Journal of Hydroinformatics

Journal of Hydroinformatics Vol 25 No 6, 2235 doi: 10.2166/hydro.2023.035

Data-driven and echo state network-based prediction of wave propagation behavior in dam-break flood

Changli Li[a], Zheng Han[a,b,*], Yange Li[a,c], Ming Li[a], Weidong Wang[a,c], Ningsheng Chen[d] and Guisheng Hu[d]

[a] School of Civil Engineering, Central South University, Changsha 410075, China
[b] Hunan Provincial Key Laboratory for Disaster Prevention and Mitigation of Rail Transit Engineering Structures, Changsha 410075, China
[c] The Key Laboratory of Engineering Structures of Heavy Haul Railway, Ministry of Education, Changsha 410075, China
[d] Key Laboratory of Mountain Hazards and Surface Processes, Institute of Mountain Hazards and Environment, Chinese Academy of Sciences, Chengdu 610041, China
*Corresponding author. E-mail: zheng_han@csu.edu.cn

ABSTRACT

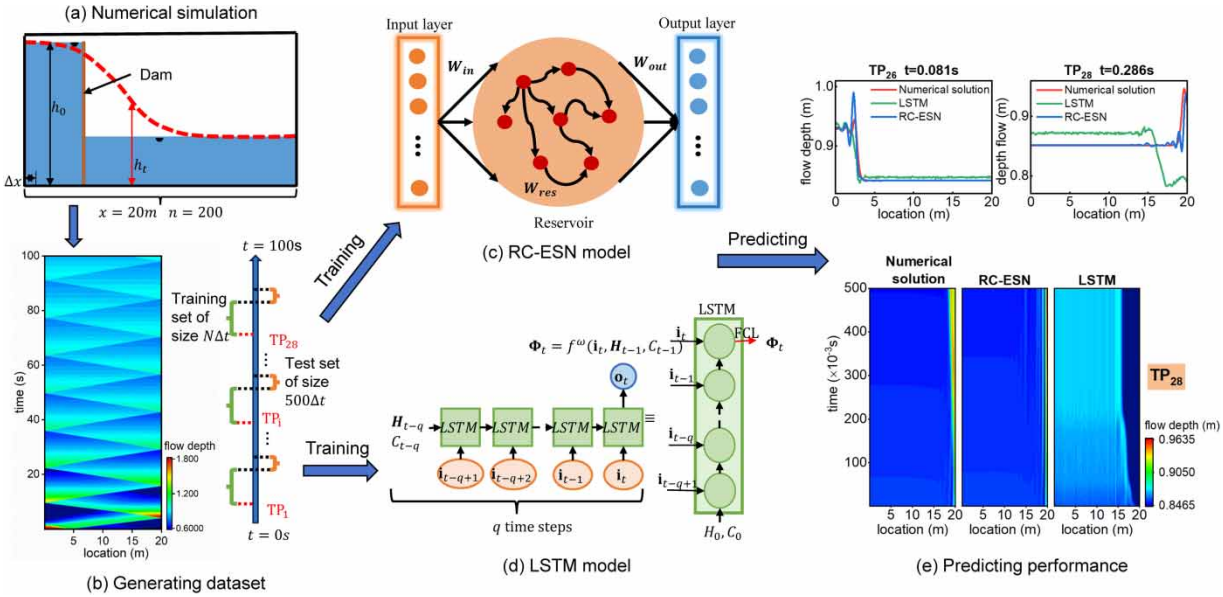
The computational prediction of wave propagation in dam-break floods is a long-standing problem in hydrodynamics and hydrology. We show that a reservoir computing echo state network (RC-ESN) that is well-trained on a minimal amount of data can accurately predict the long-term dynamic behavior of a one-dimensional dam-break flood. We solve the de Saint-Venant equations for a one-dimensional dam-break flood scenario using the Lax–Wendroff numerical scheme and train the RC-ESN model. The results demonstrate that the RC-ESN model has good prediction ability, as it predicts wave propagation behavior 286 time-steps ahead with a root mean square error smaller than 0.01, outperforming the conventional long short-term memory (LSTM) model, which only predicts 81 time-steps ahead. We also provide a sensitivity analysis of prediction accuracy for RC-ESN's key parameters such as training set size, reservoir size, and spectral radius. Results indicate that the RC-ESN is less dependent on training set size, with a medium reservoir size of 1,200–2,600 sufficient. We confirm that the spectral radius has a complex influence on the prediction accuracy and currently recommend a smaller spectral radius. Even when the initial flow depth of the dam break is changed, the prediction horizon of RC-ESN remains greater than that of LSTM.

Key words: dam-break flood, data-driven flow depth prediction, de Saint-Venant equations, echo state network, LSTM model, wave propagation

HIGHLIGHTS

- A machine learning model for predicting wave propagation in dam-break floods is presented.
- The proposed RC-ESN model well predicts wave propagation 286 time-steps ahead.
- The prediction ability of the RC-ESN model significantly outperforms the LSTM model.
- The model is not sensitive to the training dataset size but is influenced by the spectral radius.

GRAPHICAL ABSTRACT



INTRODUCTION

With the incidence of recurring heavy rainfall, snowmelt, and other extreme abnormal weather around the world in recent years, the resulting floods have become increasingly widespread (Wu *et al.* 2014). Dam-break floods have become a very important topic for engineers due to their sudden occurrence, fast expansion, and need for immediate response. In the dynamics of dam-break floods, wave propagation is responsible for catastrophic consequences such as the loss of life and properties in downstream areas (Schubert & Sanders 2012). Therefore, the computational prediction of wave propagation in dam-break floods has long been an issue in hydrodynamics and hydrological engineering practice.

The propagation of water waves in dam-break floods is a dynamic process in which the spatial–temporal variance is usually more essential than the spatial pattern that eventually formed (Li *et al.* 2017). Conventionally, real experience and knowledge, reproduction of historical events, and experiment tests, as well as physical and computer models, can be used to tackle the wave propagation problem in dam-break floods (Aureli *et al.* 2021). A dam-break wave is a flow resulting from a sudden release of a mass of water in a channel. The two-dimensional, depth-averaged Navier–Stokes (N-S) equations based on the shallow water assumption could well characterize wave propagation in a dam-break flood in most instances. For one-dimensional applications, the continuity and momentum conservation in the wave propagation process generates the de Saint-Venant (S-V) equations (Barré de Saint-Venant 1871), which are strongly related to the N-S equations.

The governing laws describing wave propagation in the dam-break flood are expressed in the forms of hyperbolic partial differential equations (PDEs) in either the one-dimensional S-V equations or the two-dimensional simplified N-S equations, requiring an adequate numerical method to solve these PDEs to obtain efficient and accurate results. Up-to-date, the finite differential method (FDM) (Gąsiorowski & Szymkiewicz 2022; Qi *et al.* 2022; Mo *et al.* 2023), finite element method (FEM) (Anisha *et al.* 2023), and finite volume method (FVM) (Hariri *et al.* 2022) have found a large fan community (Seyedashraf & Akhtari 2017). Remarkable studies refer to, e.g., the Lax–Wendroff scheme (Lax & Wendroff 1960) and the leapfrog scheme (Fauzi & Memi Mayasari 2021). Thus far, the traditional PDE-based numerical models have been the main approaches for describing the wave propagation in dam-break floods, with good agreement between the theoretical solution and flume experimental observations. Nevertheless, in order to achieve high accuracy, these modeling techniques necessitate large datasets and frequently necessitate significant tweaking and re-tuning, even with little changes in parameters. To address this issue, some researchers (Seyedashraf *et al.* 2018) are starting to try alternative approaches, rather than solving a problem directly, by offering a desirable balance between accuracy and computational cost.

Recently, artificial intelligence-driven models based on various deep neural networks have had a growing impact on assisting research, such as spatial and temporal forecasting of physical processes through PDE solutions. Due to their superior

nonlinear approximation ability, deep learning neural networks have demonstrated obvious advantages for enhancing the simulation and prediction of nonlinear dynamics systems such as fluid, climate, and dynamical system control (Kutz 2017; Gentine *et al.* 2018; Duraisamy *et al.* 2019; Peters 2019; San *et al.* 2022; Sharma *et al.* 2023). One appeal of the machine learning approach is its capacity to accelerate and enhance the prediction of complicated dynamic systems (Chattopadhyay *et al.* 2020; Drikakis & Sofos 2023). Deep learning neural networks have been proposed for spatio-temporal forecasting and are increasingly employed to assist in modeling chaotic (Pathak *et al.* 2018) and turbulent systems (Ling *et al.* 2016), with promising results (McDermott & Wikle 2017, 2019; Vlachas *et al.* 2018; Raissi *et al.* 2019). The deep learning method will also be employed in this study to make a preliminary attempt at one-dimensional dam-break wave propagation and predict the flow depth directly. Given that this is a preliminary exploration stage, the comparatively basic one-dimensional dam-break problem is studied here, and its trustworthy numerical solution can provide data for this research while reducing the influence of other factors.

Solutions of PDEs for wave propagation take the form of time-series and spatio-temporal data, which include velocities and flow depths. In this regard, the long short-term memory (LSTM) model, which is most suitable for the prediction of sequence data, has attracted a lot of attention in recent years (Goodfellow *et al*. 2016). Srivastava *et al.* (2015) employed a convolution encoder–decoder architecture with an LSTM module to propagate the latent space into the future. Sorteberg *et al.* (2020) used LSTM to construct a wave propagation prediction neural network that could reasonably predict up to 80 time-steps into the future. The LSTM is the most prominent study in overcoming the difficulties that recurrent neural networks (RNNs) have in learning long-term dependence on data as well as gradient disappearance and explosion concerns (Bengio *et al.* 1993; Pascanu *et al.* 2013; Bynagari 2020). A notable study on dam-break floods related to Fotiadis *et al.* (2020) employed LSTM as a baseline for predicting surface wave propagation under two-dimensional images. In this paper, we also use LSTM to compare with the method. In contrast to images, we will directly employ flow depth data for training.

The echo state network (ESN) (Jaeger 2001) is primarily used in this paper to predict the flow depth of a one-dimensional dam-break wave. The ESN, along with the liquid state machines (LSMs) (Maass *et al.* 2002), is referred to as reservoir computing (RC) (Verstraeten *et al.* 2007). The reservoir computing echo state network (RC-ESN) employs a single training procedure rather than a large number of repetitions as in the back-propagation through time (BPTT) algorithm (Werbos 1990). Notably, the reservoir network structure of RC-ESN is connected by a large number of neurons, and the weights of the reservoir connection matrix need to be initialized in advance, enabling the RC-ESN more stable than other neural networks, as substantiated by many prior studies (Skowronski & Harris 2007; Tong *et al.* 2007; Lin *et al.* 2009; Li *et al.* 2012). This can reduce the computational amount of training and avoid the local optimal situation in the optimization algorithm of gradient descent to a certain extent.

Our goal is to use a data-driven model based on the deep learning neural network to predict the spatio-temporal propagation of water waves in a dam-break flood. This research makes three major contributions. Firstly, we propose employing RC-ESN to predict dam-break waves directly based on numerical data, as opposed to using images to predict. Second, we compare RC-ESN to LSTM and discover that RC-ESN enhances long-term forecast accuracy. Finally, we analyze the influence of RC-ESN model parameters and the prediction effect under various initial conditions of dam break.

METHODS

In this paper, the dam-break wave is expressed in the form of the overall flow depth value at a fixed position. The RC-ESN model is used to learn the relationship between the dam-break waves at before and after moments, and this is used to continuously predict the dam-break wave. The LSTM model is used to learn the relationship between the dam-break waves at multiple moments and the next moment and the flow depth of the time series is predicted.

PDEs and the numerical solutions for dam-break flood

In this paper, we focus on wave propagation modeling in a one-dimensional dam-break flood scenario, as depicted in Figure 1(a). This classic case has served as the benchmark test for many numerical studies (Sheu & Fang 2001; Lhomme *et al.* 2010; Han *et al.* 2015; Seyedashraf & Akhtari 2017), offering a good test benchmark for controlled analysis. The de Saint-Venant (S-V) equations are commonly used as the governing equations to describe the wave propagation behavior because they exhibit a simplified mathematical structure without sacrificing the ability to consider smooth flow conditions and flow discontinuities such as hydraulic jumps, moving bores, and wave propagation on dry beds (Cozzolino *et al.* 2015). The frictional force is neglected in the one-dimensional simplified dam-break problem on the horizontal channel,

and the expression is as follows:

$$\begin{cases} \dfrac{\partial}{\partial t}h + \dfrac{\partial}{\partial x}hv = 0 \\ \dfrac{\partial}{\partial t}hv + \dfrac{\partial}{\partial x}\left(hv^2 + \dfrac{gh^2}{2}\right) = 0 \end{cases} \tag{1}$$

where $h(x, t)$ is the flow depth at position x at time t, $g = 9.81\,\mathrm{m/s^2}$ is the gravity acceleration, and $v(x, t)$ is the propagation velocity along the x-direction.

The dam-break flood is simulated in a 20 m long flume in this one-dimensional case, as shown in Figure 1(a). The flume's bottom is horizontal with a layer of 0.6 m thick water covering it initially in the downstream direction. A 1.8 m high dam exists upstream of the flume, filling water on the upstream side. The dam is suddenly removed at the start of the simulation, and the water in the reservoir is released onto the wet bed, generating a dam-break flood downstream.

The S-V equations are hyperbolic PDEs, and the exact theoretical solutions exist for some simple cases (Thacker 1981; Ancey *et al.* 2008). We chose to train the deep neural network with a numerical solution rather than the exact theoretical solution because of its ability to adapt to more complex scenarios in future studies. Therefore, we solve Equation (1) in MATLAB using the Lax–Wendroff method (Lax & Wendroff 1960). The grid number n is set to 200, illustrating that the assumed 20 m long flume is separated into 200 grids of uniform size $\Delta x = 0.1\,\mathrm{m}$. The reflective boundary condition is used on both sides of the flume. The one-dimensional S-V equations are solved with the following initial conditions, that is, the initial flow depth and velocity are:

$$h_0 = \begin{cases} 1.8\,\mathrm{m}, & n \in [1, 44] \\ 0.6\,\mathrm{m}, & n \in [45, 200] \end{cases} \tag{2}$$

$$v_0 = 0 (n \in [1, 200]) \tag{3}$$

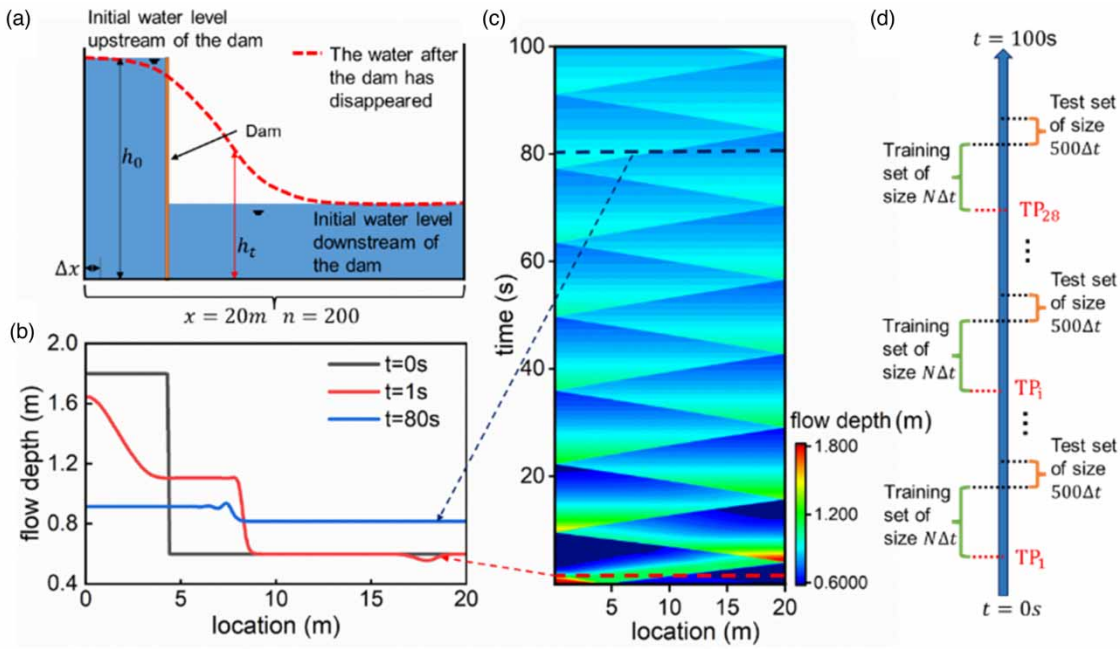


Figure 1 | Linear dam-break with an initially wet bed. (a) The linear dam-break diagram. (b) The section is shown in panel (c) by the dotted line. (b) The flow depth at each flood position at the initial times $t = 0\,\mathrm{s}$, $t = 1\,\mathrm{s}$, and $t = 80\,\mathrm{s}$. (c) Time evolution of flow depth h. The abscissa is the position, the ordinate is the time, and the color bar is the flow depth. (d) The datasets for training and testing. Different datasets generate various TPs.

To improve the simulation result, we set a relatively small time increment $\Delta t = 0.001$ s, which resulted in hundreds of dam-break flood simulations with 100,000 individual time-steps. Simulation results containing $h(x, t)$ of each time step are well recorded (as illustrated in Figure 1(b) and 1(c)). All of the simulated results have been uploaded to the GitHub repository for free access along with this paper.

We sample the flow depth $\mathbf{X} \in \mathbb{R}^{200}$ at each time step Δt to generate the training and test datasets from the above numerical solutions. The dimension of flow depth is 200, as are the dimensions of the network's input and output layers. If the flood dam break is a two-dimensional or three-dimensional scene, the flow depth data need to be integrated into a one-dimensional array according to the method in this paper, which will lead to a sharp increase in the amount of input of these models and affect the effect of these models. We construct a training set consisting of $N = 2{,}000$ sequence samples and a test set of the next 500 sequence samples from $N + 1$ to $N + 500$. In our study, we randomly selected 28 sets of such training/test data, each with a length of $(N + 500)\Delta t$. Different datasets generate different time periods (TPs) for training the machine learning model. Therefore, 28 datasets are trained and tested, namely TP_1, TP_2, …, and TP_{28}, as shown in Figure 1(d) and Table 1.

RC-ESN model

As previously stated, the RC-ESN (Jaeger 2001; Jaeger & Haas 2004) is a typical RC method. It consists of three layers: a randomly generated input layer, a high-dimensional sparse hidden layer, and a learning-specific output layer. Among them, the weights of the input layer and the hidden layer are randomly sampled from a specified distribution and kept fixed in the training stage without learning, whereas the only weight of the output layer that requires learning can be solved simply by the regression method. As the core structure, the hidden layer is commonly referred to as a 'reservoir'. Figure 2(a) depicts the RC-ESN structure diagram. In our study, we express the dam-break wave in the form of the overall flow depth value at the fixed position and use the RC-ESN model to construct the relationship between the dam-break wave at the previous time and at the next time, which is used to predict the dam-break wave at the subsequent time. We arrange the flow depth values at the previous moment in the order of coordinates into an array $\mathbf{X}(t)$ of fixed position and size, and use it as the input of the RC-ESN model, so that the output $\mathbf{X}(t + \Delta t)$ of the RC-ESN model is also an array of fixed size, and each value corresponds to the predicted flow depth at the corresponding position at the next moment.

The reservoir has a size of K and is described by a weighted adjacency matrix $\mathbf{W}_{\text{res}}$, whose largest absolute eigenvalue is the network spectral radius $\rho \leq 1$ (Jaeger 2001). For a given value of ρ, we choose the values of all the elements of $\mathbf{W}_{\text{res}}$ randomly from a uniform distribution $[U(0, 1)]$ and rescale all the values so that its largest eigenvalue is ρ, ensuring that the reservoir system meets the necessary condition of stability, namely echo state property (Jaeger 2001). The reservoir size $K = 1{,}400$ and the spectral radius $\rho = 0.1$ are used in this study to achieve the best prediction performance. Sensitivity analysis of the prediction accuracy concerning K and ρ refers to the following discussion section and Table 1.

The input connection weight matrices $\mathbf{W}_{\text{in}}$ and adjacency matrix $\mathbf{W}_{\text{res}}$ are initialized with random numbers during training, which remain unchanged during training and testing. Only the weight matrix $\mathbf{W}_{\text{out}}$ from the output layer to the reservoir is updated during training. Throughout the training phase, the training datasets of 28 TPs are respectively fed into the following equations (Jaeger 2001; Li *et al.* 2019), according to the chronological order, where Equation (4) is the echo state update

Table 1 | RC-ESN and LSTM model parameters

TP	Training set ($\Delta s = 3,000$)	Test set	RC-ESN	LSTM
TP_i $(i = 1 \sim 28)$	$(i-1)\Delta s\Delta t \sim ((i-1)\Delta s + N)\Delta t$	$((i-1)\Delta s + N)\Delta t \sim ((i-1)\Delta s + N + 500)\Delta t$	$K = 1,400$ $\rho = 0.1$ $\alpha = 0.001$ $\mathbf{e}_0 = 0$	$q = 4$ $l = 2$ $G = 800$
Sensitivity analysis				
TP_{28}			$K = 200 \sim 5,000$ $\rho = 0.01 \sim 1.00$ $\alpha = 0.001$ $\mathbf{e}_0 = 0$	$q = 1 \sim 20$ $l = 1 \sim 6$ $G = 300 \sim 1,300$

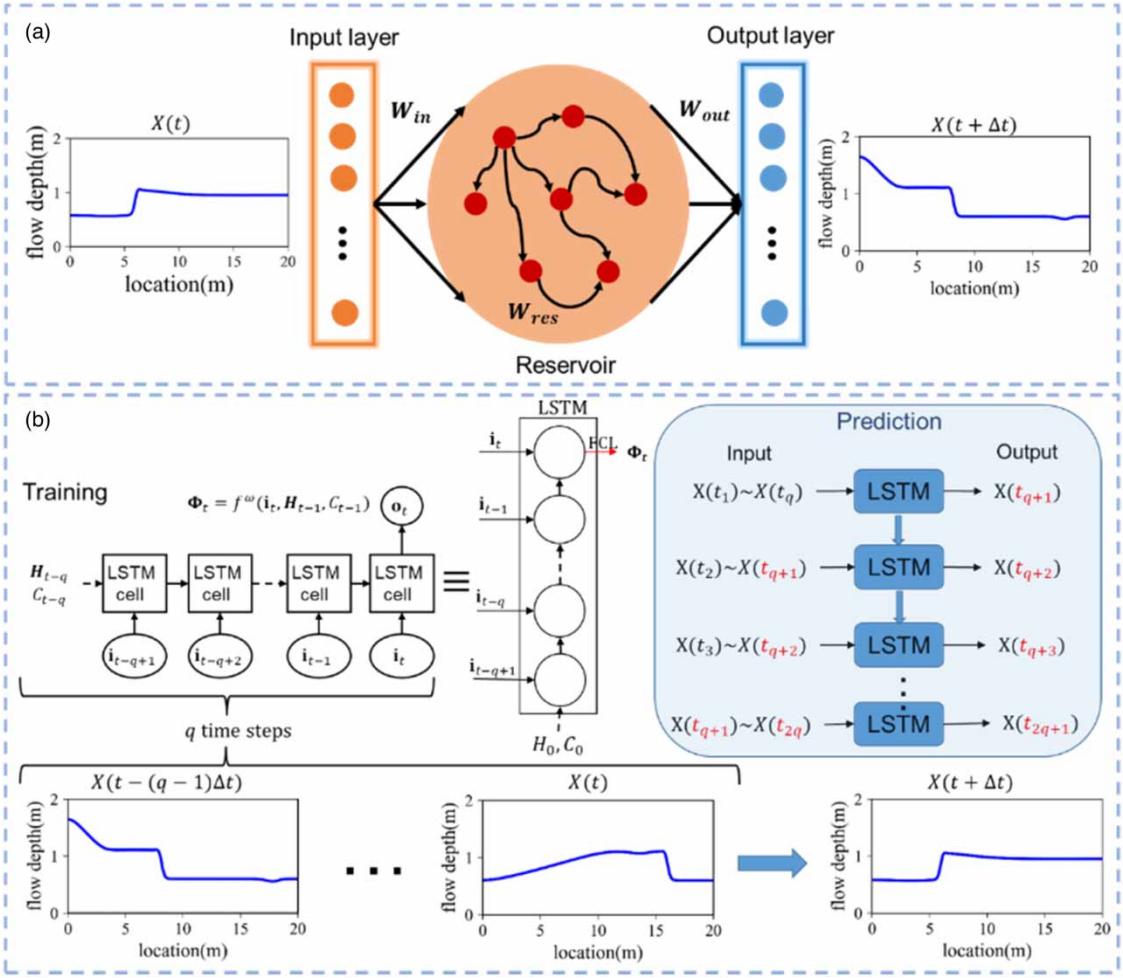


Figure 2 | A schematic diagram of the RC-ESN and LSTM methods for predicting flow depth. (a) Schematic diagram of the ESN method. $\mathbf{X}(t)$ and $\mathbf{X}(t+\Delta t)$ are the input and output of ESN. During training, $\mathbf{X}(t)$ represents the known flow depth at the last moment and $\mathbf{X}(t+\Delta t)$ represents the known flow depth at the next moment. When predicting, $\mathbf{X}(t)$ is the known flow depth or the predicted flow depth at the last moment. (b) An LSTM unfolded q time steps and prediction of flow depth. In the training stage, the flow depth of the previous q time steps $[\mathbf{X}(t-(q-1)\Delta t), \ldots, \mathbf{X}(t-\Delta t), \mathbf{X}(t)]$ is taken as the input, and the output is the flow depth of the next moment $\mathbf{X}(t+\Delta t)$. During prediction, the input can be known or previously predicted flow depth.

formula and the echo state matrix $\mathbf{e}$ is initialized to 0:

$$\mathbf{e}_{t+\Delta t} = f(\mathbf{W}_{\text{res}}\mathbf{e}_t + \mathbf{W}_{\text{in}}\mathbf{X}(t)) \tag{4}$$

$$\mathbf{W}_{\text{out}} = \arg\min_{W_{\text{out}}} \|\mathbf{W}_{\text{out}}\mathbf{e}_t - \mathbf{X}(t)\| + \alpha\|\mathbf{W}_{\text{out}}\| = \mathbf{X}(t)\mathbf{e}_t^T(\mathbf{e}_t\mathbf{e}_t^T + \alpha\mathbf{I})^{-1} \tag{5}$$

where $\mathbf{X}(t) \in \mathbb{R}^{200}$ represents the input of time step t, which is either known initial conditions or predicted. The input vector has the dimension D, where $D = 200$. $\mathbf{e}_t \in \mathbb{R}^K$ represents the echo state generated by the RC-ESN at time step t. $\mathbf{W}_{\text{res}} \in \mathbb{R}^{K\times K}$, $\mathbf{W}_{\text{in}} \in \mathbb{R}^{K\times D}$, and $\mathbf{W}_{\text{out}} \in \mathbb{R}^{D\times K}$ are the weights of the circular connection, input connection, and output connection, respectively. $f(\cdot)$ represents the activation function (e.g., $\tanh(\cdot)$) of the reservoir. $\mathbf{W}_{\text{out}}$ can be calculated using the regression scheme described in Equation (5). Here, $\|\cdot\|$ is the vector's L_2-norm and α is the L_2 regularization constant, which is 0.001.

The calculated $\mathbf{W}_{\text{out}}$ and $\mathbf{e}$ continuously updated with the predicted $\mathbf{X}$ express the relationship between the dam-break waves at two consecutive moments. The output of the prediction process is expressed as follows:

$$\mathbf{X}(t+\Delta t) = \mathbf{W}_{\text{out}}\mathbf{e}_{t+\Delta t} \tag{6}$$

According to $\mathbf{W}_{\text{out}}$ and $\mathbf{e}_{t+\Delta t}$ calculated in the training phase, the first flow depth value of the test can be predicted, and then, the predicted results are put into Equations (4) and (6) to predict the flow depth at subsequent moments.

LSTM network

To further support the long-term prediction performance of the proposed RC-ESN model, the LSTM network, commonly used in many up-to-date studies (Alizadeh *et al.* 2021; Hayder *et al.* 2022), is selected as a comparison.

The input of LSTM is a column of the time-delay-embedded matrix $\mathbf{I}_t$ of $\mathbf{X}(t)$ with the embedding dimension q, commonly known as lookback (Kim *et al.* 1999). In the training dataset of 28 TPs, we also express the dam-break wave in the form of the overall flow depth value at the fixed position and combine the flow depth of q consecutive time steps and the flow depth of one subsequent time step into a training data, where the consecutive q flow depths are used as the input of LSTM, and the output is the flow depth of the subsequent time step. In our study, the time-delay-embedded matrix has a dimension of $(200 \times q) \times M$, where M is the column number of N sequence samples $\mathbf{X} \in \mathbb{R}^{200}$ to form the delay matrix. The LSTM is calculated for forwarding propagation from a specific initial hidden state $\mathbf{H}_0$ and an initial cell state $\mathbf{C}_0$. Its cell state $\mathbf{C}_t \in \mathbb{R}^G$ and the hidden state $\mathbf{H}_t \in \mathbb{R}^G$ update formulas are presented below (Hochreiter & Schmidhuber 1997; Chattopadhyay *et al.* 2020):

$$\widetilde{\mathbf{C}_t} = \tanh\left(\mathbf{W}_C[\mathbf{H}_{t-1}, \mathbf{i}_t] + \mathbf{b}_C\right) \tag{7}$$

$$\mathbf{g}_t^f = \sigma(\mathbf{W}_f[\mathbf{H}_{t-1}, \mathbf{i}_t] + \mathbf{b}_f) \tag{8}$$

$$\mathbf{g}_t^i = \sigma(\mathbf{W}_i[\mathbf{H}_{t-1}, \mathbf{i}_t] + \mathbf{b}_i) \tag{9}$$

$$\mathbf{C}_t = \mathbf{g}_t^f\mathbf{C}_{t-1} + \mathbf{g}_t^i\widetilde{\mathbf{C}_t} \tag{10}$$

$$\mathbf{g}_t^o = \sigma(\mathbf{W}_o[\mathbf{H}_{t-1}, \mathbf{i}_t] + \mathbf{b}_o) \tag{11}$$

$$\mathbf{H}_t = \mathbf{g}_t^o\tanh(\mathbf{C}_t) \tag{12}$$

where $\mathbf{g}_t^f, \mathbf{g}_t^i, \mathbf{g}_t^o \in \mathbb{R}^{G\times(G+B)}$ are the gate signals (forget, input, and output gates), and σ represents the Sigmoid$(\cdot)$ activation function, which determines the weight ranging between 0 and 1. $\mathbf{i}_t \in \mathbb{R}^B$ is the column of the time-delay-embedded matrix $\mathbf{I}_t$ and represents the input of time step t. B is the dimension of the input vector, which is $(200 \times q)$. $\mathbf{W}_C, \mathbf{W}_f, \mathbf{W}_i, \mathbf{W}_o \in \mathbb{R}^{G\times(G+B)}$ and $\mathbf{b}_C, \mathbf{b}_f, \mathbf{b}_i, \mathbf{b}_o \in \mathbb{R}^G$ represent the weight and bias, respectively. The state's dimension G is the number of hidden units, controlling the cell's ability to encode historical information (Vlachas *et al.* 2018).

Assume that $\mathbf{\Phi}_t$ is the output of LSTM, because the output in this study is supposed to have a specified dimension of $D = 200$, we add a fully connected layer (FCL) $\mathbf{W}_{oh} \in \mathbb{R}^{D\times G}$ with no activation function, as shown in Equation (13). We use the stateless LSTM, implying that the hidden and cell states of the LSTM are updated at the start of each batch of training. By expanding the LSTM of the previous q time step and ignoring dependencies greater than q, Equation (14) can be obtained:

$$\mathbf{\Phi}_t = \mathbf{W}_{oh}\mathbf{H}_t = f^{\omega}(\mathbf{i}_t, \mathbf{H}_{t-1}, \mathbf{C}_{t-1}) \tag{13}$$

$$\mathbf{X}(t+\Delta t) \approx \mathbf{\Phi}_t = \mathcal{F}^{\omega}(\underbrace{\mathbf{i}_t, \mathbf{i}_{t-1}, \ldots, \mathbf{i}_{t-q+1}}_{\mathbf{i}_{t:t-q+1}}, \mathbf{H}_{t-q}, \mathbf{C}_{t-q}),\ \mathbf{h}_{t-q} = 0,\ \mathbf{C}_{t-q} = 0 \tag{14}$$

where $\mathcal{F}^{\omega}$ represents the iterative application of f^{ω} and computation of the LSTM state for q time steps, as shown in Figure 2(b).

This paper debugged the hidden layer number $l = 1 \sim 6$, the hidden unit number $G = 300 \sim 1,300$, and lookback $q = 1 \sim 20$, as shown in Table 1. When the parameters are the hidden layer number $l = 2$, hidden unit number $G = 800$, and lookback $q = 4$, the prediction error is small, so they are used in this study to obtain LSTM's prediction performance,

as shown in Figure 3(a). LSTM predicts $\mathbf{X}(t+\Delta t)$ based on the pre-q time step of $\mathbf{X}(t)$. The weights $\mathbf{W}_C$, $\mathbf{W}_f$, $\mathbf{W}_i$, $\mathbf{W}_o$, $\mathbf{W}_{oh}$ of LSTM are learned using Adam (Kingma & Ba 2015) optimizer via the BPTT (Werbos 1990) algorithm during training. Here, the training samples in each batch are randomly shuffled to provide an unbiased gradient estimator in the stochastic gradient descent algorithm (Meng *et al.* 2019). During the prediction, the predicted flow depth is then added to the input array for further prediction. For example, $\mathbf{X}(t+\Delta t)$ is predicted using q past observations $[\mathbf{X}(t-(q-1)\Delta t), \ldots, \mathbf{X}(t-\Delta t), \mathbf{X}(t)]$, which can be known or previously predicted. As shown in Figure 2(b), $\mathbf{X}(t_{q+1})$ is predicted by $\mathbf{X}(t_1) \sim \mathbf{X}(t_q)$, then, $\mathbf{X}(t_2) \sim \mathbf{X}(t_{q+1})$ predicts $\mathbf{X}(t_{q+2})$, and so on.

Evaluation method for prediction performance

To ensure that the initial conditions chosen do not influence optimal performance, training/test sets with different initial conditions are sampled from the sample data independently and uniformly. We employ the root mean square error (RMSE) as a comparative measure. It is defined by Equation (15) for each TP:

$$\mathrm{RMSE}_{\mathrm{TP}}(t) = \sqrt{1/n \sum_{i=1}^{n} \left(\mathbf{X}^i_{\mathrm{TP}}(t) - \tilde{\mathbf{X}}^i_{\mathrm{TP}}(t)\right)^2} \tag{15}$$

where i denotes the ith grid, and n indicates the total number of grids. $\mathbf{X}$ is the numerical solution data at time t, and $\tilde{\mathbf{X}}$ is the predicted value at time t using one of the deep learning algorithms. RMSE was calculated for flow depth under different periods at each moment, and an error curve describing the error's evolution over time was obtained.

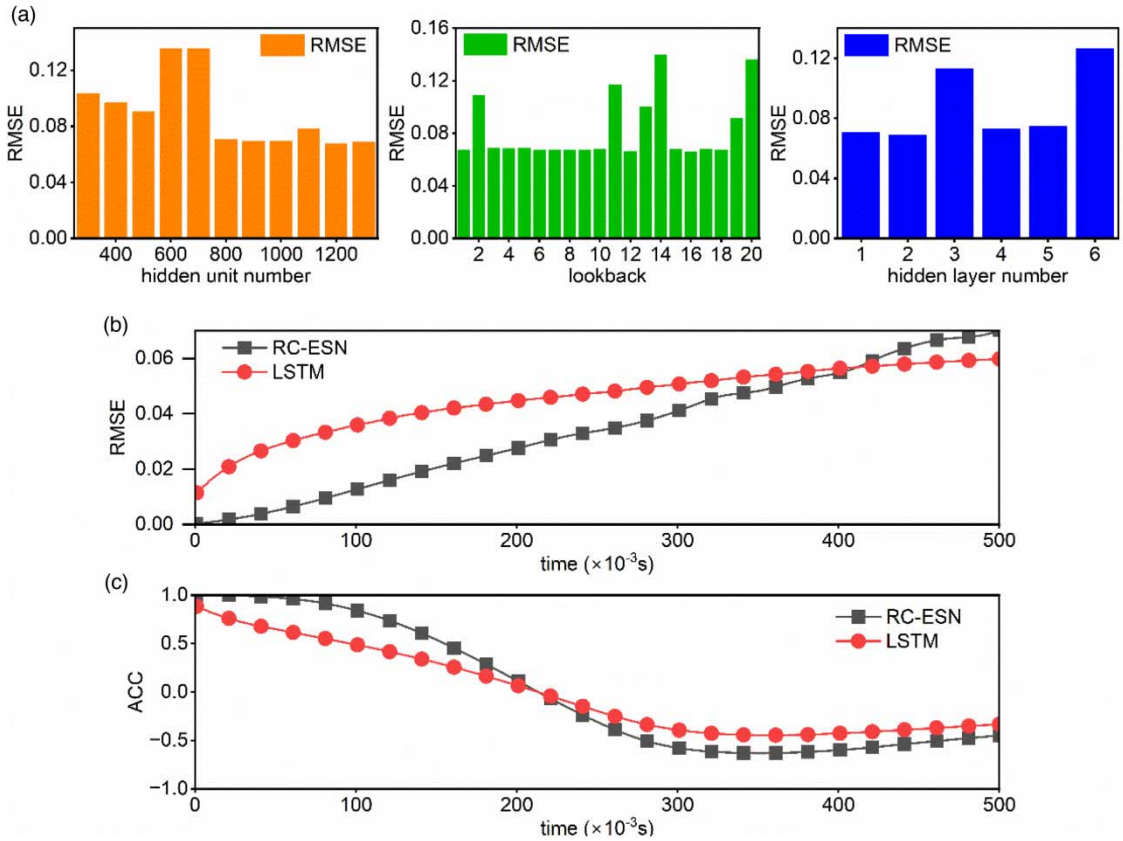


Figure 3 | Comparison of the prediction abilities among the two deep learning methods. (a) The RMSE in different LSTM parameters. (b) The RMSE growth over time for RC-ESN (black) and LSTM (red) under all TPs. (c) The ACC growth over time for RC-ESN (black) and LSTM (red) under all TPs.

Furthermore, the mean anomaly correlation coefficient (ACC) (Allgaier *et al.* 2012) of 28 TPs is used to evaluate the pattern correlation between the predicted and numerical solution. The ACC is defined as:

$$\mathrm{ACC}(t)=\frac{1}{28}\sum_{\mathrm{TP}=1}^{28}\frac{\sum_{i=1}^{n}(\mathbf{X}_{\mathrm{TP}}^{i}(t)-\bar{\mathbf{X}}_{TP}^{i})(\tilde{\mathbf{X}}_{\mathrm{TP}}^{i}(t)-\bar{\mathbf{X}}_{\mathrm{TP}}^{i})}{\sqrt{\sum_{i=1}^{n}(\mathbf{X}_{\mathrm{TP}}^{i}(t)-\bar{\mathbf{X}}_{\mathrm{TP}}^{i})^{2}\sum_{i=1}^{n}(\tilde{\mathbf{X}}_{\mathrm{TP}}^{i}(t)-\bar{\mathbf{X}}_{\mathrm{TP}}^{i})^{2}}} \tag{16}$$

Here, the temporal average of flow depths at each grid is denoted by $\bar{\mathbf{X}}_{\mathrm{TP}}^{i}$. The score ranges from -1.0 to 1.0. If the prediction is accurate, the score equals 1.0.

RESULTS

Considering the different frequencies of flow depth change during the dam-break flood development process, the prediction effect will be influenced by the flow depth at different moments. Therefore, as shown in Figure 1(d), we randomly selected 28 training/test sets from different moments and set the training set size to $N = 2{,}000$. Sensitivity analysis of the prediction accuracy concerning N refers to the following discussion section.

The prediction abilities of RC-ESN and LSTM are compared in Figures 3–5 using the identical training/test set. Figure 3(b) and 3(c) illustrates the mean value of RMSE and ACC over time of RC-ESN and LSTM under 28 distinct randomly and uniformly selected TPs, respectively, to strengthen the comparison between the two methods and to avoid the influences by TPs. The results show that as prediction steps are increased, the error of RC-ESN and LSTM increases progressively, and the model correlation between the predicted results and the simulation solution decreases continuously. RC-ESN outperforms LSTM in prediction performance in terms of prediction error and correlation between prediction results and numerical solution.

The RMSE gradually increases as the prediction time step increases. To display the prediction effect more intuitively, we refer to the prediction time step when RMSE reaches 0.01 as the prediction horizon, implying that the prediction error is acceptable within the prediction horizon. Figure 4 shows the RMSE values of the prediction and numerical solution as the prediction time step increases under two special TPs, namely the longest and shortest prediction horizons of RC-ESN and LSTM, respectively. Figure 4(a) shows that when the time period is TP_1, the prediction horizons of the RC-ESN and LSTM methods are the shortest, with the prediction horizon of RC-ESN being 49, much better than the LSTM model. When the time period is TP_{28}, the RC-ESN method has the longest prediction horizon of 286, while the LSTM method has the longest prediction horizon of 81 when the time period is TP_{26}. In addition, Figure 4(b) compares RMSE for all

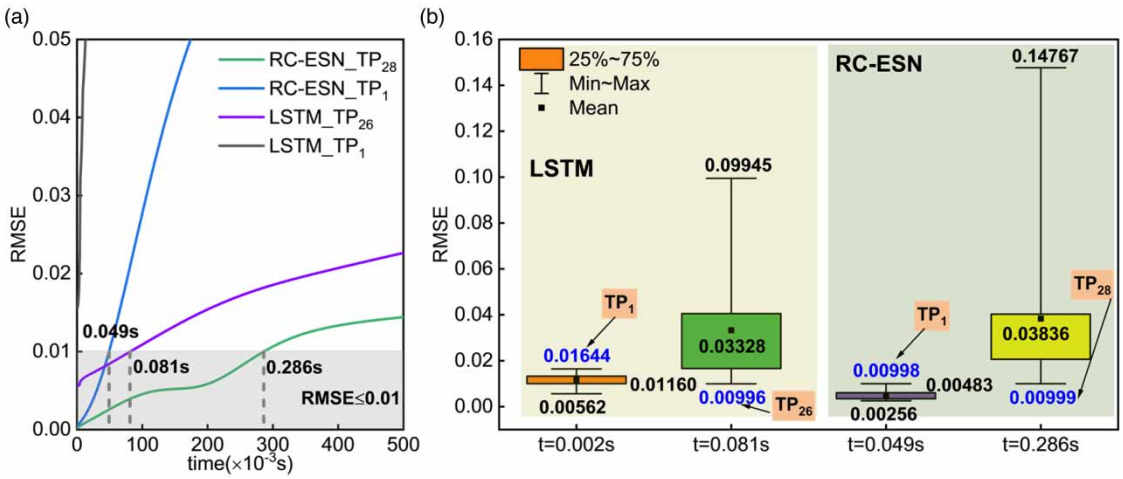


Figure 4 | The RMSE values of the prediction and numerical solution as the predicted time step increases under two special initial conditions. (a) The variation in RMSE with the prediction time step for the TP with the longest and shortest prediction horizons of RC-ESN and LSTM methods. (b) The RMSE at $t = 0.002$ s, $t = 0.081$ s, $t = 0.049$ s, and $t = 0.286$ s under all TPs.

TPs at the corresponding time for the above several prediction horizons. It can be seen that different TPs have a certain influence on the prediction effect, and the longer the prediction time step, the greater the influence.

Figure 5 presents the numerical solution and prediction flow depth of the RC-ESN and LSTM methods for the TPs and prediction horizons mentioned in Figure 4. The figure shows the prediction effect of both models and the flow depth of dam-break waves at two instantaneous time steps, $t = 0.002$ and 0.049 s, during the time period TP_1, when the prediction horizon of both models is the shortest. In this case, it is demonstrated that, despite having the worst prediction effect during such TPs, the RC-ESN model can predict wave propagation behavior because the predicted surface wave well reproduces the numerical solution. The figure also shows the prediction results under the time period TP_{28}, during which the RC-ESN performs best, as well as the flow depth of numerical solution and both models at $t = 0.286$ s. The figures demonstrate that the RC-ESN prediction result is close to the numerical solution at $t = 0.286$ s, whereas the LSTM prediction result obviously deviates. Although the local error fluctuation is observed in the RC-ESN's result at $t = 0.081$ s, the overall RMSE of the RC-ESN model is lower owing to that the dam-break wave is predicted to be stagnant in the LSTM's results during the time period TP_{26}, with which the LSTM achieves the best performance. It should be noted that the error between the numerical solution and the prediction of both models gradually grows as the number of predicted steps increases.

As a conclusion of the result analysis, the RMSE of the RC-ESN reaches 0.01 later than that of the LSTM regardless of the influence of the TPs, demonstrating that the RC-ESN model has a stronger ability for prediction than the conventional LSTM model.

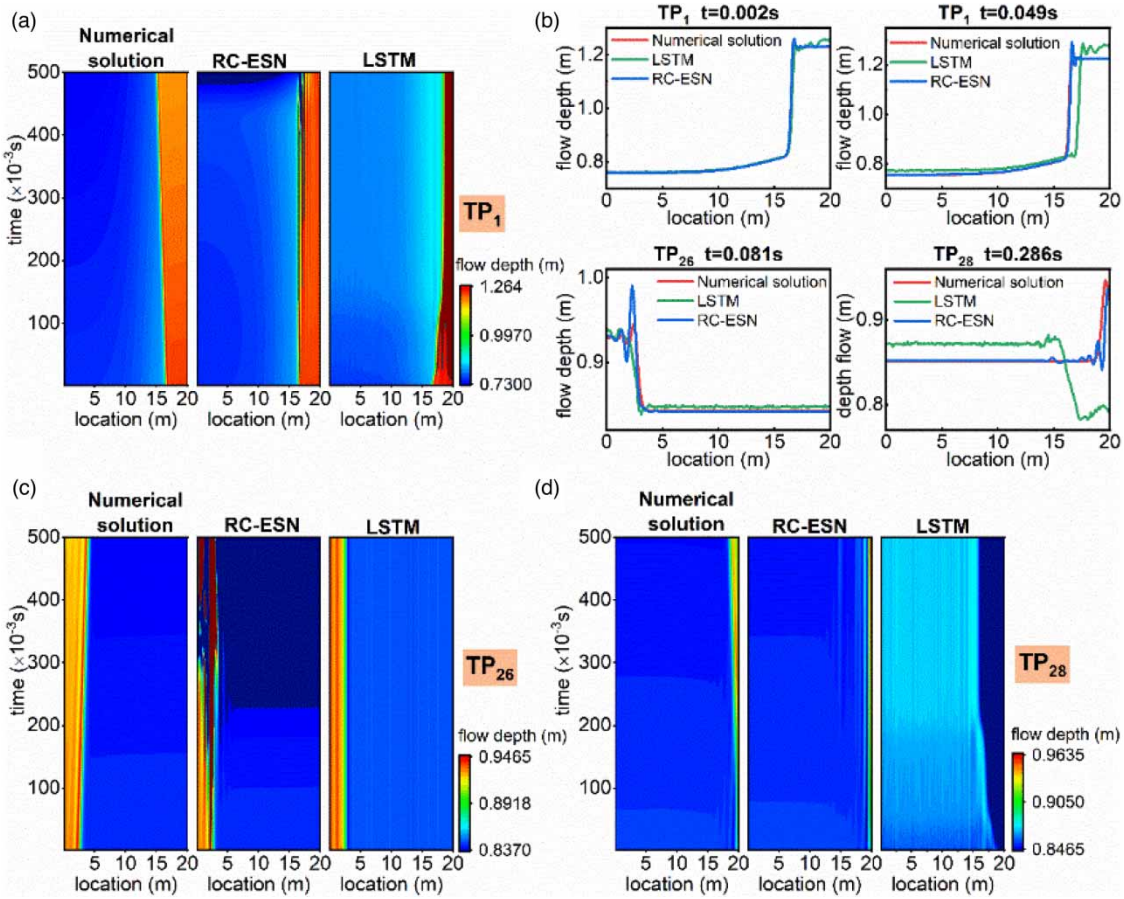


Figure 5 | Performance of RC-ESN and LSTM for spatio-temporal predictions. The numerical solution, RC-ESN, and LSTM predicted flow depth under special TPs in Figure 4 are shown, respectively. (a), (c), and (d) represent the flow depth at all locations and times during time periods TP_1, TP_{28}, and TP_{26}, respectively. The horizontal axis represents the location; the vertical axis represents the time (a total of $500\Delta t$ predicted, $\Delta t = 0.001$ s); and the color bar on the right represents the flow depth, in unit m. (b) The numerical solution and predicted flow depth at each position at $t = 0.001$ s, $t = 0.049$ s, $t = 0.061$ s, and $t = 0.286$ s for time periods TP_1, TP_{26}, and TP_{28}.

DISCUSSION

Comparison of the computational complexity of the RC-ESN and LSTM

Here, we compare the computational complexity of the RC-ESN and LSTM using multiply-accumulate operations (MACs). A MAC contains a multiplication operation and an addition operation, totaling roughly twice the number of floating-point operations (FLOPs). The MACs and the total params of the RC-ESN and LSTM are shown in Table 2 for $l = 2$, $G = 800$, $q = 4$, and $K = 1, 400$. It can be seen that RC-ESN is smaller than LSTM in both the params and MACs, especially the MACs of RC-ESN are three orders of magnitude smaller.

Sensitivity analysis on the training set size N

It is a long-standing issue in deep learning that how the model performance expands with the size of the training set has an important practical implication (Chattopadhyay *et al.* 2020) because the amount of data available for training is closely related to prediction accuracy. The impact of both models, the RC-ESN and the LSTM, on the training set size from $N = 0.1 \times 10^4$ to $N = 8.3 \times 10^4$ is investigated. The prediction horizon, as described above, is extended here, and the prediction error $\overline{\text{RMSE}}$ is defined as the average of RMSE between 0 and 100 prediction time steps ($0\tilde{1}00\Delta t$):

$$\overline{\text{RMSE}} = \frac{1}{100\Delta t} \sum_{i=0}^{i=100} \text{RMSE}(i\Delta t) \tag{17}$$

The prediction error $\overline{\text{RMSE}}$ and the prediction horizon are used to evaluate the predictive abilities of both models. Figure 6(a) shows how the $\overline{\text{RMSE}}$ varies as the size of the training set increases over 100 prediction time steps. It is demonstrated that increasing the training data size is beneficial for improving the prediction ability because the prediction error $\overline{\text{RMSE}}$ of both models decreases as the training size N increases. In general, the RC-ESN model has a lower $\overline{\text{RMSE}}$ than the LSTM method, indicating that the RC-ESN model is less affected by the size of the training set.

Figure 6(b) demonstrates the variation of the prediction horizon against the training set size N. The prediction horizon of the RC-ESN and the LSTM both arise with the growth of training set size N. The RC-ESN model clearly exhibits a positive

Table 2 | Comparison of the computational complexity of the RC-ESN and LSTM

Model	Params	MACs
RC-ESN	0.28M	2.52M
LSTM	8.49M	6.68G

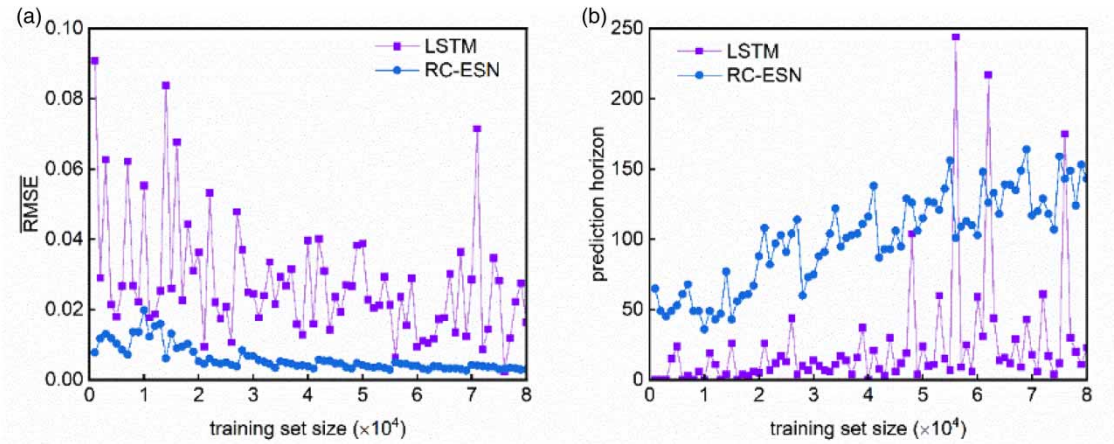


Figure 6 | Comparison of the prediction quality for RC-ESN (blue) and LSTM (purple) as the size of the training set N is changed from $N = 0.1 \times 10^4$ to $N = 8.3 \times 10^4$. (a) The average error $\overline{\text{RMSE}}$ within 100 prediction time steps ($100\Delta t$). (b) The prediction horizon (when RMSE = 0.01).

effect of training set size N on the prediction horizon. The prediction horizon expands to 200 time-steps ahead when the training set size is increased to 8.3×10^4. However, in the case of the LSTM model, this effect is not so obvious, with several exceptional uncertainties existing when the training set size is beyond 4.5×10^4. In general, when compared to the LSTM, the prediction ability and the accuracy of RC-ESN are less dependent on the training set size, which is a significant advantage when the dataset available for training is limited, as in the case of predicting wave propagation in dam-break floods.

Sensitivity analysis on the reservoir size K and spectral radius ρ

The above analysis supports that compared to the conventional LSTM model, the proposed RC-ESN model has witnessed a better performance for predicting wave propagation behavior of the one-dimensional dam-break flood. In this section, we focus on the other major concern of deep learning, i.e., the requirement for large reservoirs, which can be computationally taxing.

In fact, the low-computational consumption requirement has been demonstrated in many previous studies, for example, as a potential disadvantage of ESNs versus LSTMs for training RNNs (Jaeger 2007). In this study, we tested the reservoir sizes K ranging from 200 to 5,000 to evaluate how the prediction error RMSE varies with the number of prediction times. Figure 7 depicts the prediction performance of the RC-ESN as reservoir size K varies. Figure 7(a) shows the prediction horizon of the RC-ESN with the same initial conditions but different reservoir sizes. Figure 7(b) shows the variation of the RMSE with

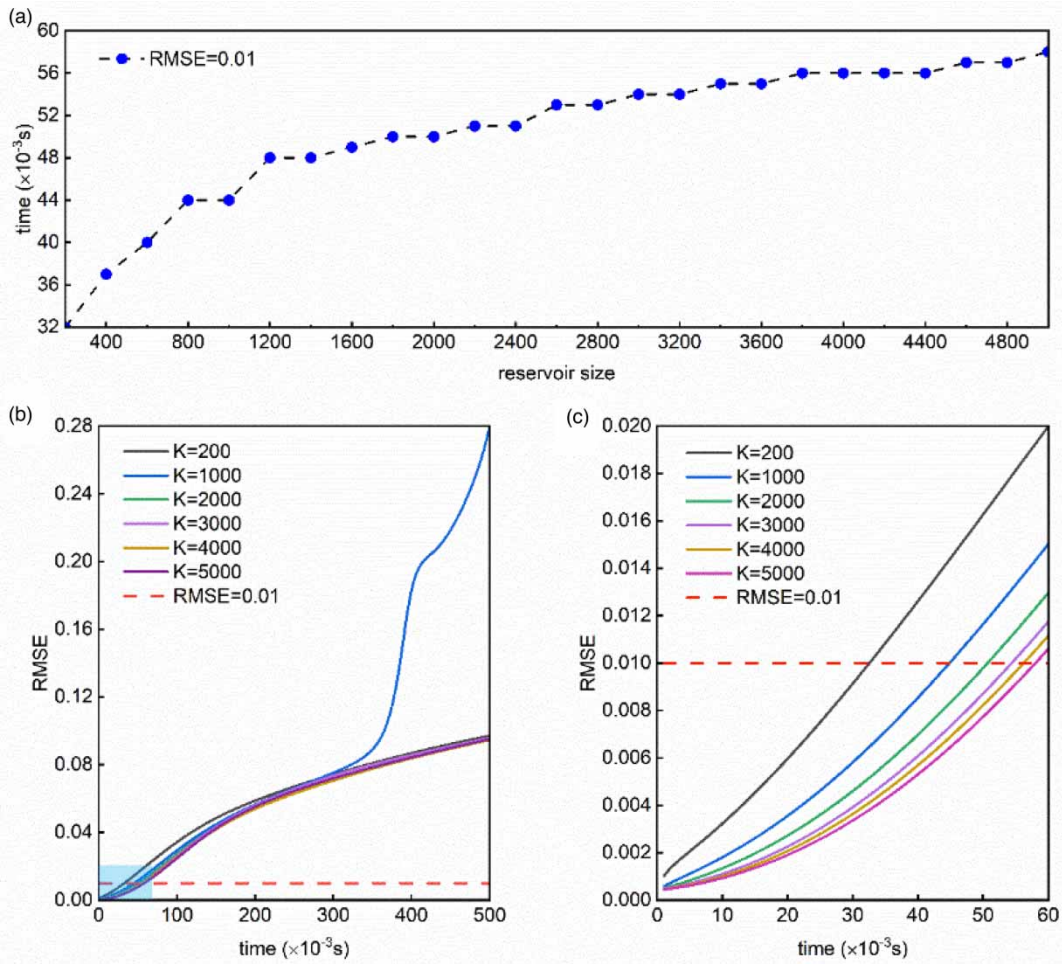


Figure 7 | Scaling of RC-ESN performance with reservoir size K. (a) The prediction horizon when the reservoir size changes from 200 to 5,000 under the same initial conditions. (b) When $K = 200 \sim 5{,}000$, the RMSE of the prediction results under the same initial conditions varies with the prediction time. Panel (c) is the enlargement of the shaded area in the lower-left corner of the panel (b), where the red dotted line is RMSE = 0.01.

predicted time steps, and Figure 7(c) is an expanded view of the shaded region in Figure 7(b). It is obvious that the number of predicted time steps gradually increases with the reservoir size, demonstrating an improvement in the RC-ESN's prediction performance. However, as the reservoir size K increases significantly beyond 2,600, the positive effect on prediction weakens and computational efficiency decreases. In this sense, we conclude that although an increasing reservoir size K is theoretically favorable for improving prediction performance, a medium reservoir size $K = 1{,}200\sim2{,}600$ is sufficient, taking into account the overall balance of computational efficiency and prediction performance.

In addition to the reservoir size K, we also explore the influence of the spectral radius ρ. In the study by Yildiz *et al.* (2012), it was pointed out that there are no generally applicable recipes for the optimal setting of the spectral radius, and an appropriate setting of the spectral radius still has to be found by task-specific experimentation. Some studies have uncovered the emergence of an interval ('valley') in the spectral radius of the neural network in which the prediction error is minimized, and such an interval appears for a variety of spatio-temporal dynamical systems described by nonlinear PDEs, regardless of the structure and edge-weight distribution of the underlying reservoir network (Jiang & Lai 2019). In the following analysis, the best-fitting reservoir size $K = 1{,}400$ is used. We tested different spectral radius ρ values ranging from 0.01 to 1.00 over the identical time period TP_{28}, then compared their effects to the predicted results. Figure 8(a) shows the RMSE variation in different prediction time steps. It is found that the influence of spectral radius is limited

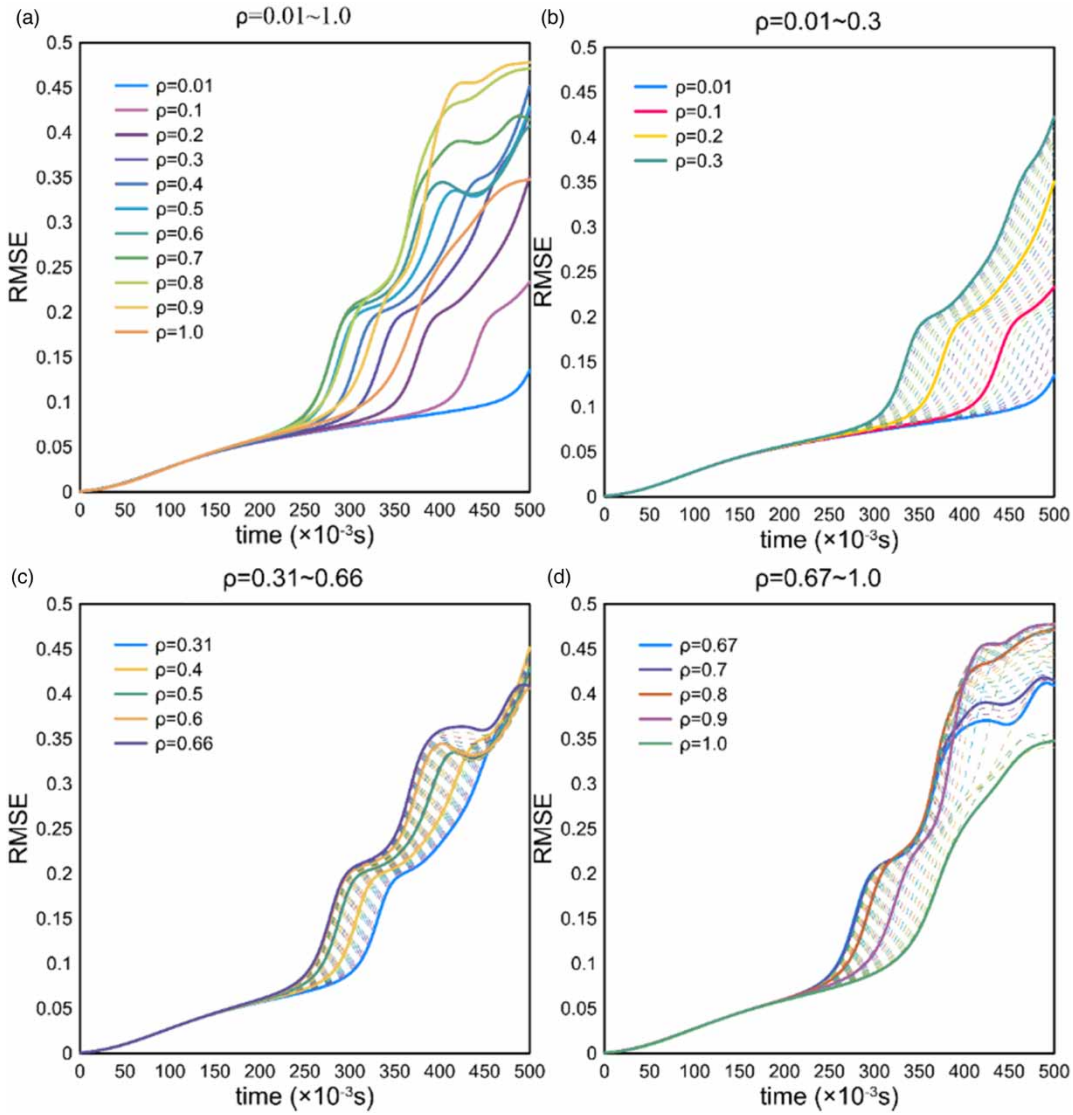


Figure 8 | Scaling of RC-ESN performance with spectral radius ρ. The spacing of dotted lines is $\rho = 0.01$.

for the cases of predicted time steps less than 200. However, as shown in the figure, the curves begin to disperse from each other once the predicted time step exceeds 200 steps, demonstrating that significant differences in RMSE appear. As the curves in Figure 8(a) are mixed, to better explore the influence of the spectral radius ρ, we separately display the curves of $\rho = 0.01 \sim 0.3$, $\rho = 0.31 \sim 0.66$, and $\rho = 0.67 \sim 1.0$ in the panels of Figure 8(b)–8(d), respectively. As illustrated in Figure 8(b) and 8(c), when $\rho = 0.01 \sim 0.66$, the RMSE value of the predicted results increases. However, when spectral radius ρ is continuously increased, $\rho = 0.67 \sim 1.0$, a reverse trend is observed, with the RMSE diminishing progressively as the spectral radius increases as demonstrated in Figure 8(d). This phenomenon implies that the spectral radius ρ has a complex influence on the prediction performance of the RC-ESN model, which deserves ongoing studies, particularly for the tasks of wave propagation prediction in dam-break floods in the future. In general, considering that $\rho = 0.01$ produces the best result for controlling RMES accumulation in Figure 8(a), we currently recommend using a smaller spectral radius ρ for better prediction performance.

Dam-break wave prediction under different initial flow depths

The method proposed in this paper is used to train and predict the propagation of dam-break waves with initial flow depths h_0^1 and h_0^2, respectively. Other parameters are consistent with those described above:

$$h_0^1 = \begin{cases} 1.5\,\text{m}, & n \in [1, 28] \\ 0.8\,\text{m}, & n \in [29, 200] \end{cases} \tag{18}$$

$$h_0^2 = \begin{cases} 1.7\,\text{m}, & n \in [51, 70] \\ 0.6\,\text{m}, & n \in [1, 50] \cup [71, 200] \end{cases} \tag{19}$$

Figure 9(a) shows the prediction horizons of RC-ESN and LSTM at different TP datasets under initial flow depths h_0^1 and h_0^2, respectively. Figure 9(b) and 9(c) respectively shows the predicted flow depths of LSTM at the time of maximum prediction horizon under h_0^1 and h_0^2 ($t = 0.052\,\text{s}$ and $t = 0.104\,\text{s}$), as well as the numerical solution and RC-ESN prediction results at this time. Figure 9(d) and 9(e) shows the predicted flow depths of RC-ESN at the time of the maximum prediction horizon under h_0^1 and h_0^2 ($t = 0.162\,\text{s}$ and $t = 0.211\,\text{s}$), as well as the numerical solution and RC-ESN prediction results at this time. As can be seen from Figure 9, after the initial flow depth changes, the prediction horizon of RC-ESN in different TP datasets is larger than that of LSTM.

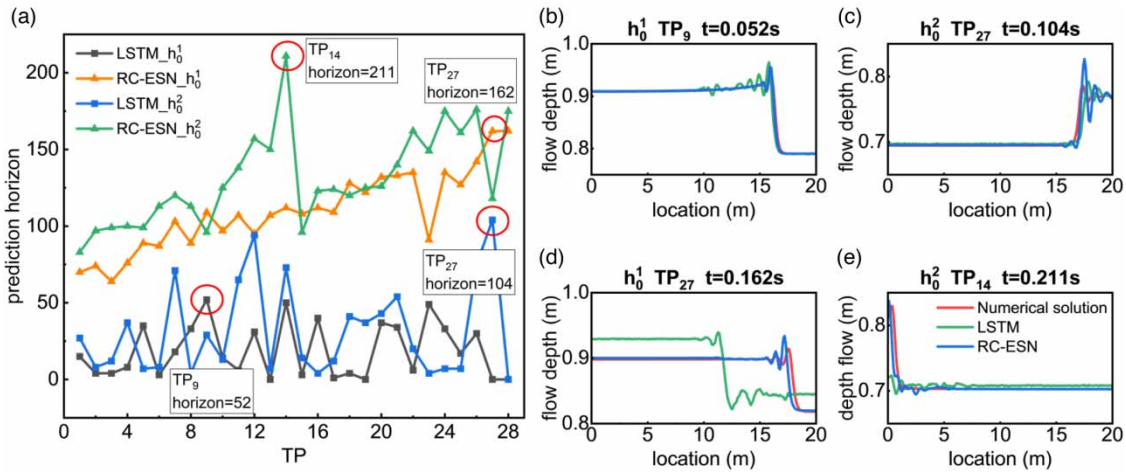


Figure 9 | Performance of RC-ESN and LSTM under different initial flow depths. (a) The prediction horizons of LSTM and RC-ESN on 28 different TP datasets of the initial flow depths h_0^1 and h_0^2, respectively. (b) The flow depth at the time of maximum prediction horizon of LSTM under initial flow depth h_0^1. (c) The flow depth at the time of maximum prediction horizon of LSTM under initial flow depth h_0^2. (d) The flow depth at the time of maximum prediction horizon of RC-ESN under initial flow depth h_0^1. (e) The flow depth at the time of maximum prediction horizon of RC-ESN under initial flow depth h_0^2.

CONCLUSION

In this paper, we show that a data-driven RC-ESN model, well-trained with numerical solutions of de Saint-Venant equations, can predict the long-term dynamic behavior of a one-dimensional dam-break flood with satisfactory accuracy. When compared to the numerical solution, the proposed RC-ESN model achieves the best performance of ahead predicting wave propagation 286 time-steps among all 28 TPs, with RMSE smaller than 0.01. It outperforms the conventional LSTM model, which reaches a comparable RMSE only 81 time-steps ahead.

According to the sensitivity analysis, the proposed RC-ESN model is less dependent on the training set size than the conventional LSTM model, which is a significant advantage when the dataset is available for training. Although increasing reservoir size K is theoretically positive for improving prediction performance, we find that a medium reservoir size $K =$ 1,200~2,600 is sufficient when considering the overall balance of computational efficiency and prediction performance. However, it is confirmed that the spectral radius ρ has a complex influence on the prediction performance of the RC-ESN model. We recommend using $\rho = 0.01$ or a smaller spectral radius ρ for better prediction performance in this one-dimensional dam-break flood scenario. In validating the preceding conclusion, we change the initial flow depth of the dam-break problem and find that the prediction horizon of RC-ESN is larger than that of LSTM in different TPs.

Our findings support the favorable role of the data-driven and RC-ESN model in predicting wave propagation in a one-dimensional dam-break flood. When the change of dam-break waves in a certain period of time is known, this work can quickly predict the evolution of dam-break waves in the following time, which may aid the rapid disaster prediction of dam-break floods. The study in this paper is limited to the one-dimensional ideal dam-break scenario. At the same time, the proposed RC-ESN prediction method has a poor prediction effect over 300 time-steps, and a certain amount of data is still needed for training before the prediction. However, for a two-dimensional or even three-dimensional scenario, which is closer to the real engineering problem, more training data, such as the wave propagation velocities along with different directions, should be included, increasing the difficulty of training a high-performance neural network. This aspect obviously needs more research and should continue to be investigated in order to achieve rapid prediction of dam-break waves in actual scenarios. In the follow-up study, we will consider the flow depth and flow velocity of the dam-break flood in two or three dimensions. Methods such as convolutional neural networks are used to encode and decode the data, and the RC-ESN model is combined for prediction.

ACKNOWLEDGEMENTS

This study was financially supported by the National Natural Science Foundation of China (Grant No. 52078493); the Natural Science Foundation of Hunan Province (Grant No. 2022JJ30700); the Natural Science Foundation for Excellent Young Scholars of Hunan (Grant No. 2021JJ20057); the Innovation Provincial Program of Hunan Province (Grant No. 2020RC3002); the Science and Technology Plan Project of Changsha (Grant No. kq2206014); the Innovation Driven Program of Central South University (Grant No. 2023CXQD033); and the Fundamental Research Funds for the Central Universities of Central South University (Grant No. 2022ZZTS0660). These financial supports are gratefully acknowledged. We also extend our gratitude to associate editor Gwo-Fong Lin and two reviewers for their insightful comments.

DATA AVAILABILITY STATEMENT

All relevant data are available from an online repository or repositories. The data that support the findings of this study are available in the GitHub repository [lcl1527/dam-break-ESN-LSTM], at HYPERLINK "https://github.com/lcl1527/dam-break-ESN-LSTM" lcl1527/dam-break-ESN-LSTM (github.com). The numerical simulation is performed using MATLAB2019. RC-ESN and LSTM training and prediction are performed using PyTorch. The source codes used in this work are freely available online in the GitHub repository: HYPERLINK "https://github.com/lcl1527/dam-break-ESN-LSTM" lcl1527/dam-break-ESN-LSTM (github.com).

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Alizadeh, B., Ghaderi Bafti, A., Kamangir, H., Zhang, Y., Wright, D. B. & Franz, K. J. 2021 A novel attention-based LSTM cell post-processor coupled with Bayesian optimization for streamflow prediction. *J. Hydrol. (Amst.)* **601**, 126526. https://doi.org/10.1016/j.jhydrol.2021.126526.

Allgaier, N. A., Harris, K. D. & Danforth, C. M. 2012 Empirical correction of a toy climate model. *Phys. Rev. E* **85**, 026201. https://doi.org/10.1103/PhysRevE.85.026201.

Ancey, C., Iverson, R. M., Rentschler, M. & Denlinger, R. P. 2008 An exact solution for ideal dam-break floods on steep slopes. *Water Resour. Res.* **44**. https://doi.org/10.1029/2007WR006353.

Anisha, A., Sahu, D. K., Sarkar, P., Mangalathu, S. & Davis, R. 2023 High dimensional model representation for flood fragility analysis of highway bridge. *Eng. Struct.* **281**, 115817. https://doi.org/10.1016/j.engstruct.2023.115817.

Aureli, F., Maranzoni, A. & Petaccia, G. 2021 Review of historical dam-break events and laboratory tests on real topography for the validation of numerical models. *Water (Basel)* **13**, 1968. https://doi.org/10.3390/w13141968.

Barré de Saint-Venant, A. J. C. 1871 Théorie du mouvement non permant des eaux, avec application aux crues des rivières et à l'introduction des marées dans leur lit. *C. R. Acad. Sci.* **73**, 147.

Bengio, Y., Frasconi, P. & Simard, P. 1993 Problem of learning long-term dependencies in recurrent networks. In: *IEEE International Conference on Neural Networks*. IEEE, pp. 1183–1188. https://doi.org/10.1109/icnn.1993.298725.

Bynagari, N. B. 2020 The difficulty of learning long-term dependencies with gradient flow in recurrent nets. *Eng. Int.* **8**, 127–138. https://doi.org/10.18034/ei.v8i2.570.

Chattopadhyay, A., Hassanzadeh, P. & Subramanian, D. 2020 Data-driven predictions of a multiscale Lorenz 96 chaotic system using machine-learning methods: Reservoir computing, artificial neural network, and long short-term memory network. *Nonlinear Process Geophys.* **27**, 373–389. https://doi.org/10.5194/npg-27-373-2020.

Cozzolino, L., Cimorelli, L., Covelli, C., della Morte, R. & Pianese, D. 2015 The analytic solution of the Shallow-Water Equations with partially open sluice-gates: The dam-break problem. *Adv. Water Resour.* **80**, 90–102. https://doi.org/10.1016/j.advwatres.2015.03.010.

Drikakis, D. & Sofos, F. 2023 Can artificial intelligence accelerate fluid mechanics research? *Fluids* **8**, 212. https://doi.org/10.3390/fluids8070212.

Duraisamy, K., Iaccarino, G. & Xiao, H. 2019 Turbulence modeling in the age of data. *Annu. Rev. Fluid Mech.* **51**, 357–377. https://doi.org/10.1146/annurev-fluid-010518-040547.

Fauzi, Y. & Memi Mayasari, Z. 2021 Theoretical study of the differential equation concept of tsunami waves using the leapfrog scheme. *J. Phys. Conf. Ser.* **1940**, 012007. https://doi.org/10.1088/1742-6596/1940/1/012007.

Fotiadis, S., Pignatelli, E., Valencia, M. L., Cantwell, C., Storkey, A. & Bharath, A. A. 2020 Comparing recurrent and convolutional neural networks for predicting wave propagation. In: *ICLR 2020 Workshop on Deep Learning and Differential Equations*.

Gąsiorowski, D. & Szymkiewicz, R. 2022 Inverse flood routing using simplified flow equations. *Water Resour. Manage.* **36**, 4115–4135. https://doi.org/10.1007/s11269-022-03244-8.

Gentine, P., Pritchard, M., Rasp, S., Reinaudi, G. & Yacalis, G. 2018 Could machine learning break the convection parameterization deadlock? *Geophys. Res. Lett.* **45**, 5742–5751. https://doi.org/10.1029/2018GL078202.

Goodfellow, I., Bengio, Y. & Courville, A. 2016 *Deep Learning*. The MIT Press, Cambridge, MA. Available at: http://www.deeplearningbook.org (accessed 3 November 2023).

Han, Z., Chen, G., Li, Y., Tang, C., Xu, L., He, Y., Huang, X. & Wang, W. 2015 Numerical simulation of debris-flow behavior incorporating a dynamic method for estimating the entrainment. *Eng. Geol.* **190**, 52–64. https://doi.org/10.1016/j.enggeo.2015.02.009.

Hariri, S., Weill, S., Gustedt, J. & Charpentier, I. 2022 A balanced watershed decomposition method for rain-on-grid simulations in HEC-RAS. *J. Hydroinf.* **24**, 315–332. https://doi.org/10.2166/hydro.2022.078.

Hayder, G., Iwan Solihin, M. & Najwa, M. R. N. 2022 Multi-step-ahead prediction of river flow using NARX neural networks and deep learning LSTM. *H2Open J.* **5**, 43–60. https://doi.org/10.2166/h2oj.2022.134.

Hochreiter, S. & Schmidhuber, J. 1997 Long short-term memory. *Neural Comput.* **9**, 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

Jaeger, H. 2001 The 'echo state' approach to analysing and training recurrent neural networks – with an erratum note. *GMD Rep.* **148**, 13.

Jaeger, H. 2007 Echo state network. *Scholarpedia* **2**, 2330. https://doi.org/10.4249/scholarpedia.2330.

Jaeger, H. & Haas, H. 2004 Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science* **304**, 78–80. https://doi.org/10.1126/science.1091277.

Jiang, J. & Lai, Y. C. 2019 Model-free prediction of spatiotemporal dynamical systems with recurrent neural networks: Role of network spectral radius. *Phys. Rev. Res.* **1**, 033056. https://doi.org/10.1103/PhysRevResearch.1.033056.

Kim, H. S., Eykholt, R. & Salas, J. D. 1999 Nonlinear dynamics, delay times, and embedding windows. *Physica D* **127**, 48–60. https://doi.org/10.1016/S0167-2789(98)00240-1.

Kingma, D. P. & Ba, J. L. 2015 Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*.

Kutz, J. N. 2017 Deep learning in fluid dynamics. *J. Fluid Mech.* **814**, 1–4. https://doi.org/10.1017/jfm.2016.803.

Lax, P. & Wendroff, B. 1960 Systems of conservation laws. *Commun. Pure Appl. Math.* **13**, 217–237. https://doi.org/10.1002/cpa.3160130205.

Lhomme, J., Gutierrez-Andres, J., Weisgerber, A., Davison, M., Mulet-Marti, J., Cooper, A. & Gouldby, B. 2010 Testing a new two-dimensional flood modelling system: Analytical tests and application to a flood event. *J. Flood Risk Manage.* **3**, 33–51. https://doi.org/10.1111/j.1753-318X.2009.01053.x.

Li, D., Han, M. & Wang, J. 2012 Chaotic time series prediction based on a novel robust echo state network. *IEEE Trans. Neural Networks Learn. Syst.* **23**, 787–799. https://doi.org/10.1109/TNNLS.2012.2188414.

Li, X., Li, D. & Liu, X. 2017 Geographical simulation and optimization system (GeoSOS) and its application in the analysis of geographic national conditions. *Cehui Xuebao/Acta Geod. Cartogr. Sinica* **46**. https://doi.org/10.11947/j.AGCS.2017.20170355.

Li, F., Wang, X. & Li, Y. 2019 Effects of singular value spectrum on the performance of echo state network. *Neurocomputing* **358**, 414–423. https://doi.org/10.1016/j.neucom.2019.05.068.

Lin, X., Yang, Z. & Song, Y. 2009 Short-term stock price prediction based on echo state networks. *Expert Syst. Appl.* **36**, 7313–7317. https://doi.org/10.1016/j.eswa.2008.09.049.

Ling, J., Kurzawski, A. & Templeton, J. 2016 Reynolds averaged turbulence modelling using deep neural networks with embedded invariance. *J. Fluid Mech.* **807**, 155–166. https://doi.org/10.1017/jfm.2016.615.

Maass, W., Natschläger, T. & Markram, H. 2002 Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural Comput.* **14**, 2531–2560. https://doi.org/10.1162/089976602760407955.

McDermott, P. L. & Wikle, C. K. 2017 An ensemble quadratic echo state network for non-linear spatio-temporal forecasting. *Stat* **6**, 315–330. https://doi.org/10.1002/sta4.160.

McDermott, P. L. & Wikle, C. K. 2019 Deep echo state networks with uncertainty quantification for spatio-temporal forecasting. *Environmetrics* **30**. https://doi.org/10.1002/env.2553.

Meng, Q., Chen, W., Wang, Y., Ma, Z. M. & Liu, T. Y. 2019 Convergence analysis of distributed stochastic gradient descent with shuffling. *Neurocomputing* **337**, 46–57. https://doi.org/10.1016/j.neucom.2019.01.037.

Mo, C., Shen, Y., Lei, X., Ban, H., Ruan, Y., Lai, S., Cen, W. & Xing, Z. 2023 Simulation of one-dimensional dam-break flood routing based on HEC-RAS. *Front. Earth Sci. (Lausanne)* **10**. https://doi.org/10.3389/feart.2022.1027788.

Pascanu, R., Mikolov, T. & Bengio, Y. 2013 On the difficulty of training recurrent neural networks. In *30th International Conference on Machine Learning, ICML 2013*.

Pathak, J., Hunt, B., Girvan, M., Lu, Z. & Ott, E. 2018 Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach. *Phys. Rev. Lett.* **120**, 024102. https://doi.org/10.1103/PhysRevLett.120.024102.

Peters, T. 2019 Data-driven science and engineering: Machine learning, dynamical systems, and control. *Contemp. Phys.* **60**, 320–320. https://doi.org/10.1080/00107514.2019.1665103.

Qi, Y., Li, Q., Zhao, Z., Zhang, J., Gao, L., Yuan, W., Lu, Z., Nie, N., Shang, X. & Tao, S. 2022 Heterogeneous parallel implementation of large-scale numerical simulation of Saint-Venant equations. *Appl. Sci.* **12**, 5671. https://doi.org/10.3390/app12115671.

Raissi, M., Perdikaris, P. & Karniadakis, G. E. 2019 Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **378**, 686–707. https://doi.org/10.1016/j.jcp.2018.10.045.

San, O., Pawar, S. & Rasheed, A. 2022 Prospects of federated machine learning in fluid dynamics. *AIP Adv.* **12**. https://doi.org/10.1063/5.0104344.

Schubert, J. E. & Sanders, B. F. 2012 Building treatments for urban flood inundation models and implications for predictive skill and modeling efficiency. *Adv. Water Resour.* **41**, 49–64. https://doi.org/10.1016/j.advwatres.2012.02.012.

Seyedashraf, O. & Akhtari, A. A. 2017 Two-dimensional numerical modeling of dam-break flow using a new TVD finite-element scheme. *J. Braz. Soc. Mech. Sci. Eng.* **39**, 4393–4401. https://doi.org/10.1007/s40430-017-0776-y.

Seyedashraf, O., Mehrabi, M. & Akhtari, A. A. 2018 Novel approach for dam break flow modeling using computational intelligence. *J. Hydrol. (Amst.)* **559**, 1028–1038. https://doi.org/10.1016/j.jhydrol.2018.03.001.

Sharma, P., Chung, W. T., Akoush, B. & Ihme, M. 2023 A review of physics-informed machine learning in fluid mechanics. *Energies (Basel)* **16**, 2343. https://doi.org/10.3390/en16052343.

Sheu, T. W. H. & Fang, C. C. 2001 High resolution finite-element analysis of shallow water equations in two dimensions. *Comput. Methods Appl. Mech. Eng.* **190**, 2581–2601. https://doi.org/10.1016/S0045-7825(00)00255-3.

Skowronski, M. D. & Harris, J. G. 2007 Noise-robust automatic speech recognition using a discriminative echo state network. In: *2007 IEEE International Symposium on Circuits and Systems*. IEEE, pp. 1771–1774. https://doi.org/10.1109/ISCAS.2007.378015.

Sorteberg, W. E., Garasto, S., Cantwell, C. C. & Bharath, A. A. 2020 Approximating the solution of surface wave propagation using deep neural networks. In: *INNSBDDL 2019: Recent Advances in Big Data and Deep Learning. Proceedings of the International Neural Networks Society*, pp. 246–256. https://doi.org/10.1007/978-3-030-16841-4_26.

Srivastava, N., Mansimov, E. & Salakhutdinov, R. 2015 Unsupervised learning of video representations using LSTMs. In *32nd International Conference on Machine Learning, ICML 2015*.

Thacker, W. C. 1981 Some exact solutions to the nonlinear shallow-water wave equations. *J. Fluid Mech.* **107**, 499. https://doi.org/10.1017/S0022112081001882.

Tong, M. H., Bickett, A. D., Christiansen, E. M. & Cottrell, G. W. 2007 Learning grammatical structure with echo state networks. *Neural Networks* **20**, 424–432. https://doi.org/10.1016/j.neunet.2007.04.013.

Verstraeten, D., Schrauwen, B., D'Haene, M. & Stroobandt, D. 2007 An experimental unification of reservoir computing methods. *Neural Networks* **20**, 391–403. https://doi.org/10.1016/j.neunet.2007.04.003.

Vlachas, P. R., Byeon, W., Wan, Z. Y., Sapsis, T. P. & Koumoutsakos, P. 2018 Data-driven forecasting of high-dimensional chaotic systems with long short-term memory networks. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **474**, 20170844. https://doi.org/10.1098/rspa.2017.0844.

Werbos, P. J. 1990 Backpropagation through time: What it does and how to do it. *Proc. IEEE* **78**, 1550–1560. https://doi.org/10.1109/5.58337.

Wu, H., Adler, R. F., Tian, Y., Huffman, G. J., Li, H. & Wang, J. 2014 Real-time global flood estimation using satellite-based precipitation and a coupled land surface and routing model. *Water Resour. Res.* **50**, 2693–2717. https://doi.org/10.1002/2013WR014710.

Yildiz, I. B., Jaeger, H. & Kiebel, S. J. 2012 Re-visiting the echo state property. *Neural Networks* **35**, 1–9. https://doi.org/10.1016/j.neunet.2012.07.005.

First received 15 February 2023; accepted in revised form 22 October 2023. Available online 2 November 2023