for 12 consecutive epochs to prevent overfitting and the learning rate decreases by a factor 10 when the validation loss has not decreased for 4 consecutive epochs to optimize the training process.

The prediction of $\hat{\text{BWM}}$ is treated as a classification task, as each cell is labelled as either water or air. The Binary Cross-Entropy loss (L_{BWM}) is applied to $\hat{\text{BWM}}$, which is commonly used in classification as it focuses on correct class prediction rather than specific numerical values. Two loss function setups are tested for the velocity grids: an unmasked loss function L_{unmasked} and a masked loss function L_{masked}. The unmasked loss function uses the mean absolute error L_{MAE}, computing the loss in the velocity field without masking. The masked loss function is a combination of a loss function L_{water} for water cells and a loss function L_{air} for air cells, using masks to ensure only errors within water or air regions are included. This separation provides control over how strongly the model focuses on improving accuracy in either water or air cells.

$$L_{\text{BWM}} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \cdot \log(\hat{y}_i) + (1-y_i)\cdot \log(1-\hat{y}_i)\right] \tag{9}$$

$$L_{\text{MAE}} = \frac{1}{N}\sum_{i=1}^{N}\left|\hat{y}_i - y_i\right| \tag{10}$$

$$L_{\text{water}} = \frac{\sum_{i=1}^{N}\left(\left|\hat{y}_i - y_i\right|\cdot \hat{\text{BWM}}_i\right)}{\sum_{i=1}^{N}\hat{\text{BWM}}_i + 10^{-6}} \tag{11}$$

$$L_{\text{air}} = \frac{\sum_{i=1}^{N}\left(\left|\hat{y}_i - y_i\right|\cdot (1-\hat{\text{BWM}}_i)\right)}{\sum_{i=1}^{N}(1-\hat{\text{BWM}}_i) + 10^{-6}} \tag{12}$$

where $\hat{y}_i$ is a predicted value in cell i, y_i is the true value in cell i, $\hat{\text{BWM}}_i$ is the value in the predicted BWM and N is the total number of cells in the predicted grid. The loss L_{water} only includes errors in water cells, for which $\hat{\text{BWM}}_i$ is 1, and omits those in air cells, where $\hat{\text{BWM}}_i$ is 0. Similarly, L_{air} only retains errors in air cells.

The two applied loss functions are

$$L_{\text{unmasked}} = w_{\text{BWM}}\cdot L_{\text{BWM}} + w_U \cdot L_{\text{MAE},U_{mag}} + w_U \cdot L_{\text{MAE},U_{dir}} \tag{13}$$

$$\begin{aligned} L_{\text{masked}} = {} & w_{\text{BWM}}\cdot L_{\text{BWM}} + w_U \cdot L_{\text{water},U_{mag}} + w_A \cdot L_{\text{air},U_{mag}} \\ & + w_U \cdot L_{\text{water},U_{dir}} + w_A \cdot L_{\text{air},U_{dir}} \end{aligned} \tag{14}$$

The weights w define how much emphasis is laid upon what component of the loss function. Weight w_{BWM} is set to 1.0, weight w_U to 2.0 and weight w_A to 0.01. This leads to L_{masked} forcing the model to focus on cells that contain water by largely neglecting the errors in the air cells. Section 4.1.2 evaluates the simulations generated by models trained with the two different loss functions.

3.4. Model implementation

The trained ViTI2I can create individual predictions for a grid of 200 × 200 cells based on an input sequence with equal dimensions (Section 3.3.1). The total wave overtopping simulation domain exceeds these dimensions, as for example the 500 × 1000 grid in Fig. 3. Therefore, the ViTI2I is applied repeatedly across the domain by dividing the domain into overlapping 200 × 200 spatial subsets. This approach forms the core of the WOSM, the framework that uses the trained ViTI2I to generate wave overtopping simulation.

An overview of the WOSM procedure is given in Fig. 5. For each time step, the simulation domain is divided into subsets for which individual predictions are made by the trained ViTI2I. All individual predicted subsets are merged to retrieve prediction $\hat{t}_5$ of the entire domain. To complete a full simulation of multiple time steps, the WOSM removes t_1 from the input sequence and appends prediction $\hat{t}_5$, creating a new input sequence (t_2, t_3, t_4, $\hat{t}_5$) to predict time step $\hat{t}_6$. Repeated application of this method allows the WOSM to complete a full simulation of multiple time steps.

To ensure continuity in the simulation, spatial subsets overlap by half their area. Without overlap, the wave never enters a new subset and cannot be propagated from one subset to the next. In overlapping regions, the two separate predictions can produce different outcomes. Two merging techniques are developed and tested: *subset averaging* and *subset priority*. *Subset averaging* averages the velocity values in regions of overlap, whereas *subset priority* preserves only the prediction of the downstream half of each predicted subset and excludes the upstream half. These approaches are illustrated in Fig. 6, showing how the different techniques merge overlapping regions. The subset merging techniques are evaluated in Section 4.1.3.

3.5. Model testing

Different combinations of step size, loss function and subset merging technique are tested in Section 4.1 before extensively validating the WOSM on hydrodynamics against CFD model results. Both optimization and validation of the WOSM involve generating WOSM simulations and comparing them to the corresponding CFD simulations. During optimization, only simulations from the test dataset are used, which ensures the WOSM is evaluated on data unseen during training.

To determine the best-performing WOSM configuration, the RMSE is calculated between the maximum and depth-averaged flow velocity profiles modelled by the WOSM and CFD model. For larger step sizes, the flow velocity profiles begin to show fluctuations, as the wavefront travels farther between consecutive time steps. Since the wavefront typically exhibits the highest velocity, the locations where these peak velocities are recorded increase in distance for larger step sizes. To mitigate this effect, the velocity profiles are smoothed by applying a Gaussian filter with a standard deviation of 10. This leads to evaluation metric

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{x=1}^{N}(U(x) - \hat{U}(x))^2} \tag{15}$$

with velocity magnitude $U(x)$ [m/s] at location x in the CFD simulation after smoothing, velocity magnitude $\hat{U}(x)$ [m/s] at location x in the WOSM simulation after smoothing and total number of measured locations N [-].

3.5.1. Procedure for validation on hydrodynamics

The best-performing WOSM configuration is used to generate additional simulations for hydrodynamic validation, which is presented in Section 4.2. This validation is performed by comparing both the maximum values and time series of flow velocity, depth-averaged velocity and water depth at four different locations along the dike profile (Fig. 7). These four locations represent typical positions found along any dike profile: the crest, the upper slope, the lower slope, and the area beyond the toe. The time series shows how well the WOSM simulates the dynamics of the overtopping wave, including the peak velocity at the wave front followed by a decreasing trend. It also reflects the ability of the WOSM to capture the overtopping duration, which is defined as the period between the arrival of the wavefront to the moment the tail of the wave passes. At the same four locations, the vertical flow structure is reviewed. This helps assess whether the WOSM can reproduce the logarithmic vertical velocity profile and capture the boundary layer. Accurate modelling of the vertical flow structure demonstrates whether the WOSM can be useful in erosion modelling, where the near-bed velocity profile influences the stress applied to the bed profile. In a simplified manner, the WOSM is evaluated for mass conservation. The simulations are two-dimensional and do not have an along-dike depth component, so mass is represented by the total surface area of the wave. The evaluation is done by comparing the total simulated water area per time step between the WOSM and the CFD model. As the WOSM is trained to replicate the CFD model, mass conservation is considered satisfied when both models simulate equal total mass per time step.

Besides comparing the maximum values for water velocity and depth between the CFD and WOSM simulation directly, they are also

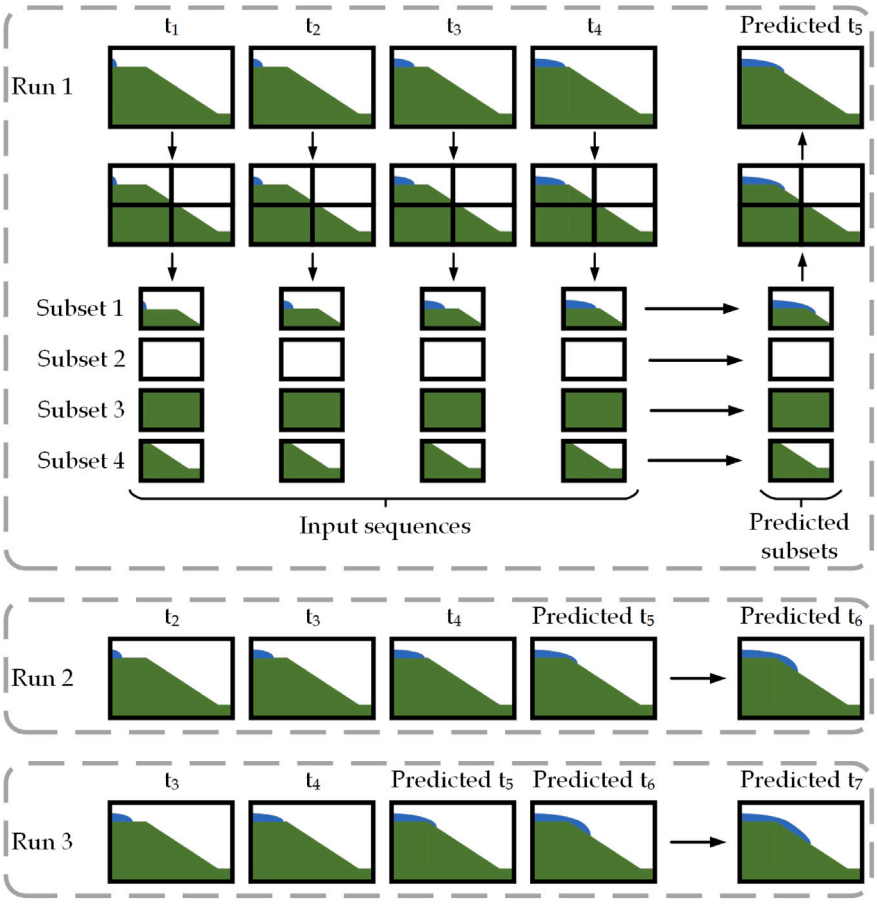

Fig. 5. Every time step is divided into spatial subsets. Every subset acts as an individual input sequence to the ViTI2I, which predicts the next frame for the specific subset. The predicted subsets are merged to return the next-frame prediction of the complete domain. New input sequences are created by removing the first time step and appending the latest predicted simulation domain to the original sequence. Repeating this process enables complete wave overtopping simulations and leads to multiple predicted time steps.

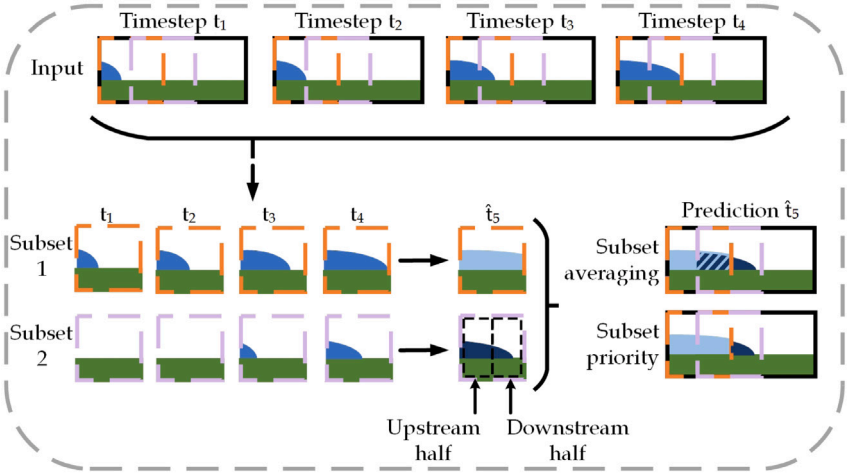

Fig. 6. The input sequence is divided into smaller subsets that overlap. Only subsets that include water in the input sequence are passed through the ViTI2I and their predictions are merged. *Subset averaging* averages the values of both predicted subsets, whereas *subset priority* only retains the downstream half of each predicted subset and excludes the upstream half.

evaluated per generated simulation by computing the normalized root mean squared error [-] (nRMSE):

$$\text{nRMSE} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2}}{\frac{1}{n}\sum_{i=1}^{n} y_i} \tag{16}$$

where $\hat{y}_i$ denotes the WOSM prediction, y_i the CFD simulation value, and n the number of measurement points, which is four (corresponding to locations D1–D4). The nRMSE expresses the RMSE relative to the observed mean value, indicating how large the prediction error is compared to the magnitude of the measured value.

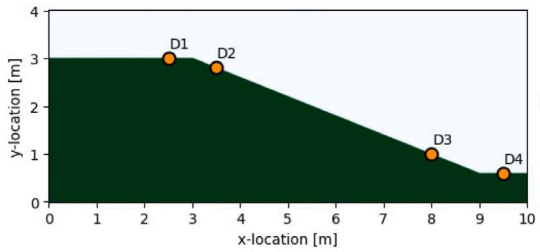

Fig. 7. The four locations along the dike profile D1 (x = 2.5 m), D2 (x = 3.5 m), D3 (x = 8 m) and D4 (x = 9.5 m) used for validation. The dike profiles in the simulations have varying slope steepnesses which influences the difference in height between the crest and locations D2, D3 and D4. However, the x-distances to transitions in geometry remain consistent.

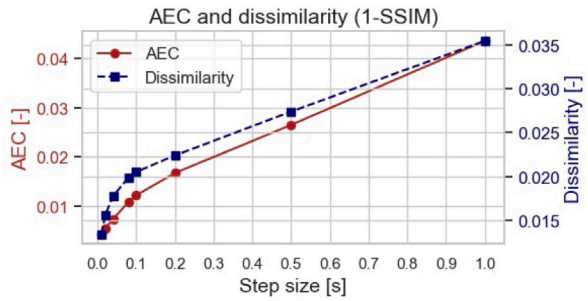

Fig. 8. The values for AEC and dissimilarity for step sizes 0.01 s, 0.02 s, 0.04 s, 0.08 s, 0.1 s, 0.2 s, 0.5 s and 1.0 s. A step size of 0.1 s is chosen as a trade-off between an increase in learnable patterns and the loss of information in the simulations.

For the validation on hydrodynamics, additional simulations are generated to provide a wider range of overtopping scenarios on which the WOSM can be evaluated. Some of these simulations originate from the validation dataset. Normally, this is undesired, but it can be justified because the ViTI2I starts to use predicted subsets as new input, which resembles data from the validation dataset but is not a direct representation. Therefore, the broader range of scenarios is considered valuable and reliable in demonstrating the performance of the WOSM.

4. Results

Section 4.1 presents the performance of the WOSM for different step sizes, loss functions and subset merging techniques. The best-performing WOSM is validated on the modelled wave overtopping hydrodynamics, which is described in Section 4.2.

4.1. Surrogate model optimization

As a benchmark, the simulation technique *subset priority* is applied to test the WOSM for different step sizes and loss functions. The subset merging techniques are assessed last, based on the defined step size and loss function.

4.1.1. Step size selection

The difference in entropy between consecutive grids is computed for various step sizes by averaging the entropy change across all grid pairs, resulting in the average entropy change (AEC). Similarly, SSIM is calculated and converted to a dissimilarity score (1 – SSIM), where higher values indicate larger differences. These metrics are used to check how larger step sizes lead to more variation between frames. The results are presented in Fig. 8 and confirm that both AEC and dissimilarity increase with step size.

Table 3
RMSE between the flow velocity profile modelled by the WOSM and CFD model for different step sizes. Values for RMSE are given for maximum velocity and depth-averaged (DA) velocity.

Step size [s]	RMSE					
	Max velocity			DA velocity		
	0.02	0.05	0.1	0.02	0.05	0.1
V = 0.5 m^3/m, s = 1:1.7	1.15	0.77	*0.40*	1.04	0.73	*0.37*
V = 1.25 m^3/m, s = 1:1.7	2.63	0.51	*0.38*	2.42	0.42	*0.39*
V = 1.5 m^3/m, s = 1:2.2	0.77	0.48	*0.29*	0.72	0.44	*0.31*

Table 4
RMSE between the flow velocity profile modelled by the WOSM and CFD model for the unmasked and masked loss functions. Values for RMSE are given for maximum velocity and depth-averaged (DA) velocity.

Loss function	RMSE			
	Max velocity		DA velocity	
	No mask	Mask	No mask	Mask
V = 0.5 m^3/m, s = 1:1.7	1.75	*0.40*	1.55	*0.37*
V = 1.25 m^3/m, s = 1:1.7	1.89	*0.38*	1.94	*0.39*
V = 1.5 m^3/m, s = 1:2.2	1.77	*0.29*	1.85	*0.31*

The increase is steepest between 0 and 0.1 s, suggesting a valuable increase in learnable patterns without compromising the practical use of the simulations. Step sizes beyond 0.1 s may lead to a loss of information and impractical simulations, since larger parts of the simulations are skipped. The ViTI2I is trained three times using datasets with step sizes of 0.02 s, 0.05 s and 0.1 s, and applying the masked loss function described in Section 3.3.2.

The performance of the WOSM is evaluated by generating simulations from the test dataset and comparing them to the original CFD simulations. For each case, the maximum and depth-averaged flow velocity profiles are extracted, and the RMSE between the WOSM and CFD profiles is calculated (see Section 3.5). Low RMSE values indicate the WOSM models a similar flow velocity profile as the CFD model. An example of the observed flow velocity profiles is given in Appendix A. The results are presented in Table 3.

The results reveal that the lowest RMSE values are achieved using a step size of 0.1 s across all three simulations, which can be attributed to this WOSM having learned more pronounced and meaningful patterns. Smaller step sizes result in minimal changes between consecutive frames, which can cause the model to rely more on copying the input rather than learning the spatiotemporal evolution. This is observed when simulating the overtopping wave over the transition from crest to slope, where the WOSM is required to rotate the direction of the flow from horizontal to diagonal. With low step sizes, this rotation must be simulated gradually with small adjustments per time step. As the model is trained on less distinctive patterns, it begins to underestimate the required rotation, causing the flow to remain horizontal longer than in the CFD simulation. Larger step sizes allow the model to observe more significant movement and clearer changes in flow direction between consecutive frames, making the rotation easier to learn, recognize and predict. As a result, the WOSM with smaller step sizes falls behind the CFD simulation, which introduces the first signs of decreased performance (Fig. 9). Based on this performance, the continuation of this study applies a WOSM configuration with a step size of 0.1 s.

4.1.2. Loss function selection

The ViTI2I is trained twice using either the unmasked or the masked loss function and both are applied in the WOSM framework, resulting in new simulations for which the RMSE is calculated. The results are presented in Table 4.

The WOSM with a masked loss consistently results in lower RMSE values, indicating better performance in capturing the flow velocity

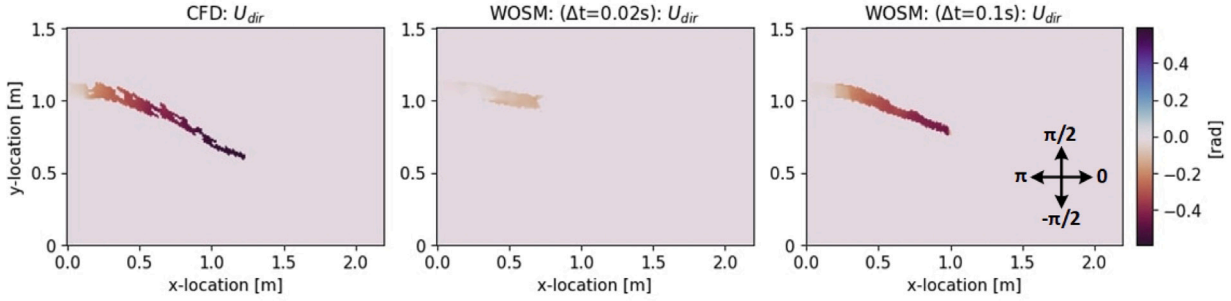

Fig. 9. The flow velocity direction field at the transition from crest to slope at t = 0.8 s for the simulation of V = 1.25 m^3/m with s = 1:1.7. The 0.02 s step size WOSM falls behind on the CFD simulation, whereas the 0.1 s WOSM more closely captures the flow rotation and trajectory.

Table 5
RMSE between the flow velocity profile modelled by the WOSM and CFD model using either *subset averaging* or *subset priority*. Values for RMSE are given for maximum velocity and depth-averaged (DA) velocity.

Simulation technique	RMSE			
	Max velocity		DA velocity	
	Averaging	Priority	Averaging	Priority
V = 0.5 m^3/m, s = 1:1.7	*0.21*	0.40	*0.21*	0.37
V = 1.25 m^3/m, s = 1:1.7	*0.37*	0.38	*0.33*	0.39
V = 1.5 m^3/m, s = 1:2.2	0.31	*0.29*	0.35	*0.31*

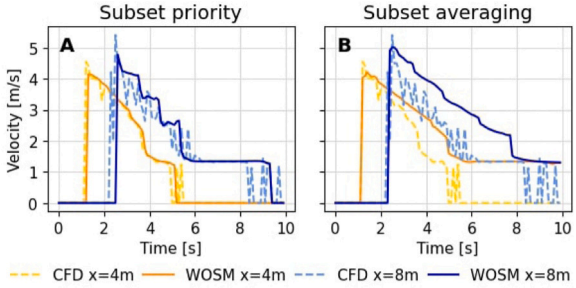

Fig. 10. The maximum flow velocity profiles over time in the simulation of V = 0.5 m^3/m with s = 1:1.7 at locations x = 4 m (yellow colour tone) and x = 8 m (blue colour tone) as simulated by the WOSM (solid lines) and CFD model (dashed lines). **Graph A** shows the WOSM simulation generated using *subset priority*, where the WOSM velocity profile closely follows the CFD profiles at both locations. **Graph B** shows the WOSM simulation generated using *subset averaging*. This simulation technique leads to the WOSM simulating higher velocities than the CFD model, as shown by the solid lines exceeding the dashed ones.

profile. This is because the masked loss function focuses specifically on improving predictions within the water region, preventing the loss of learning capacity to the correct prediction of zero velocities in air cells. Contrary, the unmasked loss function treats all grid cells equally, minimizing errors uniformly across the grid. While the masked loss function leads to larger errors in air cells, they are filtered and returned to zero during post-processing by multiplying the predicted velocity grids with BWM (see Appendix B). As a result, the ViTI2I trained with a masked loss function predicts velocity values closer to the CFD model, whereas the unmasked model underpredicts flow velocities, resulting in decreased performance and higher RMSE values. The masked-loss ViTI2I with a step size of 0.1 s is used to evaluate the two subset merging techniques.

4.1.3. Subset merging technique selection

Simulations are generated using either *subset averaging* or *subset priority*, and the RMSE between the WOSM and CFD flow velocity profile is calculated. The results are presented in Table 5, indicating similar performance between the two merging techniques.

Since the RMSE results do not indicate a clear best-performing merging technique, velocity time series are analysed for further evaluation. The velocity time series modelled by both merging techniques are presented in Fig. 10.

The figure demonstrates that using *subset priority* leads to an accurate velocity profile, as the WOSM captures a similar peak velocity and decreasing trend as the CFD simulation. The velocity profile modelled using *subset averaging* shows the WOSM maintains higher velocities for a longer period than the CFD model, even after the wave has fully passed in the CFD simulation. This occurs because *subsets priority* is better at prolonging changes that occur in the upstream region in downstream direction, improving the simulation of features such as deceleration and the tail of the wave (see Fig. 11).

Evaluation of different step sizes, loss functions and subset merging techniques demonstrated that the best-performing WOSM uses a 0.1 s step size, is trained with a masked loss function and applies *subset priority*. This WOSM generates simulations that display 10 s of wave overtopping with intervals of 0.1 s, producing 100 frames per simulation. In the next section, this best-performing WOSM is validated against the original CFD model to assess its ability to generate simulations that replicate the original hydrodynamics. For this validation, maximum values and time series of flow velocity, depth-averaged flow velocity and water depth are evaluated, vertical flow structures are compared and mass conservation is examined. The velocities and water depth are evaluated at locations D1–D4, as introduced in Section 3.5.1.

4.2. Validation of the hydrodynamics

In most cases, the WOSM captures the shape of the overtopping wave well. As shown in Figs. 12 and 13, the contours and area of the overtopping wave are similar between the WOSM and CFD model simulations. Cells where one model shows water and the other shows air naturally produce large differences, which clearly appear as dark blue or dark red areas in the error grids. In the cells where both models simulate water, the WOSM captures the velocity well. This is further described in the subsequent sections that evaluate velocity components. At times, the wave front modelled by the WOSM is slightly delayed compared to that of the CFD, as illustrated at t = 1.3 s in Fig. 12. This delay typically corresponds to a single time step, meaning the WOSM simulation reaches a given location 0.1 s after the CFD simulation. This figure also illustrates the limitation of the WOSM in predicting the entrapped air bubbles, as its simulations show a solid water body (see subplot A). In Fig. 13, an anomaly occurs where a small horizontal flow is simulated mid-slope. All results are further discussed in Section 5.1. Flow field results for the simulation of V = 0.5 m^3 with s = 1:1.7 are presented in Appendix C.

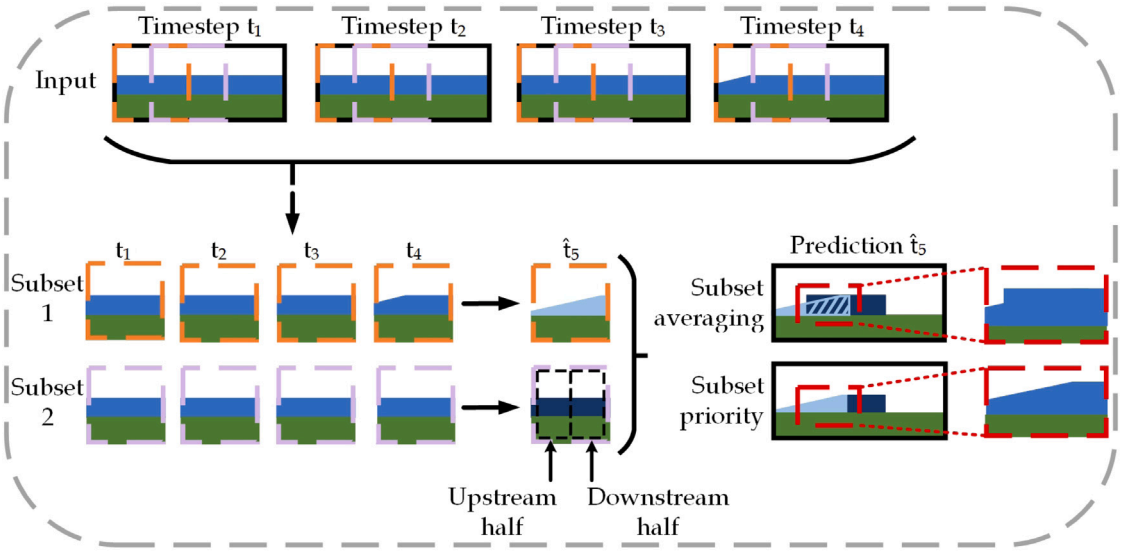

Fig. 11. At input timestep t_4, a decrease in water depth occurs at the inflow boundary, following a period of stable overtopping flow during timesteps t_1–t_3. This change in hydrodynamics does not fall within the range of subset 2. This causes the ViTI2I prediction for this individual subset 2 to again show a stable flow at $\hat{t}_5$, as it has been given no information on this upstream change in water depth. In contrast, subset 1 does extract this change, and thus the ViTI2I correctly predicts a continuation of the decrease in water depth. When *subset averaging* is applied, the incorrect prediction in the upstream half of subset 2 is incorporated into the final predicted domain. *subset priority* produces a more accurate simulation by retaining only the downstream half, thereby better propagating the upstream change in flow towards the downstream direction.

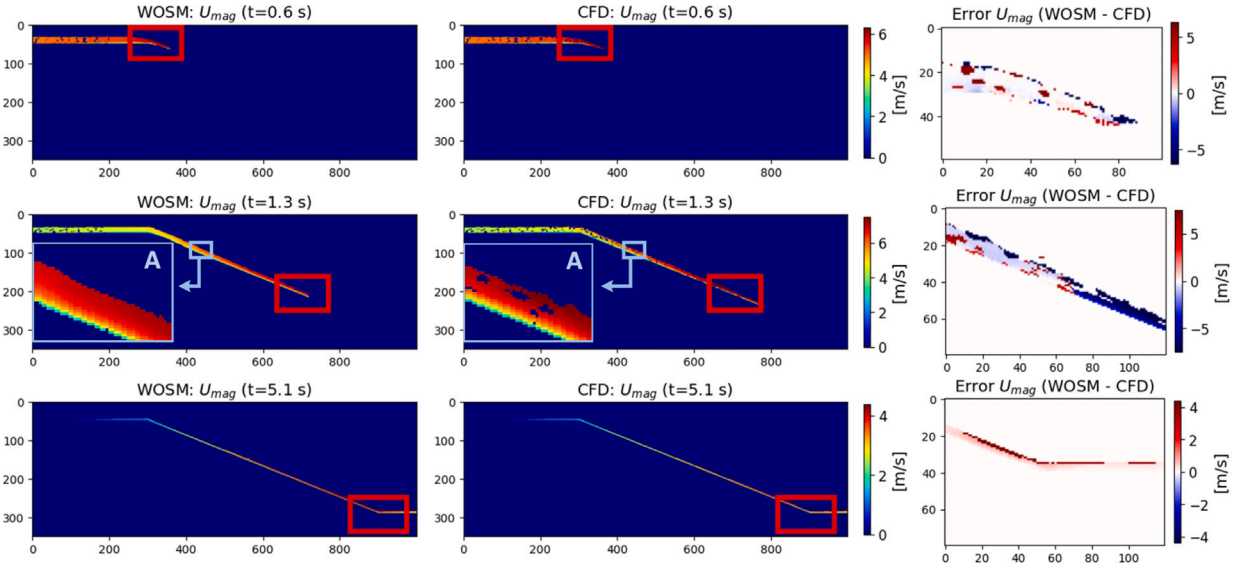

Fig. 12. Results for U_{mag} as modelled by the WOSM and CFD model for an overtopping wave of volume V = 2 m^3 over a dike with slope steepness 1:2.5 at t = 0.6 s, 1.3 s and 5.1 s. The red-outlined areas are used to calculate the error by subtracting the CFD field from the WOSM field, as shown in the right-hand column. In these plots, dark blue areas indicate areas where the CFD model simulates water, but the WOSM models air. Contrary, dark red areas indicate areas where the WOSM models water, but the CFD model simulates air. At t = 1.3 s, the wavefront simulated by the CFD model is located further along the slope than the one modelled by the WOSM, illustrating the small delay of the WOSM simulation. In subplots A, it can be seen that the WOSM simulates a solid water body whereas the CFD model results in entrapped air bubbles within the wave.

4.2.1. Maximum values

The maximum values for flow velocity and depth-averaged flow velocity modelled by the WOSM and CFD model are presented in Fig. 14. Each simulation is evaluated by computing the nRMSE (Eq. (16)) and the results are presented in Table 6.

The results show good accuracy for the two velocity components, as a large number of points lie close to the line of agreement and the nRMSE values typically range between 1%–18%. This can be considered good, as the deviations represent only a small fraction of the mean velocities. Simulating overtopping volume V = 1.5 m^3/m over slope steepness s = 1:1.5 results in larger deviations at locations D3 and D4, with the WOSM predicting zero velocity at D4. This is also represented in the increased nRMSE value, where the single outlier has a large influence. The WOSM water depths generally also align decently with the CFD model with almost no extreme outliers, except the zero water depth at D4 in the simulation of V = 1.5 m^3/m with s = 1:1.5. This outlier again causes a high value for the nRMSE. Also, some more notable deviations are observed at D2 for the simulations of V = 2.0 m^3/m with s = 1:2.5, V = 1.25 m^3/m with s = 1:2.2 and V = 1.5 m^3/m with s = 1:2.2. Such individual deviations cause

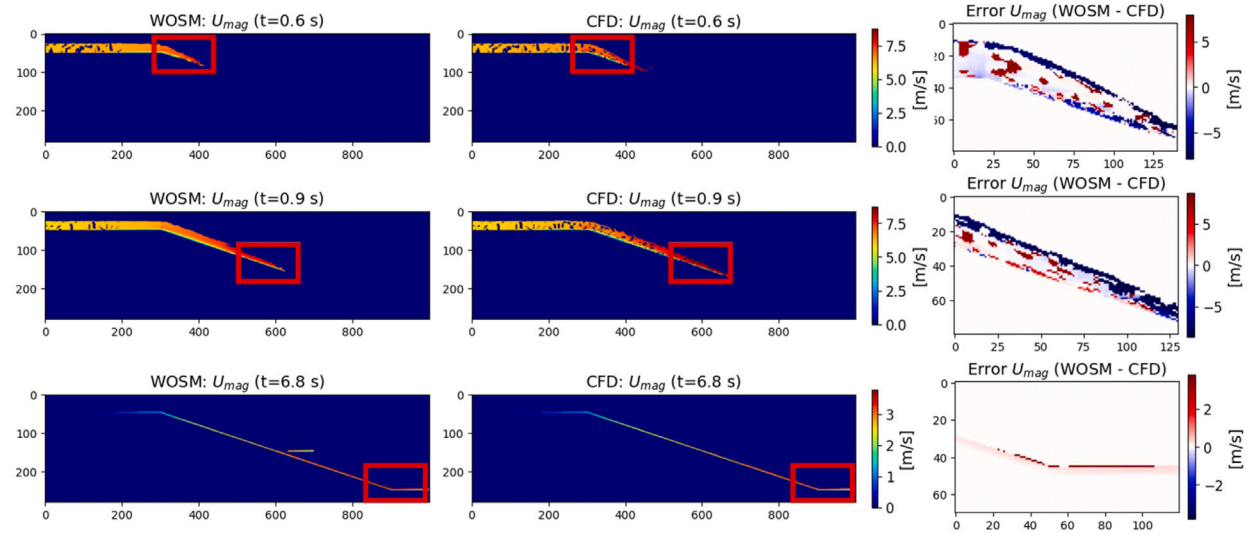

Fig. 13. Results for U_{mag} as modelled by the WOSM and CFD model for an overtopping wave of volume V = 4 m^3 over a dike with slope steepness 1:3.0 at t = 0.6 s, 0.9 s and 6.8 s. The red-outlined areas are used to calculate the error by subtracting the CFD field from the WOSM field, as shown in the right-hand column. Dark blue areas indicate areas where the CFD model simulates water, but the WOSM models air. Dark red areas indicate areas where the WOSM models water, but the CFD model simulates air. At x = 600, a small horizontal flow can be observed that occurs mid-slope.

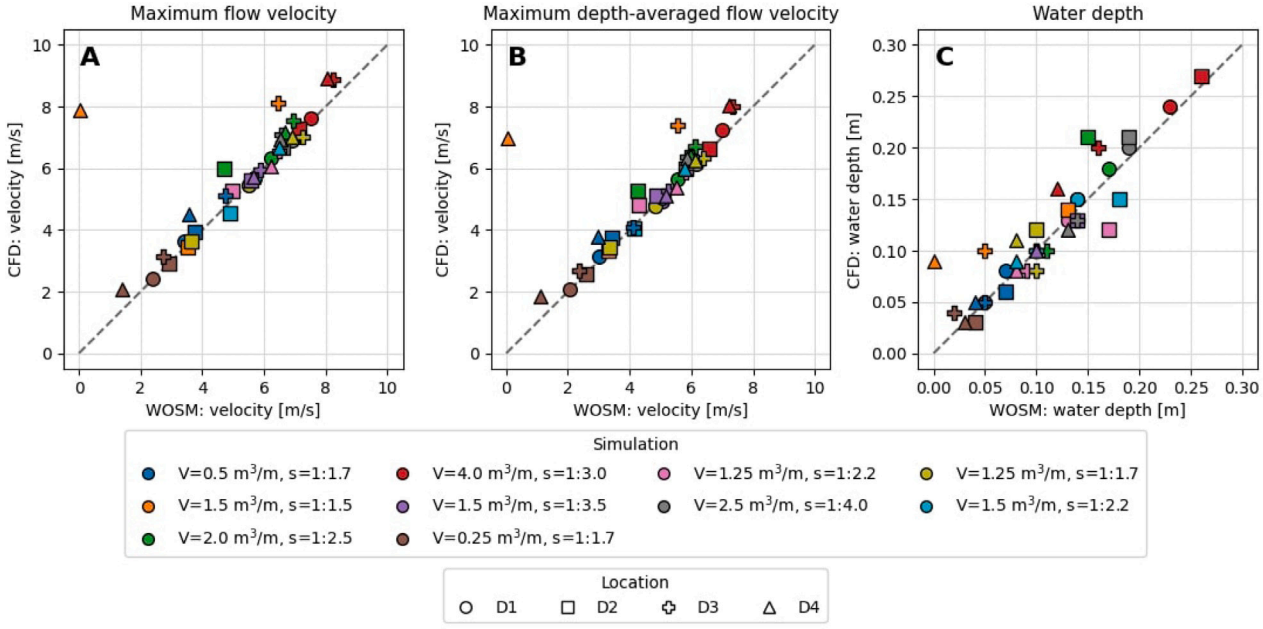

Fig. 14. Comparison of maximum values for flow velocity, depth-averaged flow velocity and water depth as simulated by the WOSM and CFD model at the four locations (D1 to D4). Each colour represents a different simulation, defined by the overtopping volume V and slope steepness s. The marker shapes indicate the location where the maximum value is recorded. Points that lie exactly on the dashed grey diagonal indicate perfect agreement between the WOSM and CFD model outputs, whereas points further away reflect larger deviations. Results are shown in **subplot A** (flow velocity), **subplot B** (depth-averaged flow velocity), and **subplot C** (water depth).

the remaining nRMSE values to range between 6%–30%, making them generally higher than their velocity-related counterparts.

4.2.2. Time series

The results for the time series of maximum velocity, depth-averaged velocity and water depth are presented in Fig. 15. For all cases except V = 2 m^3/m at location D2, the overtopping duration is accurately captured, as the arrival and passing of the wave occur at similar moments across the two simulations. The WOSM shows a large delay in modelled velocity compared to the CFD model for the simulation of an overtopping volume of 2 m^3/m at location D2, which affects the predicted overtopping duration (see graph B and F). Velocity is first recorded at t = 2.0 s in the WOSM simulation, compared to t = 0.6 s in the CFD simulation. Despite this delay, the velocity trends over time are otherwise very similar, with the WOSM capturing the initial peak velocity followed by a gradual decrease. This also holds for water depth, although larger deviations appear at location D3 and D4 for the simulation of 4 m^3/m (graph K and L). In this case, the WOSM simulates

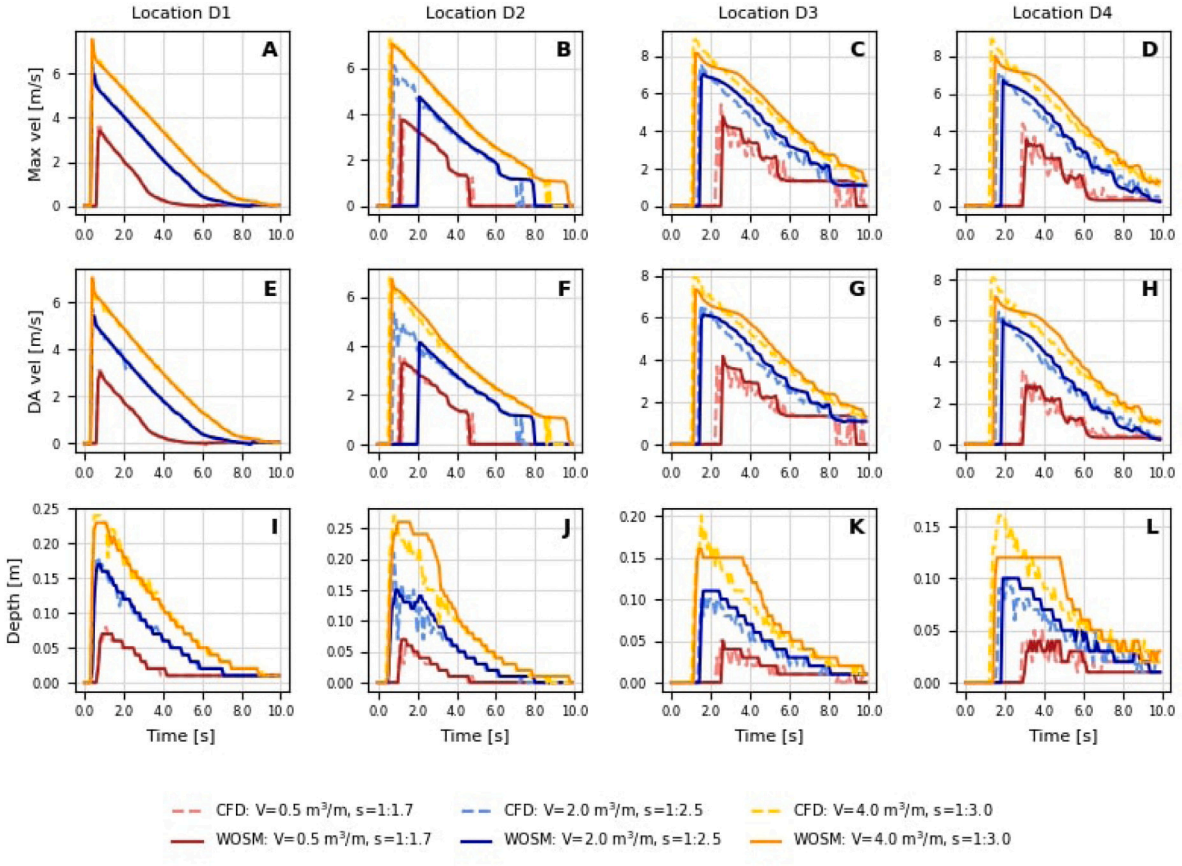

Fig. 15. The maximum velocity, depth-averaged velocity and water depth over time at locations D1, D2, D3 and D4 as modelled by the CFD model and the WOSM. Subplots A–D show the maximum velocity over time, E–H show the depth-averaged (DA) velocity over time, and I–L present the water depth over time. Columns 1 to 4 correspond to locations D1 to D4, respectively.

Table 6
The nRMSE between the maximum values for velocity, depth-averaged velocity and water depth as modelled by the WOSM and CFD model. The nRMSE is computed for each simulation individually, using the four measured locations.

Simulation	Maximum velocity	Depth-averaged velocity	Water depth
V = 0.5 m³/m, s = 1:1.7	0.120	0.121	0.144
V = 1.5 m³/m, s = 1:1.5	0.639	0.633	0.433
V = 2.0 m³/m, s = 1:2.5	0.110	0.103	0.209
V = 4.0 m³/m, s = 1:3.0	0.065	0.071	0.134
V = 1.5 m³/m, s = 1:3.5	0.010	0.025	0.059
V = 0.25 m³/m, s = 1:1.7	0.151	0.180	0.298
V = 1.25 m³/m, s = 1:2.2	0.028	0.053	0.249
V = 2.5 m³/m, s = 1:4.0	0.049	0.058	0.080
V = 1.25 m³/m, s = 1:1.7	0.023	0.018	0.189
V = 1.5 m³/m, s = 1:2.2	0.039	0.041	0.135

a water depth that initially remains constant for 2-3 seconds rather than showing a decreasing trend.

4.2.3. Vertical flow structure

The vertical flow structure modelled by the WOSM closely matches that of the CFD model (Fig. 16). The WOSM shows good performance in capturing the vertical logarithmic structure of the flow. At locations where the WOSM overpredicts the water depth, the vertical flow profiles extend higher than those in the CFD simulation. In these cases, the WOSM still predicts a realistic and plausible pattern, as can be seen in the simulation of $V = 4$ m^3 at location D2.

4.2.4. Mass conservation

Fig. 17 presents the total simulated area per time step, representing the mass in the simulation over time. The modelled area matches closely between both models for the simulation with the smaller overtopping volume $V = 0.5$ m^3/m. Deviations increase when larger volumes are simulated, and the largest errors occur during a period after the maximum area is modelled (between $t = 2$–5 s). Near the end of the simulation, the deviation decreases and the modelled areas by the WOSM and CFD model converge again.

5. Discussion

The results reveal that the WOSM is capable of accurately reproducing many key hydrodynamic patterns. Building on these findings, the discussion is structured as follows. Section 5.1 evaluates the performance of the WOSM, identifying the conditions under which performance declines and proposing possible mitigation measures. Section 5.2 describes the capabilities and limitations of the WOSM in relation to existing modelling approaches. Section 5.3 discusses how the ViTI2I architecture and the training dataset can influence WOSM performance.

5.1. Surrogate model performance

The results in Fig. 14 indicate that the WOSM captures the maximum values for flow velocity well for all simulations among the four locations, except for location D3 and D4 in the simulation of $V = 1.5$ m^3/m with $s = 1{:}1.5$. Here, the WOSM simulations deviate significantly, predicting zero velocity and water depth at location D4. This results from the WOSM correctly simulating flow separation, but failing to model the flow reattachment. After flow reattachment

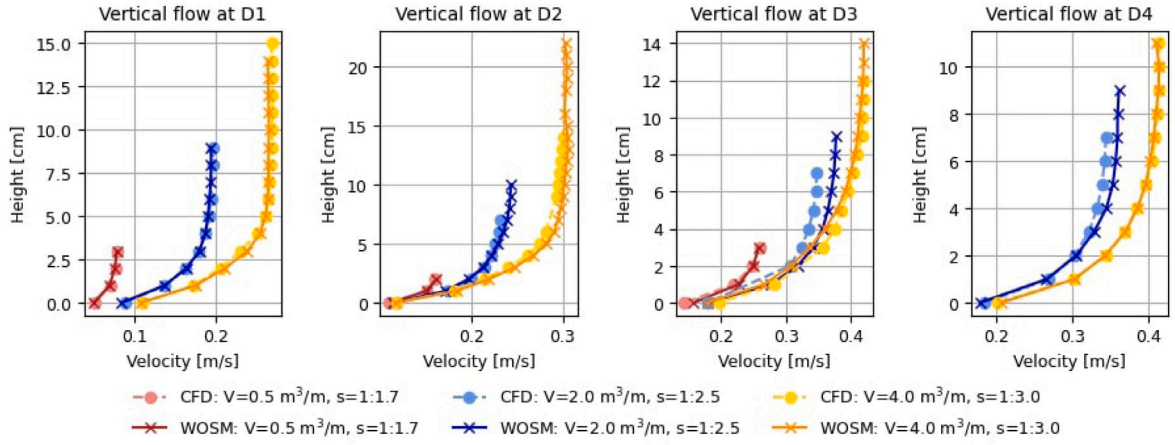

Fig. 16. The vertical flow structure at locations D1, D2, D3 and D4 at t = 2.7 s modelled by the CFD and WOSM. Simulation V = 5 m^3/m, s = 1:1.7 is not showing at location D4 because the wave has not reached this location yet.

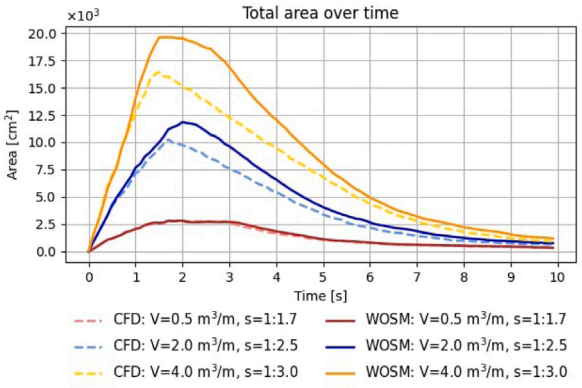

Fig. 17. The total water area over time modelled by the WOSM (solid lines) and the CFD model (dashed lines). The deviation is largest around the peak and the period after.

occurs, the WOSM no longer produces meaningful results as the flow is incorrectly modelled. This likely occurs due to a lack of training data that shows flow reattachment, as only 12 simulations in the training dataset show clear flow separation and reattachment. This means there are only 12 examples from which the ViTI2I can learn how to predict wave reattachment, which can reduce its ability to learn this behaviour. In addition, flow separation and reattachment can increase the complexity of the simulation, as a separated flow is influenced by momentum and gravity and it is not as simple as aligning the flow to the dike profile. For the remaining simulations, the one of V = 0.25 with s = 1:1.7 showed the largest nRMSE for both velocity and water depth. This overtopping volume is the smallest volume and is included only once in the training dataset. As a result, the ViTI2I has been given little data to learn from and cannot interpolate patterns observed from smaller overtopping volumes, making this simulation challenging to predict well. The remainder of the modelled water depths do not include extreme outliers, though more pronounced deviations occur at location D2, which lies just after the transition from crest to slope. This location is particularly challenging to predict, because it requires the WOSM to consider multiple aspects simultaneously—rotation of the flow direction and whether flow separation occurs. This increases the complexity of the prediction, as the same predictive capacity is required to make multiple correct decisions. Additionally, the nRMSE showed higher values compared to the velocity-related counterparts. Unlike velocity values, which are predicted by the WOSM as continuous variables and can deviate by any amount, water depth is derived from the predicted BWM, which introduces errors in steps of 1 cm. The smallest possible deviation is one grid cell with a height of 1 cm, and errors increase per additional simulated cell. Since true water depths range from 0.03 m to 0.27 m, even a misclassification of a few cells can result in a relatively high nRMSE. A potential solution can be to introduce smaller cell sizes, which is further discussed in Section 5.3.

Challenges at location D2 are also visible in the time series results for flow velocities and water depth (Fig. 15). Overall, the WOSM performs well for time series as it accurately captures the overtopping duration, peak values and the expected decreasing trend with only very small delays. Such delays are, for example, observed at location D4, where the wavefront simulated by the WOSM arrives slightly later than in the CFD simulation. This effect is also visible in Figs. 12 and 13, where the WOSM wavefront is positioned further up the slope than the CFD wavefront. The delay is attributed to the crest-to-slope transition, where the WOSM must account for multiple interacting factors. This increases prediction complexity, causing the WOSM to require an extra timestep to reach the required state of flow rotation and velocity. This can introduce a small delay before the WOSM starts propagating the wave down the slope. A large delay is observed in the predicted velocity at location D2 in the simulation of V = 2.0 m^3/m with s = 1:2.5, visible in graph B in Fig. 15. This delay results from the WOSM incorrectly predicting flow separation, when it does not occur in the original CFD simulation. This means that for some simulations, the WOSM simulates the wave front separated from the slope, before reattaching it to the dike profile. The flow velocity is only recorded when the flow is attached to the dike, so it appears in the WOSM flow velocity profile at t = 2.0 s, which is the moment it reattaches the flow to the dike. Once the flow reattaches, the flow velocity is accurately modelled, still indicating good performance for predicting flow velocities. The incorrectly simulated flow separation can again be attributed to limited training data, because the training dataset only includes 28 simulations of individual wave overtopping. As each simulation only contains one specific sequence where the wavefront travels over this transition, there are only 28 examples from which the ViTI2I can learn. Another reason why the WOSM wrongly predicts flow separation is that it determines this based on the interpolation of learned patterns and not on any inherent understanding of physics. Due to its data-driven nature, the WOSM estimates whether separation occurs based on learned patterns instead of physical principles. When faced with unseen combinations of velocity and slope steepness, it can nudge towards the closest known pattern rather than computing the outcome based on physics.

Contrary to flow velocity, values for water depth are extracted regardless of whether the flow is attached or separated, which is why no delay in water depth is visible. Fig. 15 demonstrates that water depth over time is modelled well by the WOSM, although larger deviations

occur in the simulation with overtopping volume $V = 4$ m^3/m. This overtopping volume is the largest volume and is included only once in the training dataset. This causes the same problems as encountered when simulating the lowest overtopping volume. The ViTI2I has been given less data for this particular overtopping volume, and can interpolate less patterns to create new predictions. Errors can arise more easily, potentially leading to error accumulation, where small errors in early predictions are included in new input sequences, causing the model to make predictions based on flawed input data. This can lead to an amplifying effect, where the error increases as the simulation progresses. As a result, errors observed in Fig. 15 at location D3 and D4 (graphs K and L) are larger than the errors at D1 and D2 (graphs I and J), where the wave passes earlier.

The other evaluated velocity-related component is the vertical flow structure, which is accurately predicted by the WOSM (Fig. 16). This demonstrates the ability of the WOSM to learn the effects of the applied no-slip boundary condition in the CFD model, as the near-bed velocity profile is simulated well. This can be useful in erosion modelling, where the near-bed velocity profile influences the stresses applied to the bed profile. It could be explored to use smaller cell sizes to capture the velocity profile in more detail.

Finally, a comparison of the accumulated mass between the WOSM and CFD simulation is shown in Fig. 17, indicating that the WOSM reproduces a similar mass for the case with $V = 0.5$ m^3/m and $s = 1{:}1.7$. However, deviations are observed for the simulations of larger overtopping volumes. Three factors contribute to this increased mass. First, the WOSM does not recognize the behaviour of entraped air bubbles and simulates solid water bodies, resulting in larger simulates masses (Fig. 12). Second, the WOSM can overpredict water depths, which increases the total mass. This is visible in Fig. 15 (graphs J, K and L), where the simulated water depth for $V = 2$ m^3/m is larger in the WOSM simulation than in the CFD simulation between $t = 2$–6 s. This occurs when the WOSM is uncertain about whether the flow should separate or remain aligned at the transition from crest to slope. In such cases, it may predict a combination of both, where the bottom layer aligns with the dike profile, while the top layer shows mild signs of flow separation. Although this effect is minimal, it stretches the flow vertically, increasing the water depth. Consequently, the deviation in mass is largest during the period the wavefront flows over the transition and shortly after. Once the flow velocity at the crest decreases, the WOSM recognizes that separation should no longer occur, and the stretching effect is resolved. Lastly, an anomaly appears in several simulations, where the WOSM predicts a horizontal flow mid-slope (Fig. 18). This situation arises when the ViTI2I generates a poor prediction at a subset boundary. The boundaries of predicted grids are complicated to predict well, as the ViTI2I lacks information on what lies outside these boundaries. In such cases, it can assume a horizontal dike profile below the observed domain, causing it to predict a horizontal flow at the subset boundary. A potential solution is to develop a subset merging technique that retains only the central part of each predicted subset. This approach ensures that each preserved prediction is based on surrounding context and excludes predictions of boundary regions. The increase in mass deviation for larger overtopping volumes occurs because a greater volume leads to a larger water area and greater water depths, reinforcing two of the causes of mass deviation. First, a larger water surface allows more bubbles to form, resulting in more air being filled with water. Second, when a horizontal flow is modelled mid-slope, the increased water depth from a larger overtopping volume causes greater mass deviation because more water is simulated incorrectly.

5.2. Surrogate model capabilities and limitations

The best-performing WOSM produces simulations of 100 frames, representing 10 s of individual wave overtopping. Generating one simulation with the WOSM takes 3–10 s on an NVIDIA H100 NVL GPU and 15–50 s on an Intel® Xeon® Platinum 8462Y+ CPU with 44 cores. This is a significant improvement to the long runtime of the original CFD model, which requires approximately 4 h to produce one simulation (Van Bergeijk et al., 2022). Additionally, the output of the WOSM is easily post-processed, as it consists of consistent two-dimensional grids over time that can be processed using programming software such as Python. This structured format reduces both complexity and storage, requiring around 400 MB per simulation when stored as 16-bit float, which slightly decreases numerical precision but has no meaningful impact. In contrast, the raw CFD output can be more complex, often requiring dedicated software for post-processing, and occupies around 30 GB of storage per simulated wave. It is important to note that these differences are partly due to the CFD model providing output for additional variables such as stresses and pressures at intervals of 0.01 s. While this results in a detailed dataset, it also contributes to the computational demand of the CFD model, as small step sizes are often necessary to maintain numerical stability.

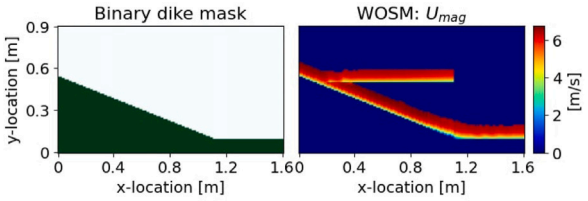

Fig. 18. The WOSM simulation of an overtopping volume $V = 2.0$ m^3/m with slope steepness $s = 1{:}2.5$ at t = 2.5 s. An incorrectly horizontal flow is predicted at x = 0.2 m. This flow stems from an incorrect prediction at a subset boundary.

The increased speed and the reduced storage size become increasingly significant when scaling up to larger experiments. Translating this into the overtopping experiment modelled by Van Bergeijk et al. (2022), the WOSM would require approximately two hours and 330 GB of storage to simulate all 858 waves in the storm. The total computational cost and storage scale linearly with the number of waves, as each individual overtopping wave within the storm is simulated independently by generating a simulation with the corresponding boundary conditions. This results in an equal number of distinct simulations. These conditions are feasible compared to those imposed by the use of CFD models and demonstrate the added value of the WOSM, allowing for the implementation in sensitivity analyses or probabilistic design. Additionally, the WOSM enables easier adjustments to the dike geometry compared to the CFD model, as it uses simple two-dimensional grids such as the binary dike mask, and does not require defining a domain and building the mesh. This makes the WOSM suitable for exploring the effects of updating the bed profile due to erosion on wave overtopping during the modelling of storm events. In future work, the WOSM framework can be extended to account for bed slope variations by coupling the predicted hydrodynamics with an erosion model. Aguilar-López et al. (2018b) and Van Bergeijk et al. (2021) proposed equations to estimate erosion depth based on the combination of load and resistance. The applied load results from the overtopping wave and, depending on the chosen erosion model, can be estimated either by computing stresses using the modelled water depth and flow velocity or by implementing velocity directly into the erosion depth equation. The resistance can be defined as a critical value for applied stress or velocity below which no erosion occurs. Applying this approach across the dike profile gives the resulting cross-profile erosion depth. Bed slope variations can be incorporated into subsequent simulations by modifying the binary dike mask according to the computed erosion depth. This is achieved by setting cells with a value of 1, representing the dike, to 0, representing air cells. In the current set-up, the grid resolution is 1 cm x 1 cm, which means erosion can only be included with steps of 1 cm. A possibility is to keep track of

the accumulated erosion until it reaches 1 cm, before implementing it into the binary dike mask. After each update, a new simulation can be generated over the eroded profile. However, this process would require additional training data (see Section 5.3). Although 1 cm steps can be sufficiently fine for many practical applications, the discretization represents a limitation as it restricts the incorporation of arbitrary erosion depths. Nevertheless, the WOSM offers new opportunities for applications such as sensitivity analyses, probabilistic design, and the exploration of bed slope variations, which are difficult to achieve with the currently available methods. The WOSM framework can provide a solution to the lack of spatiotemporal detail in formulas and the high computational demand of CFD models for large-scale studies.

While the WOSM offers valuable possibilities, it also faces limitations because it lacks any inherent understanding of physics. Unlike in CFD models, where pressures and stresses can be automatically derived, these must be manually computed in the WOSM using Python scripts. In the current set-up, this would require using water depths and flow velocities to calculate the load applied to the bed cover by the overtopping wave. This is a simplified approach compared to CFD models, which can solve the governing equations to directly compute pressures and stresses. The absence of physics also means that not all physical variables are fully represented. For instance, the water fraction is simplified to a binary value in the BWM and values for variables such as Nikuradse roughness height, fluid density and fluid viscosity are fixed by the training data. These can only be changed through retraining the model or modifying the WOSM framework. Also, the WOSM does not explicitly model turbulence, but its effects are embedded in the velocity fields. These velocity fields are generated by a CFD model that used the $k-\omega$ SST turbulence model, meaning the influence of turbulence is reflected in the resulting flow patterns. As a result, the WOSM implicitly learns the impact of turbulence through the velocity inputs, but does not model the turbulence directly. This capability could be added by including turbulence fields in the model input and output, enabling the WOSM to learn and predict turbulence values. Finally, external forces can only be applied at the start of the domain, as the remaining dynamics are processed through WOSM predictions. Overall, the data-driven structure provides speed and simplicity, but reduces versatility by limiting the applicability of the WOSM to the scope of the training data and framework characteristics.

Beyond these conceptual constraints, there can also be practical challenges for broader implementation within hydraulic engineering practices. The nature of the WOSM requires that the input data matches the format of the training data. Users must therefore supply input data in the same preprocessed shape, which can introduce an additional learning requirement to understand and reproduce the preprocessing steps. While the WOSM framework, combining the ViTI2I model with the applied simulation technique, demonstrates a potential method for developing new training procedures and applications, any specifically trained WOSM only understands input data that resembles the training data. This may hinder its direct applicability for new users, as they may first need to learn the preprocessing process before applying the WOSM to new cases.

By transforming the OpenFOAM velocity fields into grids for deep learning, the physical scale can be lost. This means that parameters like the Froude number may no longer be preserved, which is essential in numerical modelling to ensure that the simulation reflects real-world physical behaviour. In this case, the transformation is a preprocessing step to enable a data-driven approach that does not rely on physical laws. The decoupling of dimensional meaning is therefore less relevant, because the WOSM is trained to interpolate learned patterns in the preprocessed data rather than to simulate flow governed by physics. As a result, the WOSM does not incorporate any physical principles to create its predictions. It operates on learned spatiotemporal patterns, so its predictions are not derived from physical laws that ensure real-world behaviour. Instead, its performance is evaluated based on its ability to replicate the hydrodynamic behaviour present in the preprocessed data.

5.3. Deep learning model and training dataset

The final version of the ViTI2I, used in the WOSM for generating the simulations that are validated on hydrodynamics, has 1.6 million trainable parameters and is trained using a training dataset that consists of 443 thousand samples. The amount of training data is likely insufficient to reach an optimally trained model, as the training samples cover 30% of the trainable parameters. For comparison, the original ViT achieved optimal performance when trained on a dataset with a similar number of samples as trainable parameters (Dosovitskiy et al., 2020). Improving the ratio of parameters to samples can enhance the ability of the model to capture the hydrodynamic behaviour of wave overtopping. This is achieved by either reducing the size of the ViTI2I or increasing the dataset size. Reducing the size of the model involves changing the encoder or decoder. The encoder can be reduced by (1) lowering the number of embedding dimensions, (2) reducing the amount of attention computations in the transformer encoder, or (3) decreasing the number of patches by increasing the patch size or reducing the input grid dimensions. However, a lower number of embedding dimensions or a larger patch size can result in worse understanding of local patterns and less attention computations can lead to worse understanding of global patterns. Decreasing the size of the decoder is achieved by using smaller kernels or fewer layers, but this may reduce its ability to interpret the latent space and produce detailed predictions. If the model size is not reduced and the current ViTI2I architecture is retained, approximately 98 simulations with spatial and temporal dimensions comparable to the original dataset are required to reach a balanced parameter-to-sample ratio. In addition, WOSM performance could potentially be improved through hyperparameter tuning of the ViTI2I. This was not performed in the present study, as the delivered WOSM already provided sufficiently accurate results to demonstrate a novel approach to simulating wave overtopping and further optimization for end-product implementation was considered beyond the scope of this research.

Besides increasing the relative number of samples compared to the trainable parameters, the training dataset can also be tailored to further enhance the training process. For instance, the transition from crest to slope is a challenging location to predict well, and the ViTI2I would benefit from a training dataset that displays more combinations of overtopping volumes and slope steepnesses at this transition. This could improve the capacity of the WOSM to predict flow separation accurately, or mitigate the effect of the WOSM increasing the water depth when it simulates a combination of flow separation and alignment. Alternatively, the training dataset can be adapted to match the intended application of the WOSM. While the current dataset by Van Bergeijk et al. (2022) focuses on investigating flow separation, real-world overtopping events more often involve lower volumes and gentler slopes, where separation does not occur. Including more common scenarios, such as dikes with berms, can help the WOSM learn patterns that are more relevant for practical applications. If the WOSM is to be applied in erosion modelling, training data is required that includes wave overtopping simulations over dikes with erosion holes. The current training dataset solely includes dikes with straight slopes, which means the ViTI2I has not been trained to replicate the hydrodynamics of overtopping waves over slopes with erosion holes. For this research, no further expansion of the training dataset was pursued, because the size of the ViTI2I and training dataset were undetermined at the start. Only after the development of the ViTI2I and the initialization of the training dataset parameters were their sizes defined. Therefore, the goal was to develop a method given the available dataset. As the results serve as a proof-of-concept, dataset expansion and refinement is recommended for future research.

Vision Transformer-based models are inherently trained on fixed-size input grids and patch sizes. In this study, a 200 x 200 grid with 10 x 10 patches represents a 2 m x 2 m domain with patches of 0.1 m x 0.1 m . If the spatial resolution of the simulation domain increases, individual cells represent smaller areas, and the same input

grid covers a smaller spatial domain. To maintain the same spatial representation, the input grid must be enlarged, increasing the total number of patches and the computational cost of the attention mechanism. Enlarging the patch size reduces this cost, but maintaining local detail then requires increasing the embedding dimension, leading to a larger model size. This means that although the ViTI2I architecture is scalable, depending on the characteristics of the specific case, it involves trade-offs between attention mechanism computations, local accuracy, and total model size. Additionally, this study did not include an analysis of the impact of varying input grid and patch sizes, and it therefore cannot be concluded that the current setup is optimal for capturing wave overtopping hydrodynamics. However, the selected input grid and patch sizes delivered good performance for demonstrating the feasibility of the WOSM and was therefore adopted. A full sensitivity analysis of multiple ViTI2I architectures involves the time-consuming process of training additional models and was considered beyond the scope of this study, but it remains an important area for future research within the WOSM framework.

The WOSM uses the ViTI2I, which combines a ViT-based encoder with a convolutional decoder for next-frame prediction in wave overtopping simulations. Other deep learning models suited for handling spatiotemporal processes often rely on recurrent and convolutional structures, such as the ConvLSTM (Shi et al., 2015). While no direct comparison is made, the resilience of the ViTI2I against error accumulation is noteworthy and likely a result of its transformer-based architecture. In general, errors are not amplified over space and time, which demonstrates the ability of the ViTI2I to capture and propagate relevant spatiotemporal processes. This can be attributed to the attention mechanism, which links different parts of the input regardless of their distance. As a result, the model appears better able to predict complex phenomena such as flow separation and can adapt to varying conditions. The patch embeddings enable the model to first encode local features, such as different geometries or regions in the domain, before using attention to transfer relevant information across the input sequence. This supports generalization and is expected to offer advantages over ConvLSTM architectures, which process spatial and temporal patterns per time step in a coupled manner and primarily focus on local regions.

6. Conclusions and recommendations

This research proposes the WOSM, a novel method for rapidly generating two-dimensional wave overtopping simulations while preserving detailed velocity patterns. The foundation of the WOSM is a newly developed deep learning model, the ViTI2I. This model is a next-frame prediction model, predicting the next-frame based on an input sequence of previous time step frames. The ViTI2I is an encoder–decoder mechanism, for which an adapted ViT architecture is used to capture spatiotemporal patterns in an input sequence, and deconvolutional layers decode these patterns into accurate predictions.

The ViTI2I is trained with various step sizes and two different loss functions, and two different methods are tested that implement the trained ViTI2I for generating the simulations, resulting in the WOSM framework. The findings are that larger step sizes in the training dataset can create more pronounced patterns in the training dataset of which the ViTI2I can learn, which can enhance the ability of the WOSM to predict challenging locations such as the transition from crest to slope. Applying a mask in the loss function provides control over the training process and forces the ViTI2I to focus on relevant errors within the water body, which increases its performance in predicting velocity values well. Using a simulation technique that retains the downstream part of predictions leads to more accurate simulations as changes upstream are better propagated downstream, such as deceleration and the tail of the wave.

The WOSM requires 3–10 s to generate simulations when using GPU (NVIDIA H100 NVL) and 15–50 s when using CPU (Intel® Xeon® Platinum 8462Y+ with 44 cores). Validating the WOSM against a CFD model indicated good performance capturing the trajectory and shape of the overtopping wave, as well as hydrodynamic variables such as flow velocity, water depth, overtopping duration and vertical flow structure. The WOSM simulations showed anomalies, such as incorrect modelling of flow separation at the transition from crest to slope and the simulation of a horizontally directed flow along the slope. To resolve these issues, it is recommended to develop a tailored training dataset that includes challenging locations, particularly the transition from crest to slope, and to implement an alternative simulation technique that preserves the central part of each individual subset prediction, thereby ensuring that each prediction is based on surrounding context. Additionally, it is recommended to increase the ratio of dataset size to trainable parameters to further enhance the training process.

Within the context of this research, the WOSM provides a framework suitable for experimentation in probabilistic design and sensitivity analyses, due to its computational efficiency, accurate hydrodynamic modelling and the reduced complexity of generating and processing simulations. This research also serves as a proof of concept that the WOSM framework can efficiently generate detailed, complete simulations, which may be valuable in other domains facing similar limitations.

CRediT authorship contribution statement

Wouter P. Schrama: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Conceptualization. **Vera M. van Bergeijk:** Writing – review & editing, Supervision, Resources, Project administration, Conceptualization. **Patricia Mares-Nasarre:** Writing – review & editing, Supervision, Conceptualization. **Joost P. den Bieman:** Writing – review & editing, Supervision, Conceptualization. **Marcel R.A. van Gent:** Writing – review & editing, Supervision, Conceptualization. **Juan P. Aguilar-López:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Juan Pablo Aguilar-Lopez reports financial support was provided by TKI Deltatechnologie. Juan Pablo Aguilar-Lopez reports financial support was provided by Rijkswaterstaat. Given his role as Editor-in-Chief of Coastal Engineering, co-author Marcel R.A. van Gent had no involvement in the peer review of this article and had no access to information regarding its peer review. Full responsibility for the editorial process for this article was delegated to another journal editor. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by TKI Deltatechnologie and Rijkswaterstaat through the ReeinDeeR project.

Appendix A. Maximum flow velocity profile

Fig. A.19 shows the maximum flow velocity profile modelled by the WOSM and the CFD model. The solid lines show fluctuations, which are due to the applied step size causing the wave to travel further between consecutive frames. Each simulation is evaluated by calculating the RMSE between the smoothed velocity profiles (i.e. the dashed lines).

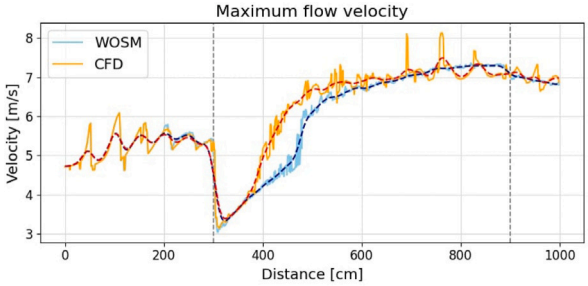

Fig. A.19. The maximum flow velocity profile along the dike profile in the CFD and WOSM simulation of $V = 1.25$ m^3/m with $s = 1{:}1.7$. The dark blue and red dashed lines show the smoothed data. The vertical grey dashed lines indicate the location of transition from crest to slope and the toe of the dike.

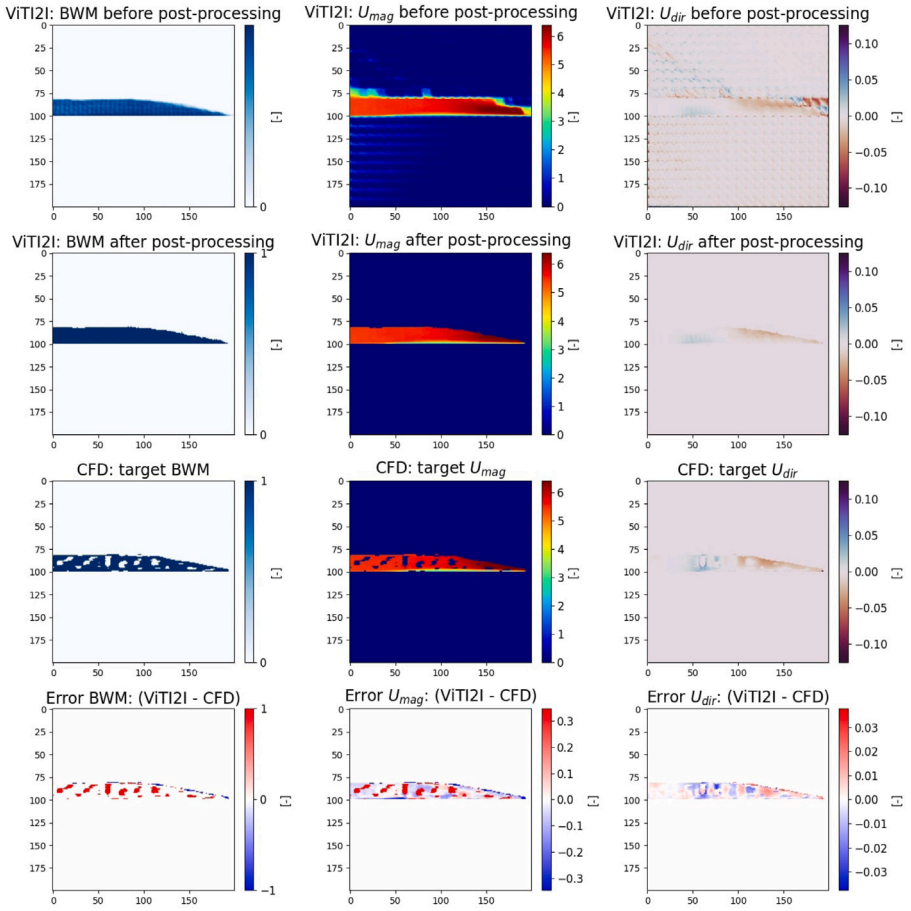

Fig. B.20. The post-processing procedure that transforms raw predictions to the final grids. First, the binary water mask is transformed into probabilities by applying a sigmoid function. The grid is then changed into a binary grid with only ones and zeros. Multiplying the binary water mask with the velocity grids leads to the post-processed velocity grids.

Appendix B. Post-processing predictions

The post-processing procedure is presented in Fig. B.20. The ViTI2I predicts three grids, the BŴM and the two velocity grids. The predicted BŴM contains confidence scores on whether a cell is water or air. As a post-processing step, these confidence scores are converted into probabilities using a sigmoid activation function. Now, values close to 1 indicate the ViTI2I is confident it is water, close to 0 that it is confident it is air and values that are in between 1 and 0 means the model is uncertain. This grid of probabilities is transformed into a binary field by changing all values of 0.5 and larger to 1, and all values below 0.5 to 0. This returns the final post-processed BŴM, which is used to post-process the predicted velocity grids. To ensure that expected air cells are applied zero velocity, the post-processed BŴM is multiplied with the predicted velocity grids. This operation returns all velocity values in expected air cells to zero and retains only velocity values in the water region, completing the final predicted frame $\hat{t}_5$.

Appendix C. Two-dimensional flow field $V = 0.5$ and $s = 1{:}1.7$

The result of the two-dimensional flow field for the simulation of $V = 0.5$ and $s = 1{:}1.7$ is presented in Fig. C.21. The water shape and contours are captured well, although the WOSM again simulates the wavefront with a small delay.

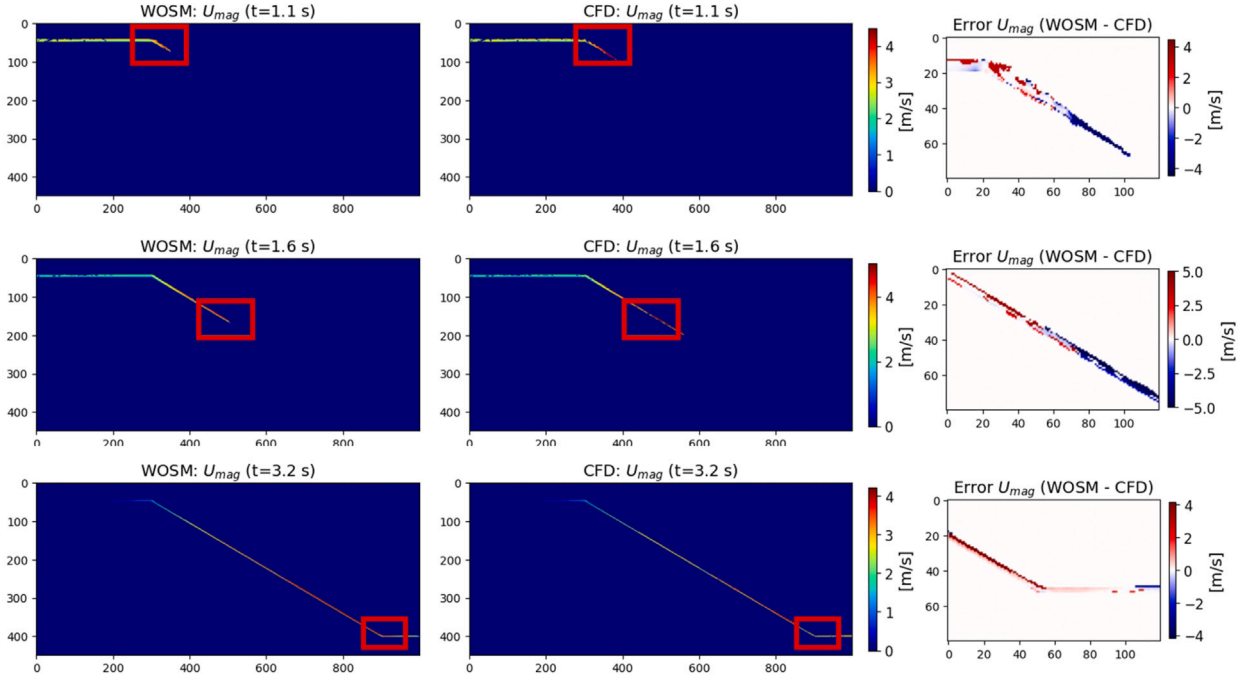

Fig. C.21. Results for U_{mag} as modelled by the WOSM and CFD model for an overtopping wave of volume $V = 0.5$ m^3 over a dike with slope steepness 1:1.7 at t = 1.1 s, 1.6 s and 3.2 s. The red-outlined areas are used to calculate the error by subtracting the CFD field from the WOSM field, as shown in the right-hand column. Dark blue areas indicate areas where the CFD model simulates water, but the WOSM models air. Dark red areas indicate areas where the WOSM models water, but the CFD model simulates air.

Data availability

OpenFOAM data will be made available on request, contact Vera van Bergeijk. The sharing of the Python code can be coordinated through Juan Pablo Aguilar-López.

References

Aguilar-López, J., Warmink, J.J., Bomers, A., Schielen, R.M., Hulscher, S.J.M.H., 2018b. Failure of grass covered flood defences with roads on top due to wave overtopping: A probabilistic assessment method. J. Mar. Sci. Eng. 6 (3), 74.

Aguilar-López, J., Warmink, J., Schielen, R., Hulscher, S.J.M.H., 2018a. Piping erosion safety assessment of flood defences founded over sewer pipes. Eur. J. Environ. Civ. Eng. 22 (6), 707–735. http://dx.doi.org/10.1080/19648189.2016.1217793.

Bertasius, G., Wang, H., Torresani, L., 2021. Is space-time attention all you need for video understanding? In: ICML. Vol. 2, p. 4.

Bomers, A., Aguilar Lopez, J., Warmink, J., Hulscher, S.J.M.H., 2018. Modelling effects of an asphalt road at a dike crest on dike cover erosion onset during wave overtopping. Nat. Hazards 93, 1–30.

Buscombe, D., Carini, R.J., Harrison, S.R., Chickadel, C.C., Warrick, J.A., 2020. Optical wave gauging using deep neural networks. Coast. Eng. 155, 103593. http://dx.doi.org/10.1016/j.coastaleng.2019.103593.

Danka, J., Zhang, L., 2015. Dike failure mechanisms and breaching parameters. J. Geotech. Geoenvironmental Eng. 141 (9), 04015039.

Den Bieman, J., Van Gent, M., Van den Boogaard, H.F., 2021. Wave overtopping predictions using an advanced machine learning technique. Coast. Eng. 166, 103830.

Den Bieman, J., Wilms, J.M., van den Boogaard, H.F., Van Gent, M., 2020. Prediction of mean wave overtopping discharge using gradient boosting decision trees. Water 12 (6), 1703.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al., 2020. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.

Forrester, A., Sobester, A., Keane, A., 2008. Engineering Design Via Surrogate Modelling: a Practical Guide. John Wiley & Sons.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press, http://www.deeplearningbook.org.

Guastoni, L., Güemes, A., Ianiro, A., Discetti, S., Schlatter, P., Azizpour, H., Vinuesa, R., 2021. Convolutional-network models to predict wall-bounded turbulence from wall quantities. J. Fluid Mech. 928, A27.

Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., Tang, Y., Xiao, A., Xu, C., Xu, Y., et al., 2022. A survey on vision transformer. IEEE Trans. Pattern Anal. Mach. Intell. 45 (1), 87–110.

Hughes, S., 2011. Adaptation of the Levee Erosional Equivalence Method for the Hurricane Storm Damage Risk Reduction System (HSDRRS). Technical Report, U.S. Army Corps of Engineers.

IPCC, 2019. Special Report on the Ocean and Cryosphere in a Changing Climate. Chapter 4: Sea Level Rise and Implications for Low-Lying Islands, Coasts and Communities. Technical Report, Intergovernmental Panel on Climate Change.

Johnson, B.D., Kobayashi, N., Gravens, M.B., 2012. Cross-Shore Numerical Model CSHORE for Waves, Currents, Sediment Transport and Beach Profile Evolution. Technical Report.

Koosheh, A., Etemad-Shahidi, A., Cartwright, N., Tomlinson, R., Van Gent, M., 2024. Wave overtopping layer thickness on the crest of rubble mound seawalls. Coast. Eng. 188, 104441.

Larson, M., Kraus, N.C., 1989. SBEACH: Numerical Model for Simulating Storm-Induced Beach Change. Report 1. Empirical Foundation and Model Development. Technical Report.

Masafu, C., Williams, R., 2024. Satellite video remote sensing for flood model validation. Water Resour. Res. 60 (1), http://dx.doi.org/10.1029/2023WR034545.

Pant, P., Doshi, R., Bahl, P., Barati Farimani, A., 2021. Deep learning for reduced order modelling and efficient temporal evolution of fluid simulations. Phys. Fluids 33 (10).

Prince, S.J., 2023. Understanding Deep Learning. The MIT Press, URL: http://udlbook.com.

Razavi, S., Tolson, B.A., Burn, D.H., 2012. Review of surrogate modeling in water resources. Water Resour. Res. 48 (7).

Schüttrumpf, H., 2001. Wellenüberlaufströmung bei Seedeichen – Experimentelle und theoretische untersuchungen.

Schüttrumpf, H., Oumeraci, H., 2005. Layer thicknesses and velocities of wave overtopping flow at seadikes. Coast. Eng. Coast Eng. 52, 473–495. http://dx.doi.org/10.1016/j.coastaleng.2005.02.002.

Schüttrumpf, H., Van Gent, M., 2004. Wave overtopping at seadikes. In: Coastal Structures 2003 - Proceedings of the Conference. pp. 431–443. http://dx.doi.org/10.1061/40733(147)36.

Shannon, C.E., 1948. A mathematical theory of communication. Bell Syst. Tech. J. 27 (3), 379–423.

Sharp, M., Wallis, M., Deniaud, F., Hersch-Burdick, R., Tourment, R., Matheu, E., Seda-Sanabria, Y., Wersching, S., Veylon, G., Durand, E., et al., 2013. The International Levee Handbook. CIRIA.

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., Woo, W.-c., 2015. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. Adv. Neural Inf. Process. Syst. 28.

Suzuki, T., Altomare, C., Veale, W., Verwaest, T., Trouw, K., Troch, P., Zijlema, M., 2017. Efficient and robust wave overtopping estimation for impermeable coastal structures in shallow foreshores using SWASH. Coast. Eng. 122, 108–123.

Van Bergeijk, V., 2022. Over the Dike Top. Modelling the Hydraulic Load of Overtopping Waves Including Transitions for Dike Cover Erosion (PhD thesis). University of Twente.

Van Bergeijk, V., Verdonk, V., Warmink, J., Hulscher, S.J.M.H., 2021. The cross-dike failure probability by wave overtopping over grass-covered and damaged dikes. Water 13, http://dx.doi.org/10.3390/w13050690.

Van Bergeijk, V., Warmink, J., Hulscher, S.J.M.H., 2020. Modelling the wave overtopping flow over the crest and the landward slope of grass-covered flood defences. J. Mar. Sci. Eng. 8 (489), http://dx.doi.org/10.3390/jmse8070489.

Van Bergeijk, V., Warmink, J., Hulscher, S.J.M.H., 2022. The wave overtopping load on landward slopes of grass-covered flood defences: Deriving practical formulations using a numerical model. Coast. Eng. 171, 104047. http://dx.doi.org/10.1016/j.coastaleng.2021.104047.

Van Bergeijk, V., Warmink, J., Van Gent, M., Hulscher, S.J.M.H., 2019. An analytical model of wave overtopping flow velocities on dike crests and landward slopes. Coast. Eng. 149, 28–38. http://dx.doi.org/10.1016/j.coastaleng.2019.03.001.

Van Der Meer, J., Hardeman, B., Steendam, G., Schüttrumpf, H., Verheij, H., 2010. Flow depths and velocities at crest and landward slope of a dike, in theory and with the wave overtopping simulator. Coast. Eng. Proc. 1 (32), 10.

Van Gent, M., 2002. Wave overtopping events at dikes. In: proceedings of the 28th International Coastal Engineering Conference. Vol. 2, http://dx.doi.org/10.1142/9789812791306_0185.

Van Gent, M., Boogaard, H., Pozueta, B., Medina, J., 2007. Neural network modelling of wave overtopping at coastal structures. Coast. Eng. 54, 586–593. http://dx.doi.org/10.1016/j.coastaleng.2006.12.001.

Vaswani, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterhiner, T., Dehgani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N., 2017. Attention is all you need. Adv. Neural Inf. Process. Syst..

Wang, S., Wang, W., Zhao, G., 2024. A novel deep learning rainfall–runoff model based on transformer combined with base flow separation. Hydrol. Res. 55 (5), 576–594.

Wei, Z., Davison, A., 2022. A convolutional neural network based model to predict nearshore waves and hydrodynamics. Coast. Eng. 171, 104044.

Yan, W., Zhang, Y., Abbeel, P., Srinivas, A., 2021. Videogpt: Video generation using vq-vae and transformers. arXiv preprint arXiv:2104.10157.

Ye, X., Bilodeau, G.-A., 2022. Vptr: Efficient transformers for video prediction. In: 2022 26th International Conference on Pattern Recognition. ICPR, IEEE, pp. 3492–3499.

Yin, L., Wang, L., Li, T., Lu, S., Tian, J., Yin, Z., Li, X., Zheng, W., 2023. U-Net-LSTM: time series-enhanced lake boundary prediction model. Land 12 (10), 1859.