A LSTM surrogate modelling approach for caisson foundations

Pin Zhang [a,b], Zhen-Yu Yin [a,*], Yuanyuan Zheng [c,d], Fu-Ping Gao [e,f]

[a] *Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China*
[b] *Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), 1119, Haibin Rd., Nansha District, Guangzhou, China*
[c] *School of Civil Engineering, Sun Yat-Sen University, Guangzhou 510275, China*
[d] *Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), China*
[e] *Key Laboratory for Mechanics in Fluid Solid Coupling Systems, Institute of Mechanics, Chinese Academy of Sciences, Beijing, 100190, China*
[f] *School of Engineering Science, University of Chinese Academy of Sciences, Beijing, 100049, China*

ARTICLE INFO

ABSTRACT

This study proposes a hybrid surrogate modelling approach with the integration of deep learning algorithm long short-term memory (LSTM) to identify the mechanical responses of caisson foundations in marine soils. The LSTM based surrogate model is first trained based on limited results generated from the SPH-SIMSAND based numerical simulations with a strong validation, thereafter it is applied to predict the mechanical responses of soil-structure interaction and the failure envelope of unknown caisson foundations with various specifications as testing. The results indicate that the LSTM based model is more flexible than macro-element method, because it can directly learn the failure mechanism of caisson foundation from the raw data, meanwhile guarantees a high computational efficiency and accuracy in comparison with physical and numerical modelling. LSTM based surrogated model shows a great potential of application in engineering practice.

1. Introduction

Since the suction caisson was first used as the foundation of an offshore wind turbine (OWT) at Frederikshave, Denmark in 2002 (Ibsen and Brincker, 2004), suction caisson foundations have drawn great attention with the increasing application in the foundation of OWT (Gelagoti et al., 2018; Skau et al., 2018b). The suction caisson is installed by pumping the trapped water within the caisson compartment after it has touched the seabed. Such process does not rely on any specialist equipment, thereby suction caisson has been commonly acknowledged as a cost-effective and eco-friendly foundation mode (Jin et al., 2019c; Zhu et al., 2018).

Numerous research works have been conducted to investigate the responses of caisson foundations to the couplings between the vertical force, the horizontal force and the bending moment through, e.g. in-situ testing (Houlsby et al., 2006), physical modelling (Byrne and Houlsby, 2001; Cassidy et al., 2002; Ibsen et al., 2014), numerical modelling (Skau et al., 2018a; Zafeirakos and Gerolymos, 2016) and analytical solutions (Li et al., 2015; Montrasio and Nova, 1997; Nova and Montrasio, 1991). Simple physical model is hard to simulate the in-situ operation condition of caisson foundation and obtain its failure mechanism; meanwhile the instrument tends to be cumbersome and

expensive. Numerical methods have thus been extensively employed to simulate the responses of caisson foundations (Jin et al., 2018; Jin et al., 2019b; Liu et al., 2014), but the nonlinear and elaborate finite element modelling is time-consuming and requires considerable skill (Jin et al., 2019c), which is suitable for a specific case. An analytical solution known as macro-element that derives from the experimental or numerical results has been proposed to explore the failure envelope and applied in engineering design due to its simplicity. In this method, soil and foundation structure are considered as a macro-element, thereby the computation is faster and simpler in comparison with finite element analysis. However, for different size of caisson foundations, certain experimental tests and numerical simulations are needed to calibrate the macro-element model, which is time-consuming and expensive. Macro-element is also limited by fixed formation and thus cannot totally replicate the results of experimental or numerical results.

An alternative method is to develop a surrogate model that is constructed using limited experimental or numerical results and at the same time it can directly learn the failure mechanism of caisson foundations from the raw experimental or numerical data. Thereafter this surrogate model can be applied to predict the mechanical responses and the failure envelope of a caisson foundation under various conditions such as different foundation size, aspect ratio, soil-structure contact surface
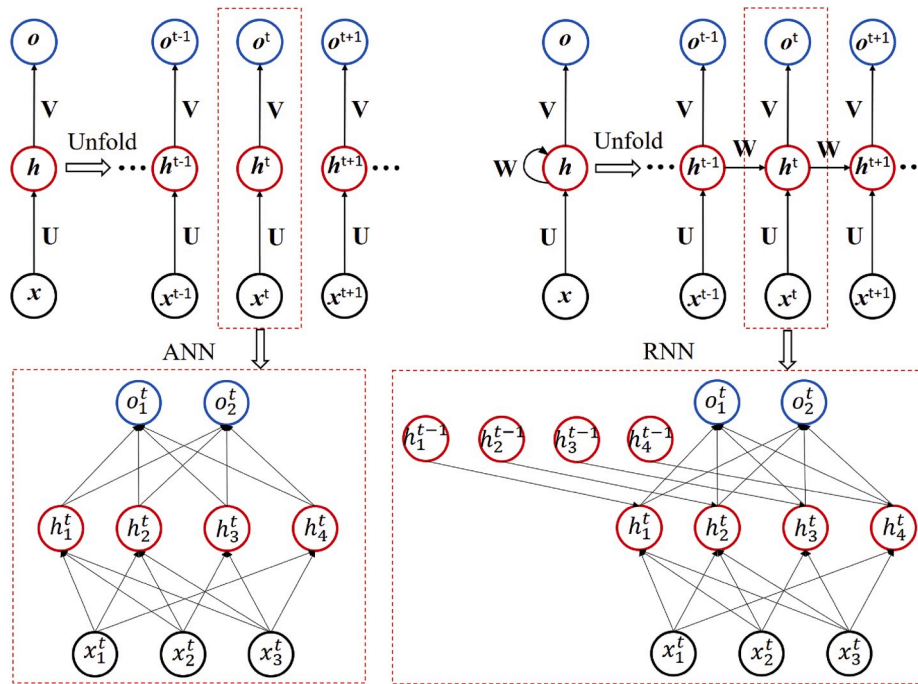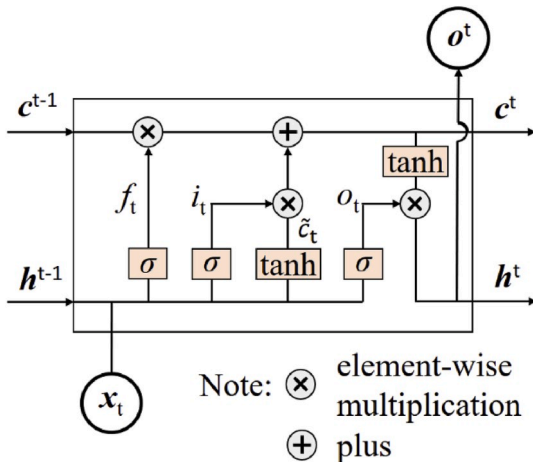
Fig. 1. Schematic view of ANN and RNN.

Fig. 2. Memory cell of LSTM.

area, thereby designing the optimum specification of a caisson foundation. Recently, the application of machine learning (ML) algorithms in geotechnical engineering have proliferated, e.g. soil parameters identification (Zhang et al., 2020a, 2020b; Zhou et al., 2016), development of constitutive models (Zhang et al., 2019a, 2020c), evaluation of soil liquefaction (Atangana Njock et al., 2020), tunneling (Chen et al., 2019a, 2019b; Elbaz et al., 2019a, 2019b; Zhang et al., 2019; Zhang et al., 2020), landslides (Huang et al., 2017; Yang et al., 2019), because the strong nonlinear mapping ability of such algorithms provides a novel methodology to tackle sophisticated problems with the interaction of multiple parameters (Sarir et al., 2019; Zhang et al., 2019a; Zhang, 2019). Most recently, Liu et al. (2019) proposed a deep neural network based failure criterion to describe the behavior of woven composites, and the predicted failure envelope matching well with the measured results. Reuter et al. (2018) compared the performance of three commonly used ML algorithms, i.e., artificial neural network, support vector machine and support vector regression (SVR) in modelling concrete failure surfaces, and found that the prediction error of SVR is

lowest. The training of ML-based model is flexible and simple as long as put the raw data. Meanwhile, once a surrogate model is well trained, the simulation of a new case can be completed within several seconds, which provides an effective method to investigate the responses of the studied object under various conditions. Nevertheless, to the best knowledge of authors, ML based models have not been developed and directly used to capture the failure envelop or mechanical responses of a caisson foundation up to now. The long short-term memory (LSTM) has been proposed to predict sequential datasets and overcome gradient vanishing and exploding problems (Hochreiter and Schmidhuber, 1997b), which means that it can account for the history of loading force or deformation. Moreover, LSTM can directly learn the failure mechanism of caisson foundations from the raw data. Therefore, a LSTM based surrogate model deserves to be developed to investigate the responses of a caisson foundation.

This study aims to develop a ML surrogate modelling approach to identify the failure envelope and mechanical responses of a caisson foundation in sand. Database is generated from an advanced numerical modelling combined the smoothed particle hydrodynamics with the SIMSAND model (SPH-SIMSAND) with a strong validation from laboratory tests to physical model tests and a field test of caisson foundations. A LSTM based surrogate model is first trained based on limited results generated from the SPH-SIMSAND, thereafter it is applied to model the mechanical behavior and the failure envelope of unknown caisson foundations with various specifications. The simulation of each new case using the surrogate model can be completed within several seconds. Therefore, the computational cost for determining the optimum design of caisson foundation specification can be remarkably reduced.

2. Deep learning based methodology

2.1. Long short-term memory neural network

The framework of a typical forward neural network is presented in Fig. 1. It can be observed that the datasets flow from the input layer to the output layer and the error is back propagated to modify the weights and biases for minimizing the discrepancy between predicted outputs
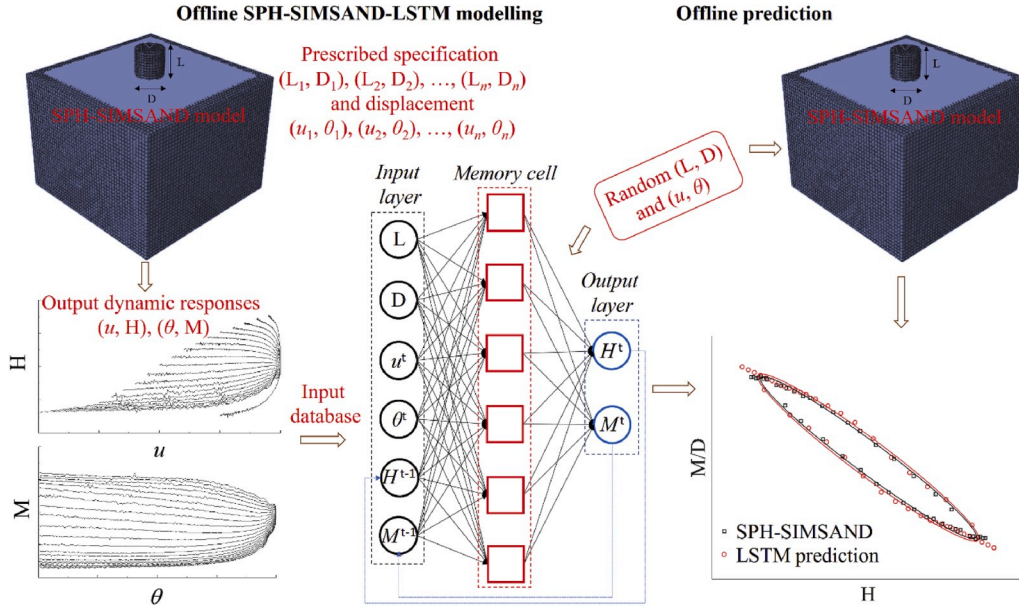
Fig. 3. Schematic view of hybrid SPH-SIMSAND and LSTM surrogate modelling process.

Table 1
Specifications of caisson foundations in the numerical model.

	Specification (L, D)
Training set	(1, 2), (1, 2.9), (1.5, 2), (1.5, 2.39), (2, 1), (2, 2), (2, 4), (2, 5.8), (2.96, 1.5), (3, 4), (3.72, 1.24), (4, 2), (4, 4), (4, 8), (4, 11.6), (4.75, 1), (6, 2.97), (6, 8), (6, 9.55), (8, 2.33), (8, 4), (8, 8), (8, 16), (10, 1.91), (10, 20), (12, 5.94), (12, 16), (15, 20), (16, 8), (16, 16), (20, 20), (20, 10)
Testing set	(1, 2.83), (1.5, 2.31), (2, 5.65), (3.56, 1.5), (4, 11.32), (4.15, 1.39), (6, 3.26), (6, 9.24)

and measured results (Rumelhart et al., 1986). In other words, the prediction of current output parameters is not affected by the previous information and it also does not affect the prediction of output parameters at the next step. Given a set of input matrix $\boldsymbol{x} = [\boldsymbol{x}_1, \boldsymbol{x}_2, …, \boldsymbol{x}_n]$, the output of hidden and output layers can be obtained by:

$$h = f(\mathbf{U}x + b_1) \tag{1}$$

$$o = g(\mathbf{V}h + b_2) \tag{2}$$

where $\mathbf{U}$, $\mathbf{V}$ = matrix connecting the input and hidden layers, hidden and output layers, respectively; $\boldsymbol{b}_1$, $\boldsymbol{b}_2$ = biases vectors in the input and hidden layers, respectively; f, g = activation functions in the hidden and output layers, respectively.

The main departure of RNN is a cyclic connection topology is adopted, as presented in Fig. 1. It is clear that the predicted output at the current step depends on current values of input parameters and the information transferred from the former hidden layer, which can be obtained by:

$$h^t = f(\mathbf{U}x + \mathbf{W}h^{t-1} + b_1) \tag{3}$$

where $\mathbf{W}$ = matrix connecting hidden layers at adjacent steps.

The history information is stored and it is applied to predict the next status, such history-dependent characteristic makes RNNs applicable to investigate problems with sequential datasets, such as language transformation, speech recognition, and the prediction of load–deformation responses (Wang and Sun, 2018; Zhu et al., 1998). However, training RNNs has proved to be problematic because the back-propagated gradients either grow or shrink at each time step, resulting in exploding or vanishing gradients (LeCun et al., 2015), that is, the learning efficiency

of the hidden layers in the front of the architecture is poorer than the later hidden layers.

To overcome gradients exploding and vanishing in the conventional RNNs, a memory cell is thus added in the architecture of LSTM in place of the neurons used in conventional RNNs. Such memory cell can store information over extended time intervals and handle long-time-lag tasks (Hochreiter and Schmidhuber, 1997a) by using a novel entity termed as "gate", as presented in Fig. 2. Three gates, i.e., forget, input and output gates are included in the memory cell to control the flow of information and the state of the cell. Forget gate decides which information is discarded from the memory cell, input gate decides which information is stored in the memory cell, and output gate decides ultimate output values. The outputs of forget and input gates at the tth step are obtained by:

$$\mathbf{f}_j^t = \sigma\left(\left[\mathbf{U}_f\mathbf{x}^t\right]_j + \left[\mathbf{W}_f\mathbf{h}^{t-1}\right]_j + \left[\mathbf{b}_f\right]_j\right) \tag{4}$$

$$\mathbf{i}_j^t = \sigma\left(\left[\mathbf{U}_i\mathbf{x}^t\right]_j + \left[\mathbf{W}_i\mathbf{h}^{t-1}\right]_j + \left[\mathbf{b}_i\right]_j\right) \tag{5}$$

where σ = *sigmoid* function. In the forget gate, $\sigma = 1$ and 0 represent all information is maintained or discarded, respectively. In the input gate, $\sigma = 1$ and 0 represent all information is selected or discarded, respectively.

Based on the forget and input information, the memory cell state at the tth current step is thus updated by:

$$\tilde{\mathbf{c}}_j^t = tanh\left(\left[\mathbf{U}_c\mathbf{x}^t\right]_j + \left[\mathbf{W}_c\mathbf{h}^{t-1}\right]_j + \left[\mathbf{b}_c\right]_j\right) \tag{6}$$

$$\mathbf{c}^t = \mathbf{f}^t \odot \mathbf{c}^{t-1} + \mathbf{i}^t \odot \tilde{\mathbf{c}}^t \tag{7}$$

where *tanh* is the activation function; $\odot$ = elementwise product; $\mathbf{f}^t \odot \mathbf{c}^{t-1}$ represents the discarded information; $\mathbf{i}^t \odot \tilde{\mathbf{c}}^t$ represents newly selected information. The update of memory cell status with an addition format can avoid the gradients vanishing and exploding. Thereafter, output of the hidden layer at the tth step is obtained by:

$$\mathbf{o}_j^t = \sigma\left(\left[\mathbf{U}_o\mathbf{x}^t\right]_j + \left[\mathbf{W}_o\mathbf{h}^{t-1}\right]_j + \left[\mathbf{b}_o\right]_j\right) \tag{8}$$

$$\mathbf{h}^t = \mathbf{o}^t \odot tanh(\mathbf{c}^t) \tag{9}$$

A multiplicative input gate unit is employed to protect the memory contents stored at the current step from perturbation by irrelevant inputs, and a multiplicative output gate unit is employed to protect other
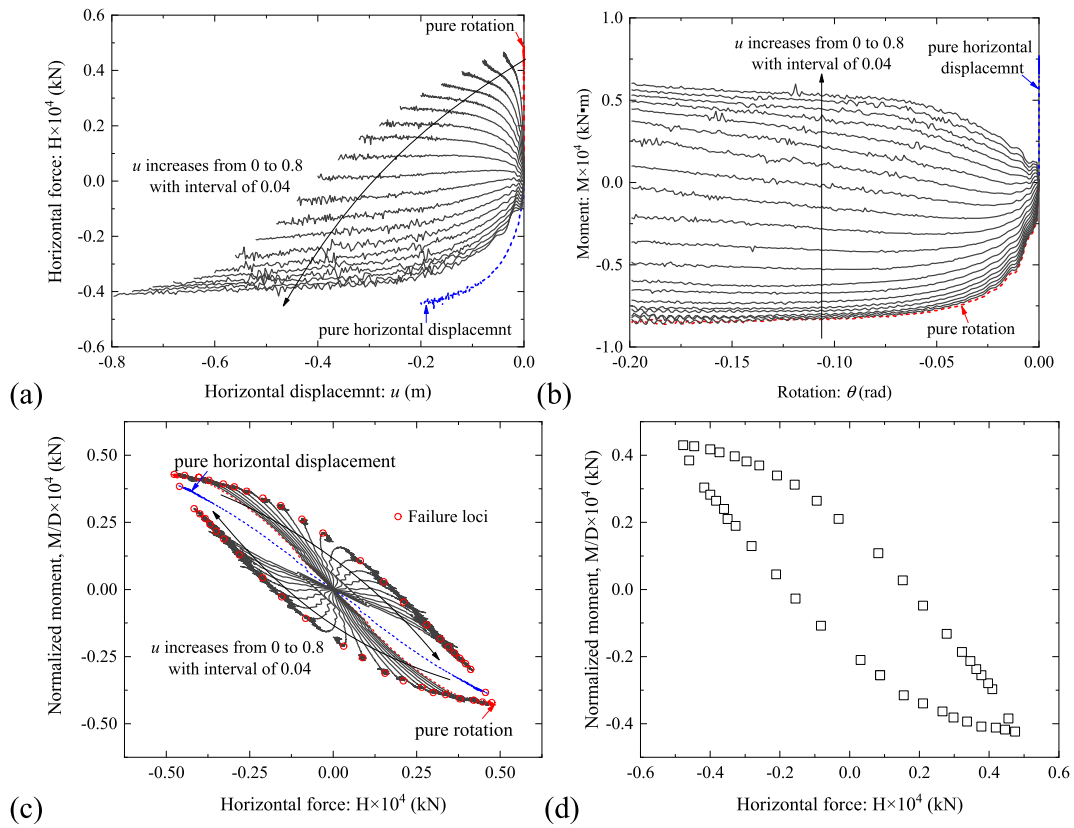
Fig. 4. Results of SPH-SIMSAND numerical modelling: (a) u–H; (b) θ–M; (c) H–M/D; (d) failure envelope.

units from perturbation by currently irrelevant memory contents stored at the current step (Hochreiter and Schmidhuber, 1997a). It should be noted that numerous LSTM variants have been thereafter proposed such as gated recurrent unit (GRU) (Cho et al., 2014). LSTM with numerous weights and biases that is beneficial to predict high-dimensional issues is selected to explore its feasibility in capturing caisson foundations responses.

2.2. Proposed hybrid surrogate model

The development of a hybrid surrogate model is categorized into two phases: offline and online modelling. The objective of the offline modelling is to bridge the numerical modelling platform of SPH-SIMSAND by Yin et al., 2018b and deep learning algorithm LSTM. Hence, a LSTM based surrogate model that can entirely replace the numerical modelling to reduce computational cost is further developed. This phase starts from the calibration of numerical model including the parameters of SPH and SIMSAND constitutive model for a given soil type (Jin et al., 2016, 2017, 2019a; Yin et al., 2013, 2016, 2018a), thereafter several cases with different specifications of caisson foundations are computed for creating a synthetic database. It should be noted that the loading paths are consistent among all cases. Herein, 80% of datasets are used to train the LSTM based surrogate model, and the remaining 20% of datasets are used to test the model. The online modelling aims to utilize the surrogate model to predict the mechanical responses and the failure envelope of a caisson foundation with a random specification under various loading paths. Meanwhile the SPH-SIMSAND is also used to simulate the same case to validate the accuracy of the surrogate model.

The framework of training a LSTM based model is presented Fig. 3. The input parameters consist of the length L and diameter D of a caisson foundation, the horizontal displacement u^t and rotational angle θ^t at the current step, the sequence of history values of horizontal force H^{t-1} and bending moment M^{t-1}. The output parameters are the horizontal force

H^t and bending moment M^t. The training performance of the LSTM based model is evaluated by the mean square error (MSE) values on both training and test sets, meanwhile 10-fold cross-validation method is applied to enhance model robustness, thereby the loss function can be obtained by:

$$\text{MSE} = \frac{1}{10r} \sum_{i=1}^{10} \sum_{j=1}^{r} \left(y_i^m - y_i^p \right)^2 \tag{10}$$

where y_i^m = measured result at the ith point; y_i^p = predicted result at the ith point; r = a total of datasets in one cross-validation set.

2.3. Evaluation indicators

Two commonly used indicators "Mean Absolute Percentage Error (MAPE)" and "Nash–Sutcliffe model Efficiency coefficient (NSE)" are employed to evaluate the performance of the LSTM based model. MAPE is an unbiased measure to compute the average prediction error of the model, and NSE can assess the accuracy and precision of the model (Nash and Sutcliffe, 1970). The expression of two measures can be obtained by

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i^m - y_i^p}{y_i^m} \right| \tag{11}$$

$$\text{NSE} = 1 - \frac{\sum_{i=1}^{n} \left(y_i^p - y_i^m \right)^2}{\sum_{i=1}^{n} \left(y_i^m - \overline{y_i^m} \right)^2} \tag{12}$$

where n = a total of datasets; μ = mean value of y_i^p / y_i^m; δ = standard deviation of y_i^p / y_i^m. The combination of MAPE and NSE enables to comprehensively evaluate the model performance. Low value of MAPE and high value of NSE indicate an excellent model performance.
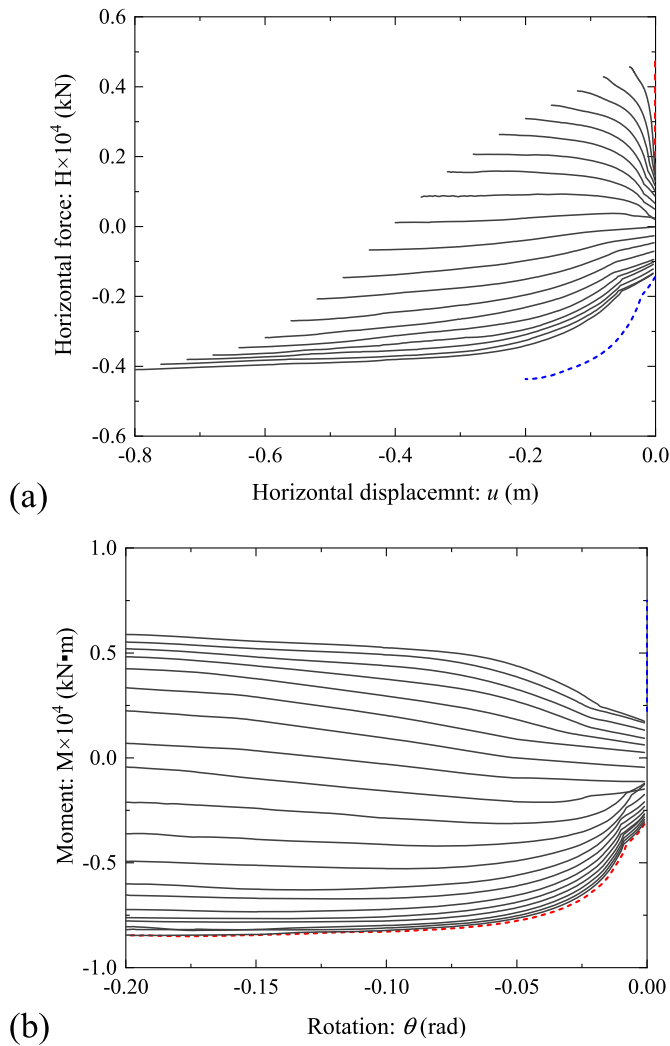
(a)

(b)

Fig. 5. Data smoothing: (a) u–H; (b) θ–M.

(a)

(b)

Fig. 6. Loss values yielded by LSTM model with dropout layer on the: (a) training set; (b) testing set.

Table 2
Main hyper-parameters during training LSTM.

Hyper-parameter	Description	Value
N_h	Number of hidden layers	3
N_n	Number of nodes in the hidden layer	80; 80; 50
Activation	Activation function to use	*tanh*; ReLU; ReLU
Dropout	Fraction of the units to drop	0
Optimizer	Algorithm for optimizing weights and biases	Adam
η	Learning rate in the optimizer	0.0002–0.002; period = 20
Batch_size	Number of training samples	200
Epoch	Number of iterations during training	200

3. Database design

3.1. Data source

This study directly utilizes the numerical results of SPH-SIMSAND conducted by Jin et al., 2019b. In such research work, the parameters of the SIMSAND model were first calibrated using triaxial tests on Baskarp sand. Meanwhile a cone penetration test and model tests on a reduced scale, and a full scale field test of caisson foundation were simulated to v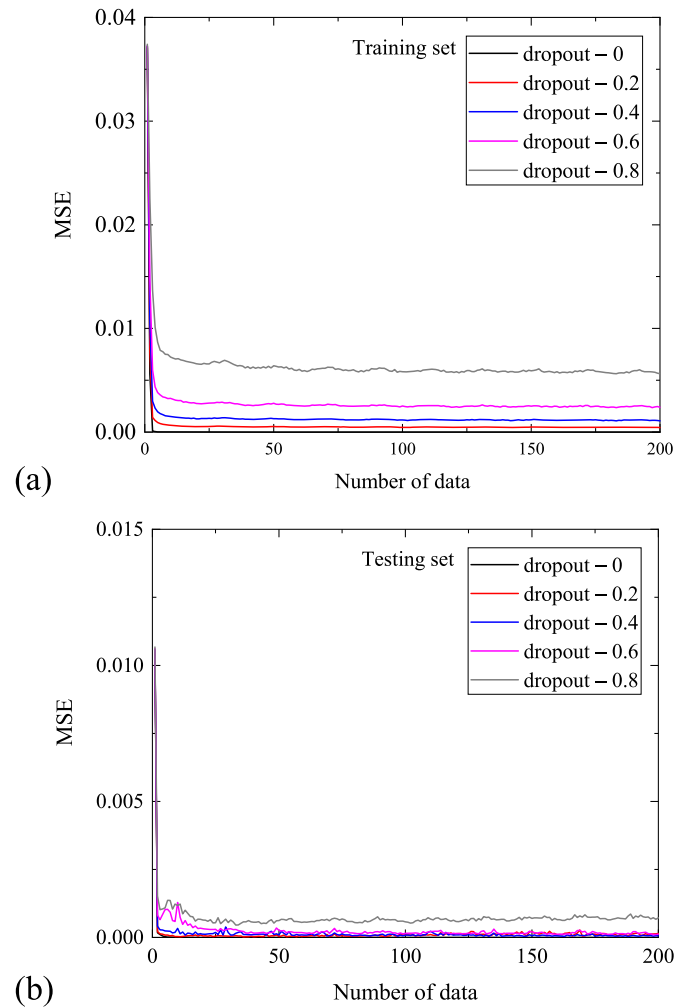alidate SPH-SIMSAND. The excellent agreement between the numerical and experimental results indicates the reliability of SPH-SIMSAND modelling method to investigate the responses of a caisson foundation. Therefore, the results from such numerical modelling are used to establish database in this study.

Because of the lightness of a caisson foundation, the horizontal and overturning moment bearing capacities are important for the design. Jin et al., 2019b thus investigated the failure envelope of caisson foundations with various specifications in the H-M plane. Numerical radial displacement tests in which the ratio between the applied displacements or the combined rotation-displacement increments is kept constant are adopted as the main loading control over the course of numerical modelling (Gottardi et al., 1999). This study aims to identify the mechanical responses and failure envelop of caisson foundations with various specifications, thereby a total of 40 numerical models with 176000 datasets are constructed and the detailed specifications of studied caisson foundations are presented in Table 1. Herein, 32 numerical modelling results with 140800 datasets are used to train the LSTM based model, and the remaining 8 numerical modelling results with 35200 datasets are used to test the performance of the LSTM based model. All datasets are ultimately stored in comma-separated values (CSV) file for fast importing into Python.

As an example, a typical simulating result of a caisson foundation with an outer diameter (D) of 2 m and a skirt length (L) of 2 m is presented in Fig. 4. It can be observed that a total of 22 loading paths including a pure rotation and a pure horizontal displacement paths are considered in each case, where the horizontal displacement increases
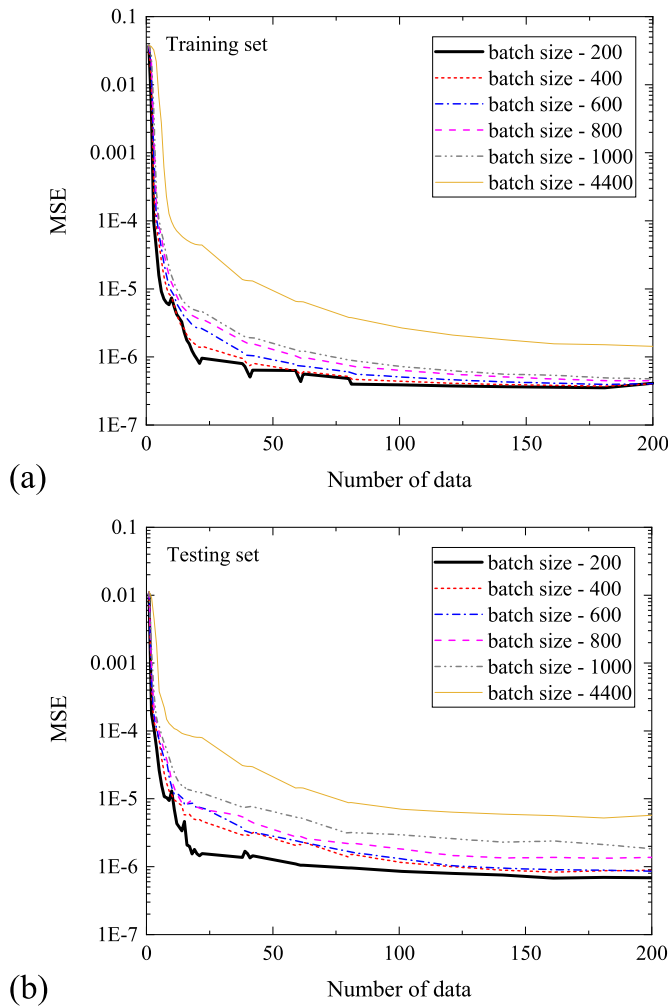
Fig. 7. Loss values yielded by LSTM model with different batch size on the: (a) training set; (b) testing set.

Table 3
Indicator values for the training and testing sets.

Parameter	Training set		Testing set	
	MAPE	NSE	MAPE	NSE
H	18.36%	0.99	25.07%	0.93
M	36.57%	0.99	42.15%	0.92

from 0 to 0.8 m with an interval of 0.04 and the rotation maintains in the range of 0 to 0.2 rad (see Fig. 4(a) and (b)). Each loading path includes 200 data points. To determine the bearing capacity, the loading paths are plotted in the H-M plane, and the ultimate bearing capacity is determined by the inflexion of loading paths, i.e., failure loci, as presented in Fig. 4(c). It should be noted that the ultimate bearing capacity is hard to reach for some loading paths. To unify the basis of determination, the ultimate bearing capacities of such cases are represented by the ends of loading paths (Jin et al., 2019c), based on which the final failure envelope can be obtained by connecting the failure loci, as presented in Fig. 4(d).

3.2. Data preprocessing

It can be seen that the simulating loading paths of SPH-SIMSAND are noisy, as presented in Fig. 4. The large variance of sequential data has an adverse impact on the training process and the model performance (Xu

et al., 2019). This study thus introduces a sliding window approach to smooth data before such datasets applied to train the LSTM based model. The value of smoothed x_n can be obtained by:

$$x_n = \frac{1}{t} \sum_{i=n-t+1}^{n} x_i \tag{13}$$

where t = window size. The average value of datasets within a window is assigned as the new value of the studied parameter. The datasets within a window consist of current and the former $(t–1)$ values. It should be noted that the first $(t–1)$ and the last $(t–1)$ points cannot form a complete window, thereby values of such points maintain constant. Larger window size can generate smoother sequential curve, but it is much more likely to deviate from the original curve. Considering the variance of u–H and θ–M relationships presented in Fig. 4 is small, the window size is thus set as two in this study for maintaining the reliability and smoothness of the datasets. The smoothed relationships of u–H and θ–M are presented in Fig. 5. It can be observed that the magnitude and trend of sequential curves are roughly identical to the original results presented in Fig. 4, meanwhile the smoothness is improved dramatically.

The different scales of input and output parameters also affect model performance. After smoothing all datasets, a remedy is to normalize all datasets into a same scale [–1, 1] to eliminate the scale effect using Eq. [14].

$$x_{norm} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} (\bar{x}_{\max} - \bar{x}_{\min}) + \bar{x}_{\min} \tag{14}$$

where $x_{\max}$ and $x_{\min}$ = measured maximum and minimum of the parameter x; $\bar{x}_{\max}$ and $\bar{x}_{\min}$ = 1 and −1, respectively. The ultimate database can be downloaded at the Appendix.

4. Offline training of a hybrid surrogate model

4.1. Determination of hyper-parameters

Training the LSTM based model means numerous hyper-parameters need to be determined in advance. The main hyper-parameters over the course of training are summarized in Table 2. Herein, grid search method is used to search for the optimum architecture of LSTM including the number of hidden layers, the number of nodes and the activation function in each layer. Considering the dropout layer can overcome the overfitting problems, the performance of LSTM based models with various dropout rates (0, 0.2, 0.4, 0.6, 0.8) is thus investigated. Adaptive moment estimation (Adam) optimizer is utilized in this study due to its superiority (Ruder, 2016). The learning rate of Adam controls the updated step size of weights and its default value is 0.001. This parameter needs to be finely tuned for a complex problem with numerous saddle points (Smith, 2017). Considering the highly non-linear responses of a caisson foundation, a cyclical learning rate proposed by Smith, 2017 is used to finely optimize the weights and biases of LSTM. Batch size represents the number of samples to be fed at each training step. Because each loading path consists of 200 datasets as mentioned in the section 3.1, the batch size is thus set as 200n (n = 1, 2, …) to ensure that datasets from an entire loading path can be simultaneously used to train model. This study investigates the performance of LSTM based models with batch size of 200, 400, 800, 1000 and 4400 (a total of datasets in a case is 4400). The number of epochs requires to be sufficiently large to ensure the loss value can converge at a constant value. Herein, orthogonal initializer is used to generate an initial random orthogonal weights matrix and zeros initializer is used to generate an initial zero biases vector.

Regarding the implementation, Keras, that is a high-level deep learning library based on Python programming language, is leveraged to design the architecture of the LSTM based model. Tensorflow as the back engine supports to implement operations in Keras such as tensor calculus. The matrix construction and computation are achieved by using
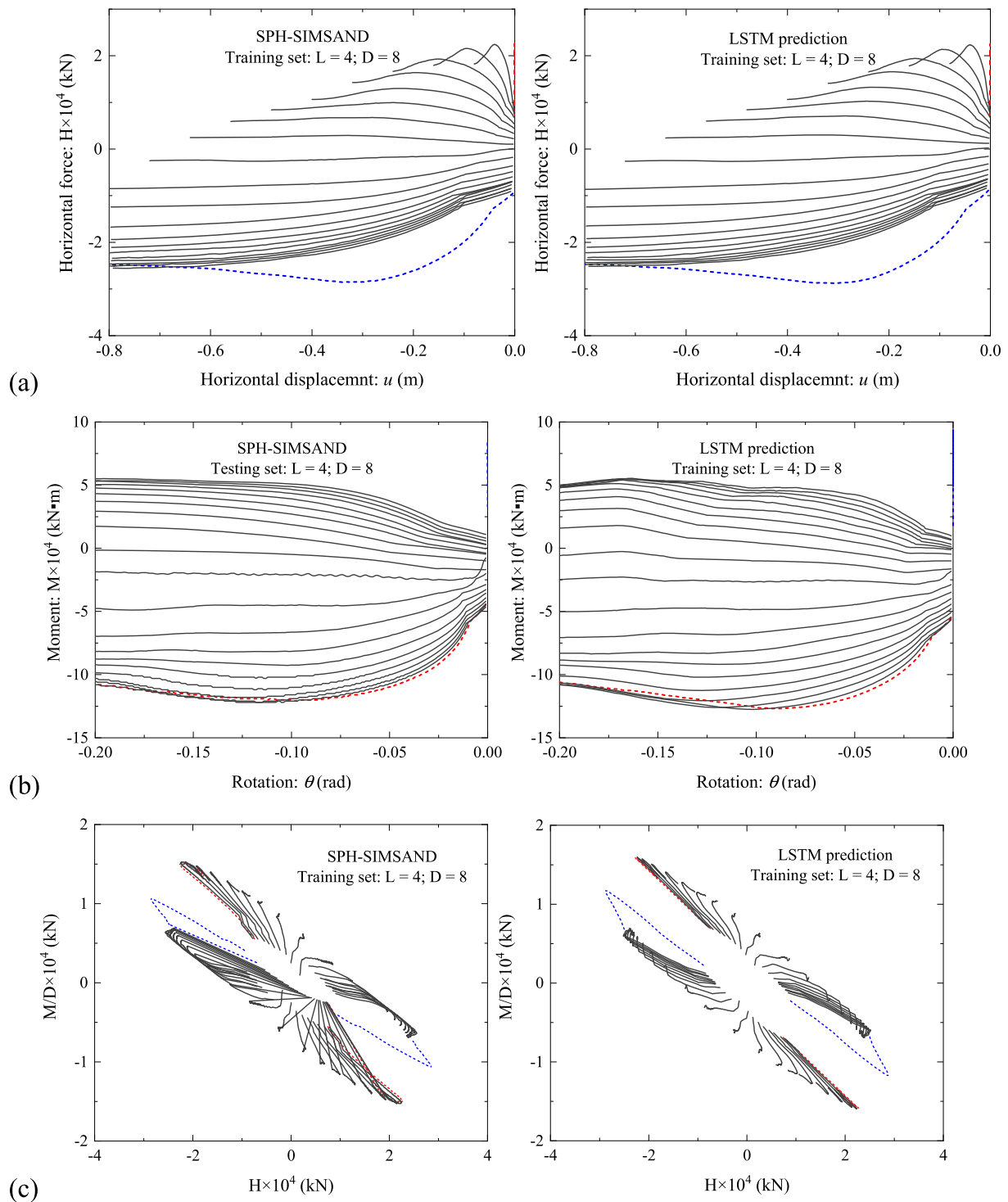
Fig. 8. Predicted loading using LSTM based surrogate model for the training set L = 4, D = 8, in comparison with numerical results: (a) *u*–H; (b) *θ*–M; (c) H–M/D.

NumPy library. Data mining and analysis toolbox Pandas is employed to import CSV datasets file.

The results of grid search indicate that the LSTM model with three hidden layers can produce the lowest loss value. The number of nodes in each layer is 80, 80 and 50, respectively, and the corresponding activation functions are *tanh*, ReLU and ReLU, respectively. Therefore, the number of weights and biases are 104820 and 842, respectively. Over the course of training, the learning rate first increases from 0.0002 to 0.002 within 10 epochs and thereafter decreases from 0.002 to 0.0002 within 10 epochs, thereby each period includes 20 epochs. Such strategy

improves the optimization process to escape from the local optima and saddle points. The loss value roughly maintains steadily and converges at a constant value within 200 epochs, thereby the maximum number of epochs is set as 200.

4.2. Underfitting and overfitting examination

The examination of underfitting and overfitting is a key step to guarantee the reliability of the LSTM based model. Learning curves of both loss values on the training and testing sets have been successfully
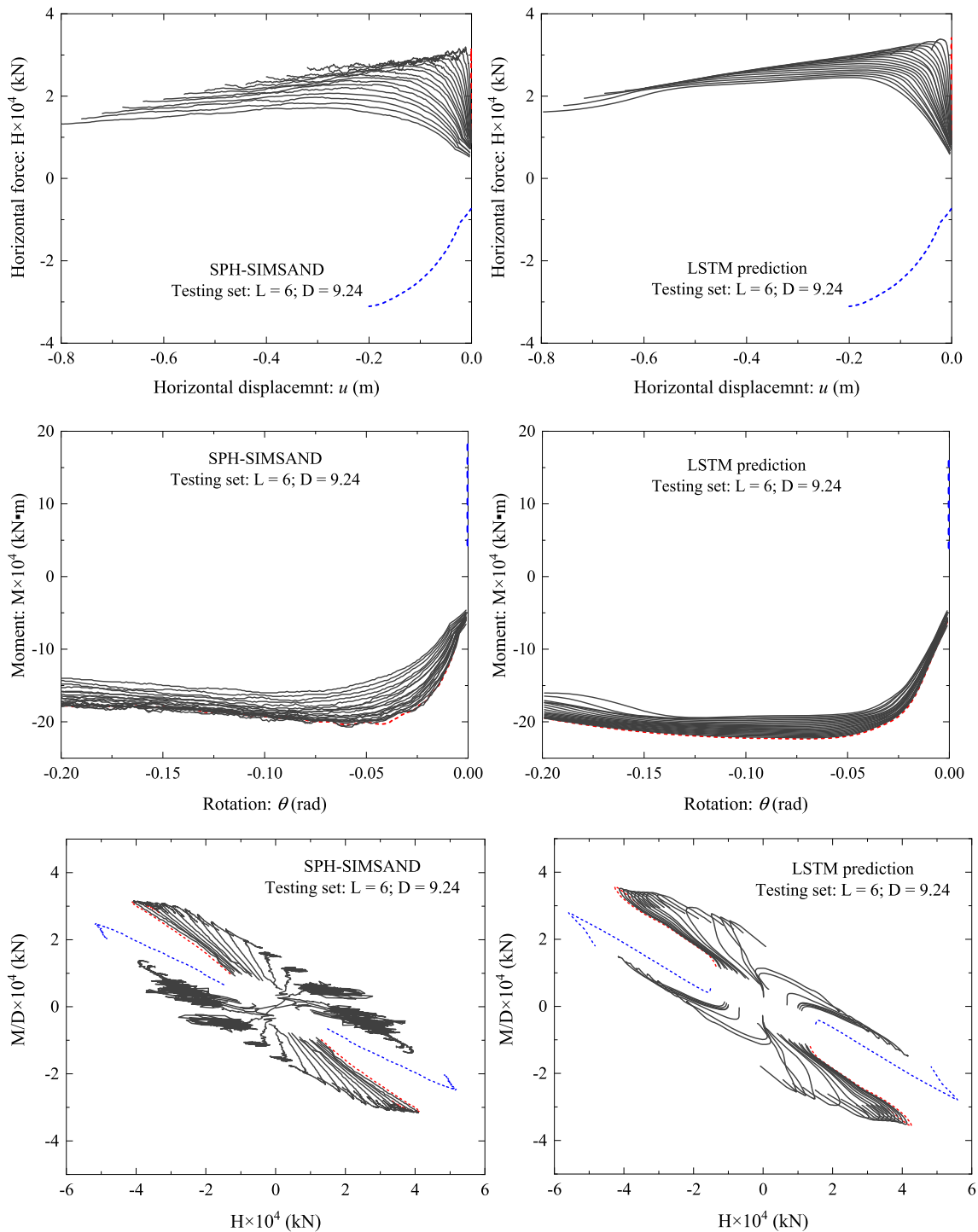
Fig. 9. Predicted loading using LSTM based surrogate model for the testing set L = 6, D = 9.24, in comparison with numerical results: (a) u–H; (b) θ–M; (c) H–M/D.

used to evaluate the underfitting and overfitting problems (Hassan et al., 2020), because it can reflect how well a behavior of neural network is improved with the increasing number of training samples or complexity of neural network (Murata et al., 1993). Large loss values on both training and testing sets represent that the LSTM based model exists underfitting problem. The large loss value on the training set and the low loss value on the testing set represent that the LSTM based model has overfitting problem. This study thus uses learning curve to examine the potential underfitting and overfitting issues.

Numerous research works have demonstrated that dropout family methods give significant advantages over other regularization methods

such as L_1 and L_2 penalties(Moradi et al., 2019; Srivastava et al., 2014). Therefore, a dropout method is used to avoid potential overfitting of LSTM in this study. The effect of dropout rates on the prediction performance of LSTM based models can be observed in Fig. 6. In comparison with the model with dropout rates of 0, 0.2, 0.4, 0.6 and 0.8, it is clear that the loss values on both training and testing sets increase with the increasing dropout rates. Meanwhile the increasing dropout rates can cause the variation of loss values. The minor effectiveness of the dropout layer indicates that the LSTM based model can well suppress the overfitting problem over the course of training and provide accurate prediction. Therefore, the dropout rate is set as 0 in this study.
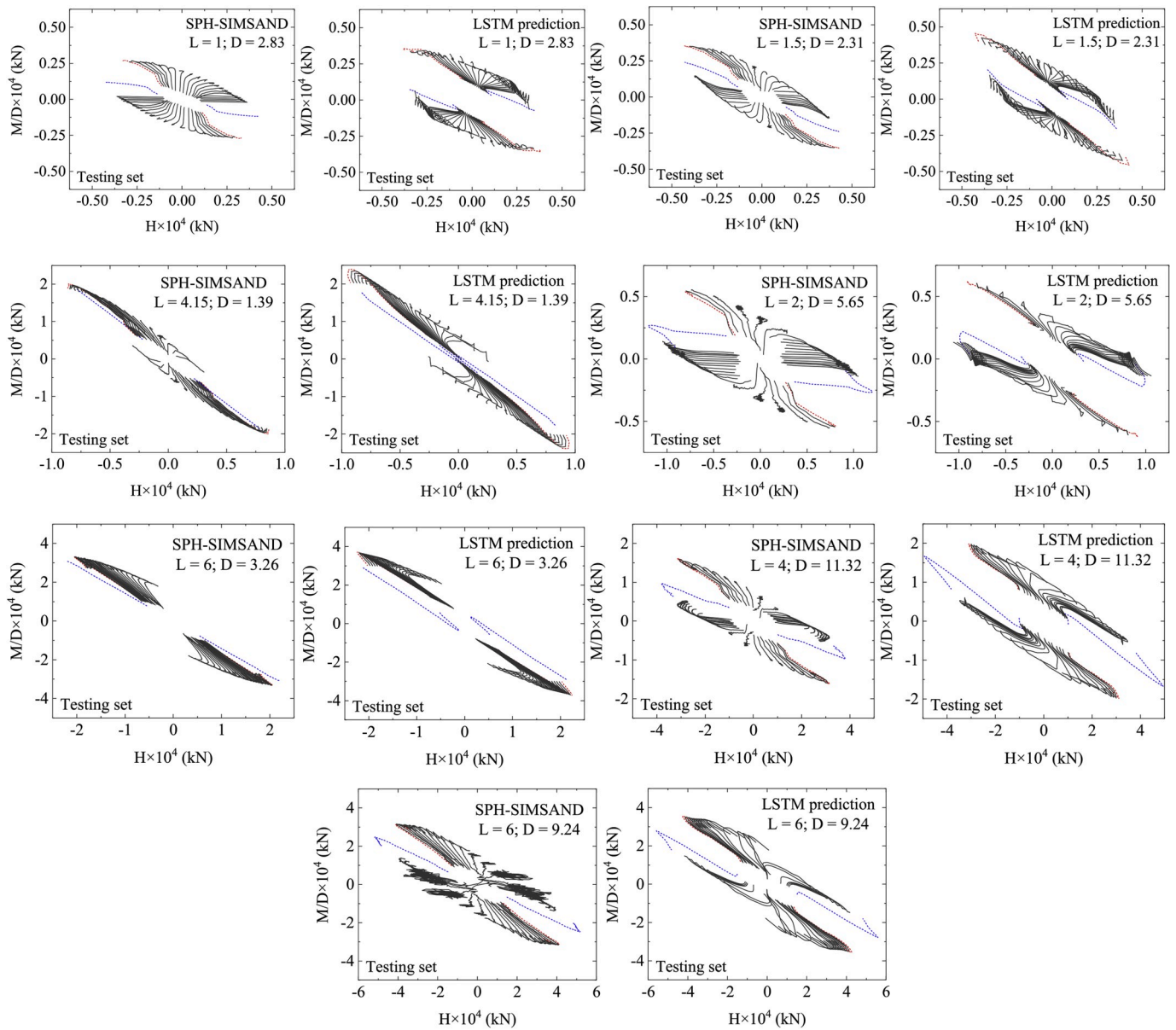
Fig. 10. Comparison between loading paths of testing sets yielded by SPH-SIMSAND and LSTM.
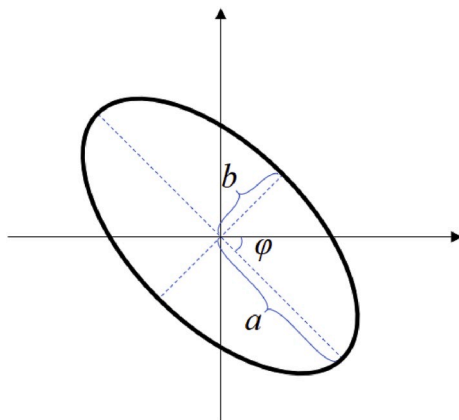

Fig. 11. Notation convention of failure envelope.

Fig. 7 presents the evolution of loss values on both training and testing sets produced by LSTM based models with various number of batch size. Loss value in logarithm is used to highlight the difference of model performance because of the small loss value. The loss value roughly holds steadily as the epochs reaches 200, and it clearly increases with the increasing number of batch size. The increasing training samples bring about difficulties in optimizing weights and biases to reduce loss value, but the model trained with larger batch size possesses better generalization ability, thereby the prediction error on both training and testing sets decreases continuously throughout the training process. The model trained with small batch size ensures to shrink the difference between the most measured and predicted results, thereby the loss value is small, but such model suffers from poor generalization ability. Consequently, the LSTM model with optimized weights and biases which presents an excellent prediction performance on cross-validation sets may produce large error on both training and testing set, thereby it can be seen from Fig. 7 that the loss value varies as the number of batch size decreases. Considering the model trained with batch size of 200 produces the lowest loss value and the variation is acceptable, thereby the batch size of 200 is applied to train the LSTM based model in this
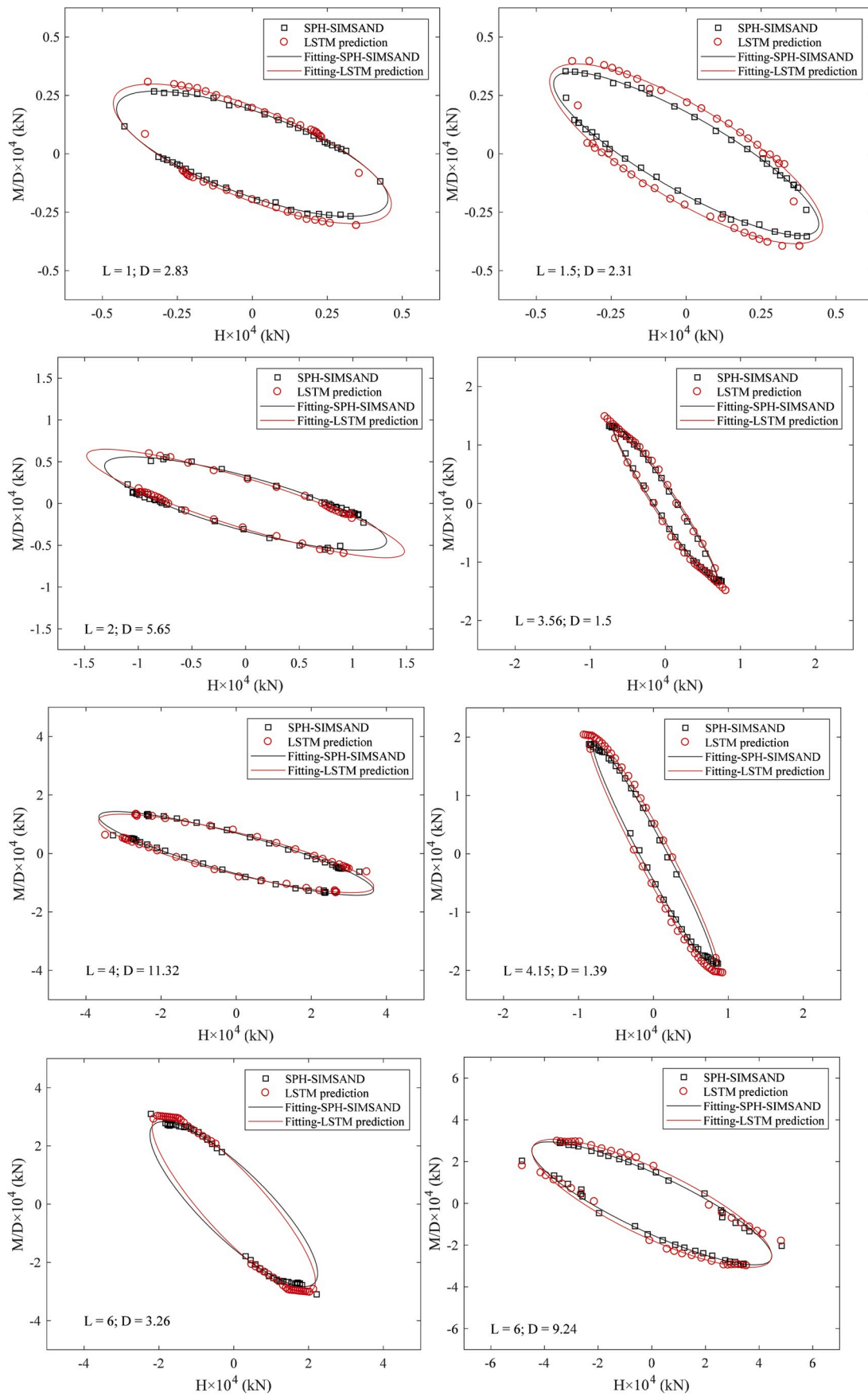
Fig. 12. Comparison between failure envelopes of testing sets yielded by SPH-SIMSAND and LSTM.

Table 4
Values of parameters used in failure envelope.

L	D	SPH-SIMSAND			LSTM based model		
		$a\ (\times 10^4)$	$b\ (\times 10^4)$	$\phi\ (°)$	$a\ (\times 10^4)$	$b\ (\times 10^4)$	$\phi\ (°)$
1	2.83	0.50	0.17	26.30	0.52	0.18	28.96
1.5	2.31	0.54	0.15	37.27	0.57	0.18	39.06
2	5.65	1.39	0.31	19.95	1.59	0.29	21.86
3.56	1.5	1.49	0.15	62.98	1.54	0.17	63.31
4	11.32	3.88	0.66	19.48	3.84	0.69	17.74
4.15	1.39	2.06	0.18	67.5	2.23	0.23	67.75
6	3.26	3.47	0.99	52.84	3.63	0.89	55.51
6	9.24	5.20	1.29	31.69	5.20	1.51	32.39

study. Meanwhile the loss value decreases continuously for both training and test sets, and the convergence value is roughly identical. Such factors indicate that the constructed LSTM model can well overcome underfitting and overfitting problems.

4.3. Evaluation of surrogate model performance

All of optimum values of hyper-parameters are determined as mentioned in the former two sections. Table 2 summarizes such values and the model is trained based on this set of parameters. The indicator values for describing the prediction performance of the model are presented in Table 3. For the training set, MAPE values are low on both horizontal force and moment predictions, meanwhile NSE values are roughly identical to 1. The LSTM based model shows an excellent performance in capture loading paths of caisson foundations.

For brevity, the predicted loading path of one training set with L = 4 m and D = 8 m is presented as a typical example to illustrate the training performance of the LSTM based model, as shown in Fig. 8. Such results are obtained within several seconds. Remarkably, the LSTM based model is capable of replicating the u–H, θ–M and H–M/D relationships with negligible error. The results presented in Fig. 8(a)–(b) and (e)–(f) indicate that the softening behavior can be captured by the LSTM based model. The excellent repeatability provides a basis for the LSTM based surrogate model to replace numerical modelling for investigating the mechanical responses of caisson foundations with lower computational cost.

5. Online prediction using LSTM surrogate model

5.1. Loading paths prediction

To test the reliability of the LSTM based surrogate model to guarantee its application in engineering practice, the responses of additional eight caisson foundations are investigated using the LSTM based model developed in the former section, meanwhile the SPH-SIMSAND platform

is also used to simulate the same cases for comparison. Fig. 9 presents the predicted loading paths of a caisson foundation with L = 6 m and D = 9.24 m. It can be observed that the LSTM based model has an excellent performance in reproducing the u–H relationship, but the prediction errors of initial θ–M and H–M/D relationships are large, which is attributed to the loss function, i.e., MSE value. Such indicator focuses on eliminating the discrepancy of large output values, whereas the small output values are less important, thereby the trained LSTM based model shows larger prediction error in predicting initial loading paths. The in-depth study of loss function selection to achieve the tradeoff of predicting large and small values, and further improve the model generalization ability is important, but is out of the scope of this paper. These studies will be conducted in a future dedicated work. Overall, the prediction performance of the LSTM based model on the mechanical responses of unknown caisson foundations is reliable.

Fig. 10 presents the predicted H–M/D relationships of the remaining seven testing sets. The loading paths generated from the LSTM based model show good agreement with the numerical results. Small MAPE and high NSE values are generated on the testing set, as presented in Table 3. The simulations using the LSTM based model are completed without using any internal variables to capture the responses of caisson foundations. Such model is thus ready to be used to predict the failure envelope of caisson foundations with various specifications on a given soil type in engineering practice.

5.2. Prediction of failure envelope in the H–M plane

As presented in Fig. 4(d), the failure envelope in the H–M plane has an elliptical shape. Following Villalobos et al., 2009, Jin et al., 2019c proposed an ellipse formulation with only three parameters a, b and ϕ to describe the failure envelope of a caisson foundation in the H–M plane. Fig. 11 illustrates the notation convention of failure envelope, in which a and b are the major and minor axis of the ellipse, respectively, and ϕ is the rotation of the ellipse. The formulation can be obtained by:

$$A_1 X^2 + A_2 XY + A_3 Y^2 + A_4 = 0 \tag{15}$$

$$\begin{cases} A_1 = a^2 (\sin \phi)^2 + b^2 (\cos \phi)^2 \\ A_2 = 2(b^2 - a^2) \sin \phi \cos \phi \\ A_3 = a^2 (\cos \phi)^2 + b^2 (\sin \phi)^2 \\ A_4 = -a^2 b^2 \end{cases} \tag{16}$$

where X and Y denote the horizontal force H $\times 10^4$ and normalized moment M/D $\times 10^4$ in this study.

The failure loci of eight testing cases obtained from the numerical modelling and the LSTM based model is plotted together in Fig. 12. The predicted points are close to the numerical results. Using Eqs. 15 and 16 to fit these failure loci, it can be observed that the fitted failure envelope based on failure loci obtained from the LSTM based models exhibit good
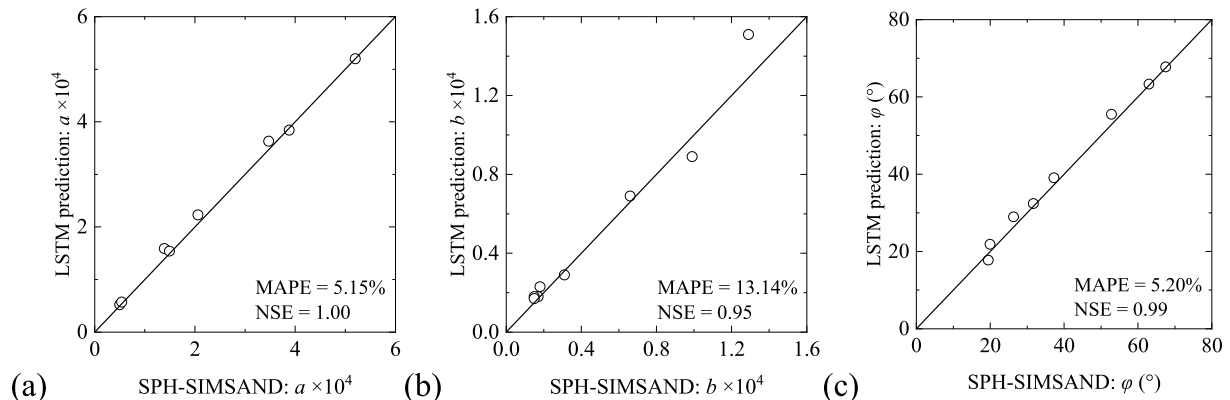

Fig. 13. Values of parameters used in failure envelope: (a) major axis a; (b) minor axis b; (c) rotation ϕ.

agreement with numerical results. The corresponding *a*, *b* and *φ* values generated by the numerical modelling and the LSTM based model are presented in Table 4. Scatter plots of the predicted and actual *a*, *b* and *φ* values are presented in Fig. 13 with the MAPE and NSE values. It is obvious that all points are close to the line with slope of 1. MAPE values of *a* and *φ* are only around 5%, the larger MAPE value (13.14%) in predicting *b* is attributed to the smaller value of *b*. All NSE values are larger than 0.95. Such results clearly demonstrate the capacity of the LSTM based model in predicting the failure envelop of unknown caisson foundations with various specifications. It can reproduce the mechanical responses of caisson foundations with low computational cost and high accuracy, which is a significant improvement over numerical and analytical methods.

6. Conclusions

This study presents the development of a hybrid surrogate model with the integration of numerical modelling technique SPH-SIMSAND and deep learning algorithm long short-term memory (LSTM) to identify the mechanical responses and failure envelope of a caisson foundation in sand. The LSTM based surrogate model was first trained based on limited results generated from the numerical modelling, thereafter it was applied to simulate the mechanical behavior and the failure envelope of unknown caisson foundations with various specifications. The underfitting and overfitting problems that commonly exist during the development of machine leaning based models have been well tackled, which ensures the robustness and generalization ability of the LSTM based model.

The predictive ability on the sequential data allowed the LSTM based model to accurately reproduce the mechanical responses of a caisson foundation including the relationships between horizontal displacement and force, rotation and moment, horizontal force and moment. Such surrogate model has the capacity to memorize and interpret history-dependent events without using additional parameters. It means that LSTM based model is more flexible than macro-element method, because it can directly learn the failure mechanism of caisson foundation from the raw data. The failure envelopes of caisson foundations can be rapidly obtained using the LSTM based surrogate model, which agree well with the actual results. Therefore, LSTM based model also guarantees a high computational efficiency and accuracy in comparison with physical and numerical modelling.

Overall, LSTM based surrogated model shows a great potential of application in engineering practice. Engineers can first obtain several responses of caisson foundations using experiments or numerical modelling, thereafter a LSTM based model is built with such datasets and further use the LSTM based model to obtain the responses of caisson foundations under different conditions. Thereby an optimum design of caisson foundation can be obtained with less experimental or computational costs.

The proposed method of developing the LSTM based model can be extended to more conditions (i.e. different soil properties, different vertical forces applied to the caisson foundation) if database is available. Future work will focus on the application of the method and the model using experimental observations.

Appendix

The datasets and the optimum LSTM based model used in this study can be freely downloaded at following link. There are three documents entitled as "databasesmooth.csv", "Caisson_Mech_Response.h5" and "validation.py". The data can be stored in the "databasesmooth.csv", and "Caisson_Mech_Response.h5" is the optimum LSTM based surrogate model. "Validation.py" is the main code. Researchers and engineers can directly run this code to replicate the results presented in this study and apply it in engineering practice.

https://www.researchgate.net/publication/338983602_LSTM_

based_model_for_predicting_caisson_foundations_responses.

Declaration of competing interest

We declare that we have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this manuscript "A LSTM Surrogate Modelling Approach for Caisson Foundations".

CRediT authorship contribution statement

Pin Zhang: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Visualization. **Zhen-Yu Yin:** Resources, Conceptualization, Methodology, Resources, Writing - review & editing, Supervision, Project administration, Funding acquisition. **Yuanyuan Zheng:** Methodology, Writing - review & editing. **Fu-Ping Gao:** Methodology, Writing - review & editing.

Acknowledges

References

Atangana Njock, P.G., Shen, S.-L., Zhou, A., Lyu, H.-M., 2020. Evaluation of soil liquefaction using AI technology incorporating a coupled ENN/t-SNE model. Soil Dynam. Earthq. Eng. 130, 105988.

Byrne, B., Houlsby, G.T., 2001. Observations of footing behaviour on loose carbonate sands. Geotechnique 51 (5), 463–466.

Cassidy, M., Byrne, B., Houlsby, G.T., 2002. Modelling the behaviour of circular footings under combined loading on loose carbonate sand. Geotechnique 52 (10), 705–712.

Chen, R.P., Zhang, P., Kang, X., Zhong, Z.Q., Liu, Y., Wu, H.N., 2019a. Prediction of maximum surface settlement caused by EPB shield tunneling with ANN methods. Soils Found. 59 (2), 284–295.

Chen, R.P., Zhang, P., Wu, H.N., Wang, Z.T., Zhong, Z.Q., 2019b. Prediction of shield tunneling-induced ground settlement using machine learning techniques. Front. Struct. Civ. Eng. 13 (6), 1363–1378.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, v., Bougares, F., Schwenk, H., Bengio, Y., 2014. Learning Phrase Representations Using RNN Encoder–Decoder for Statistical Machine Translation arxiv 1406.1078.

Elbaz, K., Shen, S.-L., Zhou, A., Yuan, D.-J., Xu, Y.-S., 2019a. Optimization of EPB shield performance with adaptive neuro-fuzzy inference system and genetic algorithm. Appl. Sci. 9 (4).

Elbaz, K., Shen, S.L., Zhou, A.N., Yin, Z.Y., Lyu, H.M., 2019b. Prediction of disc cutter life during shield tunnelling with AI via incorporation of genetic algorithm into GMDH-type neural network. Engineering (in press).

Gelagoti, F., Georgiou, I., Kourkoulis, R., Gazetas, G., 2018. Nonlinear lateral stiffness and bearing capacity of suction caissons for offshore wind-turbines. Ocean Eng. 170, 445–465.

Gottardi, G., Houlsby, G.T., Butterfield, R., 1999. Plastic response of circular footings on sand under general planar. Geotechnique 49 (4), 453–469.

Hassan, M.M., Gumaei, A., Alsanad, A., Alrubaian, M., Fortino, G., 2020. A hybrid deep learning model for efficient intrusion detection in big data environment. Inf. Sci. 513, 386–396.

Hochreiter, S., Schmidhuber, J., 1997a. Long short-term memory. Neural Comput. 9, 1735–1780.

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

Houlsby, G.T., Kelly, R.B., Huxtable, J., Byrne, B.W., 2006. Field trials of suction caissons in sand for offshore wind turbine. Geotechnique 56 (1), 3–10.

Huang, F., Huang, J., Jiang, S., Zhou, C., 2017. Landslide displacement prediction based on multivariate chaotic model and extreme learning machine. Eng. Geol. 218, 173–186.

Ibsen, L.B., Barari, A., Larsen, K.A., 2014. Adaptive plasticity model for bucket foundations. J. Eng. Mech. 140 (2), 361–373.

Ibsen, L.B., Brincker, R., 2004. Design of a New Foundation for Offshore Wind Turbines, 22nd International Modal Analysis Conference. MI, USA., Detroit.

Jin, Y.-F., Wu, Z.-X., Yin, Z.-Y., Shen, J.S., 2017. Estimation of critical state-related formula in advanced constitutive modeling of granular material. Acta Geotechnica 12 (6), 1329–1351.

Jin, Y.-F., Yin, Z.-Y., Shen, S.-L., Hicher, P.-Y., 2016. Selection of sand models and identification of parameters using an enhanced genetic algorithm. Int. J. Numer. Anal. Methods GeoMech. 40 (8), 1219–1240.

Jin, Y.-F., Yin, Z.-Y., Zhou, W.-H., Horpibulsuk, S., 2019a. Identifying parameters of advanced soil models using an enhanced transitional Markov chain Monte Carlo method. Acta Geotechnica 14 (6), 1925–1947.

Jin, Z., Yin, Z.Y., Kotronis, P., Jin, Y.F., 2018. Numerical investigation on evolving failure of caisson foundation in sand using the combined Lagrangian-SPH method. Marine Georesources & Geotechnology 37 (1), 23–35.

Jin, Z., Yin, Z.-Y., Kotronis, P., Li, Z., 2019b. Advanced numerical modelling of caisson foundations in sand to investigate the failure envelope in the H-M-V space. Ocean Eng. 190 (15), 106394.

Jin, Z., Yin, Z.-Y., Kotronis, P., Li, Z., Tamagnini, C., 2019c. A hypoplastic macroelement model for a caisson foundation in sand under monotonic and cyclic loadings. Mar. Struct. 66, 16–26.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444.

Li, Z., Kotronis, P., Escoffier, S., Tamagnini, C., 2015. A hypoplastic macroelement for single vertical piles in sand subject to three-dimensional loading conditions. Acta Geotechnica 11 (2), 373–390.

Liu, X., Gasco, F., Goodsella, J., Yua, W.B., 2019. Initial failure strength prediction of woven composites using a new yarnfailure criterion constructed by deep learning. Compos. Struct. 230, 111505.

Liu, M., Yang, M., Wang, H., 2014. Bearing behavior of wide-shallow bucket foundation for offshore wind turbines in drained silty sand. Ocean Engineering 82, 169–179.

Montrasio, L., Nova, R., 1997. Settlements of shallow foundations on sand geometrical effects. Geotechnique 47 (1), 49–60.

Moradi, R., Berangi, R., Minaei, B., 2019. A survey of regularization strategies for deep models. Artif. Intell. Rev. https://doi.org/10.1007/s10462-019-09784-7.

Murata, N., Yoshizawa, S., Amari, S., 1993. Learning curves, model selection and complexity of neural networks. In: Hanson, S.J., Cowan, J.D., Giles, C.L. (Eds.), Advances in Neural Information Processing Systems, 5. Morgan Kaufmann, San Mateo, CA, pp. 607–614, 1993.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I - a discussion of principles. J. Hydrol. 10 (3), 282–290.

Nova, R., Montrasio, L., 1991. Settlements of shallow foundations on sand. Geotechnique 41 (2), 243–256.

Reuter, U., Sultan, A., Reischl, D.S., 2018. A comparative study of machine learning approaches for modeling concrete failure surfaces. Adv. Eng. Software 116, 67–79.

Ruder, S., 2016. An Overview of Gradient Descent Optimization arXiv preprint, arXiv: 1609.04747v04742.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. Nature 323 (9), 533–536.

Sarir, P., Shen, S.-L., Wang, Z.-F., Chen, J., Horpibulsuk, S., Pham, B.T., 2019. Optimum model for bearing capacity of concrete-steel columns with AI technology via incorporating the algorithms of IWO and ABC. Eng. Comput. 1–11.

Skau, K.S., Chen, Y., Jostad, H.P., 2018a. A numerical study of capacity and stiffness of circular skirted foundations in clay subjected to combined static and cyclic general loading. Geotechnique 68 (3), 205–220.

Skau, K.S., Grimstad, G., Page, A.M., Eiksund, G.R., Jostad, H.P., 2018b. A macro-element for integrated time domain analyses representing bucket foundations for offshore wind turbines. Mar. Struct. 59, 158–178.

Smith, L.N., 2017. Cyclical learning rates for training neural networks. In: IEEE Winter Conference on Applications of Computer Vision (WACV), Santa Rosa, California.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15, 1929–1958.

Villalobos, F.A., Byrne, B.W., Houlsby, G.T., 2009. An experimental study of the drained capacity of suction caisson foundations under monotonic loading for offshore applications. Soils Found. 49 (3), 477–488.

Wang, K., Sun, W., 2018. A multiscale multi-permeability poroplasticity model linked by recursive homogenizations and deep learning. Comput. Methods Appl. Mech. Eng. 334, 337–380.

Xu, P., Du, R., Zhang, Z., 2019. Predicting pipeline leakage in petrochemical system through GAN and LSTM. Knowl. Base Syst. 175, 50–61.

Yang, B., Yin, K., Lacasse, S., Liu, Z., 2019. Time series analysis and long short-term memory neural network to predict landslide displacement. Landslides 16 (4), 677–694.

Yin, Z.-Y., Huang, H.-W., Hicher, P.-Y., 2016. Elastoplastic modeling of sand–silt mixtures. Soils Found. 56 (3), 520–532.

Yin, Z.-Y., Jin, Y.-F., Shen, J.S., Hicher, P.-Y., 2018a. Optimization techniques for identifying soil parameters in geotechnical engineering: comparative study and enhancement. Int. J. Numer. Anal. Methods GeoMech. 42 (1), 70–94.

Yin, Z.-Y., Jin, Z., Kotronis, P., Wu, Z.-X., 2018b. Novel SPH SIMSAND–based approach for modeling of granular collapse. Int. J. GeoMech. 18 (11).

Yin, Z.-Y., Xu, Q., Hicher, P.-Y., 2013. A simple critical-state-based double-yield-surface model for clay behavior under complex loading. Acta Geotechnica 8 (5), 509–523.

Zafeirakos, A., Gerolymos, N., 2016. Bearing strength surface for bridge caisson foundations in frictional soil under combined loading. Acta Geotechnica 11 (5), 1189–1208.

Zhang, P., Shen, S.-L., Zhou, A., Xu, Y.-S., 2019a. Investigation on performance of neural networks using quadratic relative error cost function. IEEE Access 7, 106642–106652.

Zhang, P., 2019. A novel feature selection method based on global sensitivity analysis with application in machine learning-based prediction model. Appl. Soft Comput. 85, 105859.

Zhang, P., Chen, R.P., Wu, H.N., 2019b. Real-time analysis and regulation of EPB shield steering using Random Forest. Autom. ConStruct. 106, 102860.

Zhang, P., Wu, H.N., Chen, R.P., Chan, T.H.T., 2020. Hybrid meta-heuristic and machine learning algorithms for tunneling-induced settlement prediction: A comparative study. Tunnelling and Underground Space Technology 99, 103383.

Zhang, P., Yin, Z.-Y., Jin, Y.-F., Chan, T.H.T., 2020a. A novel hybrid surrogate intelligent model for creep index prediction based on particle swarm optimization and random forest. Eng. Geol. 265, 105328.

Zhang, P., Yin, Z.Y., Jin, Y.F., Chan, T., 2020b. Intelligent modelling of clay compressibility using hybrid meta-heuristic and machine learning algorithms. Geosci. Front. (in press).

Zhang, P., Yin, Z.Y., Jin, Y.F., Ye, G.L., 2020c. An AI-based model for describing cyclic characteristics of granular materials. Int. J. Numer. Anal. Methods GeoMech. 1–21.

Zhou, W.-H., Garg, A., Garg, A., 2016. Study of the volumetric water content based on density, suction and initial water content. Measurement 94, 531–537.

Zhu, F., Bienen, B., O'Loughlin, C., Morgan, N., Cassidy, M.J., 2018. The response of suction caissons to multidirectional lateral cyclic loading in sand over clay. Ocean Eng. 170, 43–54.

Zhu, J.-H., Zaman, M.M., Anderson, S.A., 1998. Modeling soil behavior with a recurrent neural network. Can. Geotech. J. 35, 858–872.