Deep Learning-Based Prediction of Wind-Induced Lateral Displacement Response of Suspension Bridge Decks for Structural Health Monitoring

Zhi-wei Wang [a,b], Xiao-fan Lu [a], Wen-ming Zhang [a,*], Vasileios C. Fragkoulis [c], Yu-feng Zhang [d], Michael Beer [b,e,f]

[a] *Key Laboratory of Concrete and Prestressed Concrete Structures of the Ministry of Education, Southeast University, Nanjing 211189, China*

[b] *Institute for Risk and Reliability, Leibniz Universität Hannover, Hannover 30167, Germany*

[c] *Department of Civil and Environmental Engineering, University of Liverpool, Liverpool L69 3GH, United Kingdom*

[d] *State Key Laboratory of Safety and Health for In-service Long Span Bridges, Jiangsu Transportation Institute Co. Ltd., Nanjing 211112, China*

[e] *Institute for Risk and Uncertainty and School of Engineering, University of Liverpool, Liverpool L69 7ZF, United Kingdom*

[f] *International Joint Research Center for Resilient Infrastructure & International Joint Research Center for Engineering Reliability and Stochastic Mechanics, Tongji University, Shanghai 200092, China*

Abstract: Monitoring the wind-induced lateral displacement (WLD) of the bridge deck is crucial for structural health monitoring (SHM) of suspension bridges. An accurate WLD prediction model can aid the bridge SHM systems in abnormal data detection and reconstruction, structural response estimation under specific wind events, and structural condition assessment. However, WLD prediction faces challenges due to stochastic wind action and complex aerodynamic effects acting on the bridge deck. To address this, a deep learning-based framework was proposed for predicting the WLD response of the suspension bridge deck. This framework decomposed the WLD response into two components, namely the quasi-static and the dynamic one. Two separate deep-learning tasks were employed to predict these components using the lateral wind speed as input. In Task 1, a recurrent neural network (RNN) based on the gated recurrent unit (GRU) was built, whereas a fully convolutional neural network (CNN) based on U-Net was built in Task 2. Novel loss functions tailored to each task were established to facilitate accurate predictions. Measured data from the SHM system of the Jiangyin Yangtze River Bridge, China, was used as a case study to verify the proposed predictive framework's feasibility and high accuracy. The extreme value-weighted loss function in Task 1 enhanced the prediction accuracy for the extreme quasi-static WLD, while the time-frequency cross-domain loss functions in Task 2 effectively integrated the prediction accuracies in both time and frequency domains for the dynamic component of WLD. However, trade-offs were identified between the prediction errors of extreme and non-extreme values, as well as between the time- and frequency-domain prediction accuracies.

Keywords: Suspension bridge deck; Wind-induced lateral displacement; Structural health monitoring; Deep learning; Extreme values; U-Net.

* Corresponding author.

E-mail addresses: zhiwei_wang@seu.edu.cn (Z. Wang), lu_xiaofan@seu.edu.cn (X. Lu), zwm@seu.edu.cn (W. Zhang), Vasileios.Fragkoulis@liverpool.ac.uk (V.C. Fragkoulis), kolya@jsti.com (Y. Zhang), beer@irz.uni-hannover.de (M. Beer).

Nomenclature

Abbreviations

CDF	Cumulative distribution function
CNN	Convolutional neural network
GRU	Gated recurrent unit network
HPO	Hyperparameter optimization
LSD	Log-spectral distance
LSD_{Norm}	Min-max normalized LSD with range [0, 1]
MAE	Mean absolute error
MSE	Mean squared error
PDF	Probability density function
PSD	Power spectral density
RMSE	Root mean square error
$\text{RMSE}_{\text{Norm}}$	Min-max normalized RMSE with range [0, 1]
RNN	Recurrent neural network
SF	Sampling frequency
SHM	Structural health monitoring
STFT	Short-time Fourier transform
WLD	Wind-induced lateral displacement

Symbols

a	Maximum weight factor in $\mathcal{L}_{\text{W-MSE}}$
$F(\lvert\mathbf{y}(n)\rvert)$	CDF of the absolute value of measured quasi-static WLD
$\mathcal{L}_{\text{MSE}}$	MSE loss function in Task 1
$\mathcal{L}_{\text{W-MSE}}$	Extreme value-weighted MSE loss function in Task 1
$\mathcal{L}^{\text{T}}_{\text{MAE}}$	Time-domain MAE loss function in Task 2
$\mathcal{L}^{\text{F}}_{\text{Mag}}$	Frequency-domain spectral magnitude MAE loss function in Task 2
$\mathcal{L}^{\text{TF}}_{\text{MAE-Mag}}$	Time-frequency cross-domain loss function in Task 2 (combination of $\mathcal{L}^{\text{T}}_{\text{MAE}}$ and $\mathcal{L}^{\text{F}}_{\text{Mag}}$)
$\mathcal{L}_{\text{Re}}$	MAE loss function of $\mathbf{Y}_{\text{Re}}(u,v)$ in Task 2
$\mathcal{L}_{\text{Im}}$	MAE loss function of $\mathbf{Y}_{\text{Im}}(u,v)$ in Task 2
$\mathcal{L}^{\text{TF}}_{\text{RI-Mag}}$	Time-frequency cross-domain loss function in Task 2 (combination of $\mathcal{L}^{\text{F}}_{\text{Mag}}$, $\mathcal{L}_{\text{Re}}$, and $\mathcal{L}_{\text{Im}}$)
$\mathbf{w}(n)$	Extreme value weight factor in $\mathcal{L}_{\text{W-MSE}}$
$\mathbf{y}(n)$	Measured time-domain quasi-static or dynamic component of WLD
$\mathbf{Y}(u,v)$	STFT result of WLD's dynamic component $\mathbf{y}(n)$
$\mathbf{Y}_{\text{Re}}(u,v)$	Real part of $\mathbf{Y}(u,v)$
$\mathbf{Y}_{\text{Im}}(u,v)$	Imaginary part of $\mathbf{Y}(u,v)$
$\mathbf{Y}_{\text{Mag}}(u,v)$	Time-dependent magnitude spectrum of WLD's dynamic component $\mathbf{y}(n)$ obtained via STFT
α	Time-frequency combination factor in $\mathcal{L}^{\text{TF}}_{\text{MAE-Mag}}$
β	Time-frequency combination factor in $\mathcal{L}^{\text{TF}}_{\text{RI-Mag}}$

1. Introduction

The suspension bridge's deck in service undergoes frequent lateral reciprocating motions under external excitation (Fenerci and Øiseth, 2018; Fenerci et al., 2023). Several studies have shown that the lateral displacement response of the bridge deck is mainly determined by wind action (Wang et al., 2021a, 2021b; Zhang et al., 2024). Due to its significant flexibility, the suspension bridge has a greater wind-induced lateral displacement (WLD) response. For instance, numerical analyses indicated that the maximum WLD response of the deck of the Golden Gate Bridge (suspension bridge) could reach up to 1.7 m (Vincent, 1958), while that of the Sutong Bridge (cable-stayed bridge) was 1.2 m under extreme wind conditions (Xu et al., 2013). Clearly, the excessive WLD of the deck impairs driving comfort and safety or may even threaten structural safety (Yang et al., 2022; Zhu et al., 2021). Moreover, the prolonged lateral swing of the bridge hangers driven by the deck's dynamic WLD can further cause fatigue damage to the hanger anchors (Liu et al., 2017). The short hangers of the Egongyan Rail-transit Suspension Bridge were broken after two years of operation due to the lateral fatigue failure of the rigid hanger anchor. Therefore, the WLD response monitoring of the deck is a non-negligible part of the suspension bridge's SHM. Besides detecting and reconstructing abnormal or missing WLD monitoring data and estimating the WLD response highly efficiently under specific wind conditions without using finite element analysis (Castellon et al., 2021), an accurate data-driven WLD response prediction model can assist the bridge SHM system in many aspects. Indicatively, it can be used to predict the fatigue life of bridge hangers through rapid fatigue simulation analysis, to assess the change in structural stiffness by tracking the extent to which the measured WLD response deviates from the model predictions, as well as to provide guidance for bridge operation and maintenance.

There are three primary methods for the WLD response prediction of the bridge decks. The first one is the analytical approach, which features clear physical interpretability. Based on the deflection theory of suspension bridges, Cheng and Xiao (2006) used the equivalently simplified beam method to estimate the deck's WLD response induced by static wind action. Further, Zhang et al. (2022) proposed an analytical method to derive the deck's static WLD response by solving the suspension bridge's analytical governing equations established based on geometric compatibility, unstrained length conservation, and force balance. However, the current analytical approaches can only solve the static WLD response under a static wind load. Moreover, the results usually diverge from the real values since the analytical methods involve making many necessary modeling assumptions. The second method for WLD response prediction is the finite element theory-based numerical simulation method. This approach converts the wind speed time history into

buffeting force time history, with the latter being the bridge finite element model input for time-domain buffeting analysis (Minh et al., 1999; Tang et al., 2023; Yuan et al., 2023). In this context, the dynamic WLD response of the bridge deck can be solved. However, the numerical approach often relies on precise finite element modeling and accurate structural aerodynamic parameters, and wind tunnel tests are usually required to obtain the latter. For large-scale suspension bridges, it is rather challenging to update accurately on a real-time scale the time-varying material properties and boundary constraints, in addition to the inevitable modeling errors; let alone that finite element calculation is also a time-consuming process. Finally, the third method for WLD response prediction is the data-driven method, which builds the predictive model for the WLD response of the bridge deck from a large amount of measured data. The data-driven method is expected to overcome the above limitations of analytical and numerical methods.

With the popularity of bridge SHM systems, some conventional data-driven methods have found increasing applications in WLD response modeling. For example, Wang et al. (2021b) established two Bayesian predictive models for the deck's WLD response of the Tsing Ma Bridge (suspension bridge) using Global Positioning System (GPS) data and wind data from the SHM system. Unfortunately, the proposed models mainly focused on the quasi-static component of the WLD, while the dynamic component was simplified through a statistical peak factor. Besides this, the literature on conventional data-driven methods was mostly concerned with WLD prediction for cable-stayed bridges. For example, Wang et al. (2021a) established a regression model between the lateral wind speed and the corrected WLD response of the bridge deck for the Anqing Yangtze River Bridge (cable-stayed bridge). However, this predictive model was only a linear model with limited accuracy. Considering the Sutong Bridge (cable-stayed bridge) as an example, Wang and Ding (2014) established the correlation model between the WLD's quasi-static component and static wind action by combining the Fourier series and ARMA models. Meanwhile, the WLD's dynamic component was simulated based on the fitted power spectrum under the assumption that a stable power spectrum exists in the time series of the dynamic WLD. However, the fitted power spectra are inadequate for the accurate prediction of the dynamic WLD of the bridge deck under different wind conditions in the future.

Deep learning-based data-driven methods have been proven to have the ability to mine and learn the complex nonlinear correlation between related variables (Wang et al., 2023). Recurrent neural networks (RNNs) and convolutional neural networks (CNNs) are two classical deep learning models. Typical RNNs excel in capturing the long-term temporal dependencies in time series data and have been widely applied to

time series regression and prediction for SHM (Xu et al., 2023a; Deng et al., 2022; Wang et al., 2022a). Further, CNNs are more suitable for extracting deep, hidden features from data and have been widely used in structural response prediction (Lei et al., 2022), SHM signal reconstruction (Oh and Kim, 2021; Tang et al., 2021; Wang et al., 2022b), and computer vision-based damage detection (Jiang et al., 2022; Rahman et al., 2021). Besides, there has been a recent exploration into physics-informed neural networks aimed at improving generalization and interpretability, wherein physical knowledge is integrated either by combining physics losses with data losses during training or by directly encoding physical principles into the network architecture (Cuomo et al., 2022; Faroughi et al., 2024). Related studies involve structural seismic response prediction with a physics-informed RNN (Zhang et al., 2020), bridge's traffic-induced response prediction using a physics-informed CNN (Ni et al., 2022), and wind turbine vibration modeling utilizing physics-informed residual RNN or wavelet CNN (Li and Zhang, 2022; Xu et al., 2023b).

Nevertheless, existing deep learning-based studies are mostly focused on modeling the vertical displacement of the bridge deck, and few of them are concerned with lateral displacement modeling. Indicatively, an RNN-based correlation model between the vertical displacement of the bridge deck, vehicle load, and temperature of the Nanxi Bridge (suspension bridge) was proposed in the work of Deng et al. (2022) by leveraging the concept of the influence line; the vertical displacement-temperature mapping model for the bridge deck of the Egongyan Rail-transit Bridge (suspension bridge) was built in the work of Wang et al. (2022a) using an RNN; and the acceleration and strain data were used as inputs in the work of Ni et al. (2022) and a physics-informed CNN was applied to build the prediction model for vehicle-induced vertical displacement of the bridge. However, considering that there are significant differences between the vertical displacement and WLD of the bridge deck in data features, excitation source, and load-response correlation mechanism, the above deep learning models for estimating the vertical displacement response are no longer suitable for modeling the WLD response. To the best of the authors' knowledge, only Lei et al. (2022) studied the application of CNN for predicting the WLD of the bridge deck of the Anqing Yangtze River Bridge (cable-stayed bridge). However, their model was only concerned with predicting the quasi-static component of the WLD response, while neglecting the dynamic component.

The purpose of the present study is to build a predictive model for the WLD response of the suspension bridge deck using SHM data. The methods described in the literature are rather inadequate in this case for the following reasons. First, both the analytical and numerical approaches have their intrinsic limitations, as demonstrated above. Then, it is noted that the conventional data-driven methods have limited accuracy in

predicting the dynamic component of the WLD response of the bridge deck and focus mostly on cable-stayed bridges. Finally, there is currently a lack of studies on WLD response prediction of the suspension bridge deck based on deep learning. The existing deep-learning models for predicting the bridge deck's vertical displacement are inadequate for lateral displacement.

To overcome the above limitations, we proposed a predictive framework for the WLD response of the suspension bridge deck based on deep learning. This model decomposed the WLD response into a quasi-static component and a corresponding dynamic component. Using lateral wind speeds as inputs, the herein proposed RNN model and the CNN model efficiently predicted the two components. The contributions of the present study can be summarized as follows:

(1) To the best of the authors' knowledge, this is the first application of a deep learning-based framework to predict the WLD response of the suspension bridge deck for SHM.

(2) An extreme value-weighted loss function was proposed when building the RNN-based predictive model for the quasi-static component of the WLD. Its main feature pertains to that increasing the weights of extreme events could potentially increase the prediction accuracy for extreme values of the quasi-static WLD. Four probability distribution-based extreme value weighting schemes were proposed and compared.

(3) Two time-frequency cross-domain loss functions were constructed when building the CNN-based predictive model for the WLD's dynamic component. The prediction accuracy was increased by considering both the time- and frequency-domain prediction errors of the dynamic component. An integrated normalized error metric was proposed to compare the comprehensive predictive performance in both time and frequency domains between different loss functions.

(4) A model accuracy evaluation method was proposed from the perspective of fatigue analysis for bridge hangers undergoing lateral swing, i.e., comparing the 'swing amplitude-cycle count' histograms of the bridge deck (obtained using the rain flow counting method) corresponding to the predicted and measured WLD time histories.

2. Methodology

2.1 General framework

The proposed framework for the WLD response prediction of the deck of the suspension bridge based on deep learning is shown in Fig. 1. According to the bridge wind-induced vibration (buffeting) analysis

theory (Tao et al., 2020), the quasi-static component of the WLD response of the bridge deck is correlated to the time-varying mean wind speed (i.e., aerostatic wind action) in the lateral direction of the bridge. Further, the dynamic component of the WLD response is correlated to both the time-varying mean wind speed and the fluctuating wind speed in the lateral direction. In this study, the WLD response of the bridge deck was decomposed into quasi-static and dynamic components, which were predicted by two deep-learning tasks. The deep learning model was chosen depending on the task attributes and model features.

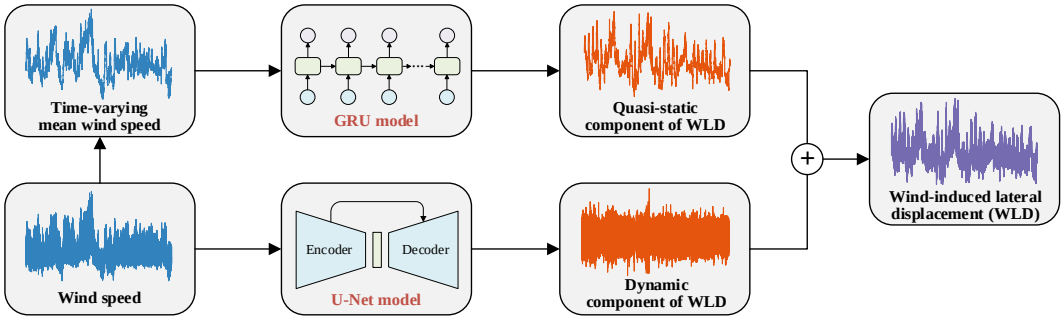

Fig. 1. The proposed deep learning-based framework.

Task 1: The lateral time-varying 10-min mean wind speed was used as the model input, and then the quasi-static component of the WLD response of the bridge deck was predicted. Considering the long-term dependencies between the timesteps and the relatively simple correlation structure between the input and output, an RNN model based on the gated recurrent unit (GRU) network was built, which has the advantages of temporal memory capacity and faster learning.

Task 2: The actual lateral wind speed (consisting of the time-varying mean wind speed and the fluctuating wind speed) was used as the model input, and then the dynamic component of the WLD response of the bridge deck was predicted. This task had a much more complex correlation structure between the input and output data. Therefore, a CNN model based on the U-Net was built for mining deep-level data features and establishing complex, nonlinear mapping.

2.2 Task 1: Predicting WLD's quasi-static component using a GRU model

2.2.1 Introduction to the GRU network

GRU is a type of RNN and a simplified version of the long short-term memory (LSTM) network (another typical type of RNN), which performs similarly to LSTM but has a faster computational speed (Chung et al., 2014). As shown in Fig. 2(a), the conventional RNN is designed to store information from previous time steps through a unique design called the 'hidden state'. This temporal memory capability makes it adept at handling time series prediction problems. On this basis, as shown in Fig. 2(b), the GRU

network uses two learnable ‘gates’, namely, the reset gate and the update gate, to implement the gating of the hidden state. The concept of the gated hidden state introduces a learnable mechanism into the GRU network to autonomously decide if the hidden state is to be updated or reset. Such a design mitigates the vanishing or exploding gradient problem that may occur during error backpropagation in the conventional RNN model. In addition, the gating mechanism increases the model’s capacity to capture the dependencies between sequence elements at larger time distances (Jiang et al., 2021).

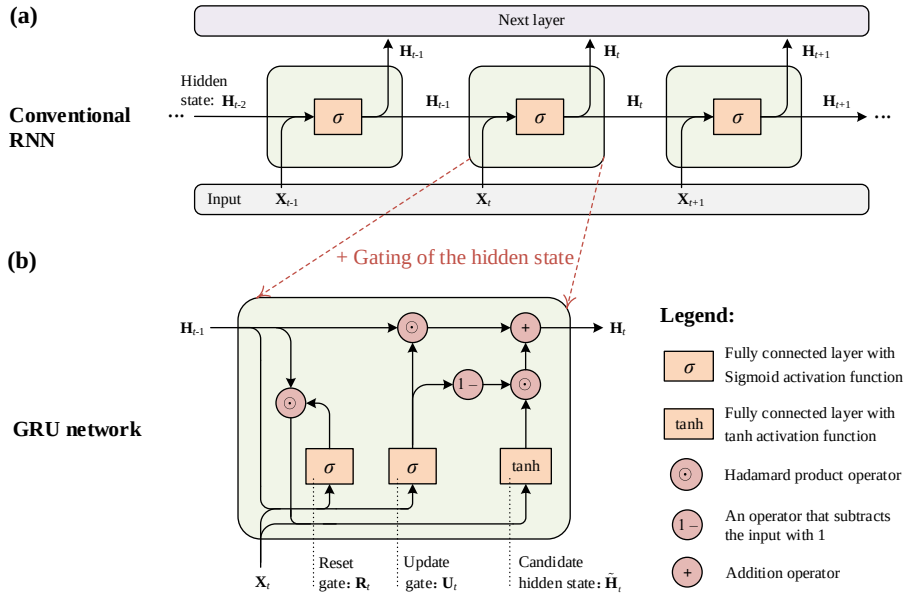

Fig. 2. Comparison of the internal structures of (a) the conventional RNN and (b) the GRU network.

Specifically, consider a mini-batch of the input $\mathbf{X} \in \mathbb{R}^{B\times d\times l}$ of the GRU model, where B is the mini-batch size, d is the number of input features, and l is the sequence length. Let the input at time step t $(1 \le t \le l)$ be $\mathbf{X}_t \in \mathbb{R}^{B\times d}$, and the hidden state transmitted from the previous time step be $\mathbf{H}_{t-1} \in \mathbb{R}^{B\times h}$, with h denoting the number of hidden neurons. Then, the reset gate $\mathbf{R}_t \in \mathbb{R}^{B\times h}$ and the update gate $\mathbf{U}_t \in \mathbb{R}^{B\times h}$ execute the following calculations:

$$\mathbf{R}_t = \sigma(\mathbf{X}_t\mathbf{W}_{\mathrm{xr}} + \mathbf{H}_{t-1}\mathbf{W}_{\mathrm{hr}} + \mathbf{b}_{\mathrm{r}}) \tag{1}$$

$$\mathbf{U}_t = \sigma(\mathbf{X}_t\mathbf{W}_{\mathrm{xu}} + \mathbf{H}_{t-1}\mathbf{W}_{\mathrm{hu}} + \mathbf{b}_{\mathrm{u}}) \tag{2}$$

where $\mathbf{W}_{\mathrm{xr}} \in \mathbb{R}^{d\times h}$ and $\mathbf{W}_{\mathrm{hr}} \in \mathbb{R}^{h\times h}$ are the weight parameters of $\mathbf{X}_t$ and $\mathbf{H}_{t-1}$ in the reset gate, respectively; $\mathbf{W}_{\mathrm{xu}} \in \mathbb{R}^{d\times h}$ and $\mathbf{W}_{\mathrm{hu}} \in \mathbb{R}^{h\times h}$ are the weight parameters of $\mathbf{X}_t$ and $\mathbf{H}_{t-1}$ in the update gate, respectively; $\mathbf{b}_{\mathrm{r}} \in \mathbb{R}^{B\times h}$ and $\mathbf{b}_{\mathrm{u}} \in \mathbb{R}^{B\times h}$ are the offset parameters in the reset gate and the update gate, respectively; and σ is the sigmoid activation function that maps $\mathbf{R}_t$ and $\mathbf{U}_t$ to the interval (0, 1).

The reset gate $\mathbf{R}_t$ decides to what extent the hidden state $\mathbf{H}_{t-1}$ transmitted from the previous to the current time step will be retained. A value of $\mathbf{R}_t$ closer to 1 indicates that the previous hidden state tends to be retained, while a value closer to 0 indicates that the previous hidden state tends to be reset. The candidate hidden state $\tilde{\mathbf{H}}_t \in \mathbb{R}^{B\times h}$ at time step t is obtained using $\mathbf{R}_t$ and is calculated as follows:

$$\tilde{\mathbf{H}}_t = \tanh(\mathbf{X}_t \mathbf{W}_{\text{xh}} + (\mathbf{R}_t \odot \mathbf{H}_{t-1})\mathbf{W}_{\text{hh}} + \mathbf{b}_{\text{h}}) \tag{3}$$

where $\odot$ is the Hadamard product (i.e., element-wise multiplication); $\mathbf{W}_{\text{xh}} \in \mathbb{R}^{d\times h}$ and $\mathbf{W}_{\text{hh}} \in \mathbb{R}^{h\times h}$ are weight parameters of the candidate hidden state; $\mathbf{b}_{\text{h}} \in \mathbb{R}^{B\times h}$ is the offset parameter; and the tanh activation function is used to map the data to the interval (–1, 1). Notably, the above weight and offset parameters are shared across different time steps.

The update gate $\mathbf{U}_t$ determines the amount of the past information (from previous time steps) that is passed from the previous hidden state $\mathbf{H}_{t-1}$ and the candidate hidden state $\tilde{\mathbf{H}}_t$ to the new hidden state $\mathbf{H}_t$. The final hidden state $\mathbf{H}_t$ output at time step t is given by:

$$\mathbf{H}_t = \mathbf{U}_t \odot \mathbf{H}_{t-1} + (1 - \mathbf{U}_t) \odot \tilde{\mathbf{H}}_t \tag{4}$$

A value of $\mathbf{U}_t$ closer to 1 indicates that the hidden state is not updated, and the previous one, $\mathbf{H}_{t-1}$, is retained, whereas a value close to 0 indicates that the hidden state is completely updated to $\tilde{\mathbf{H}}_t$.

2.2.2 Architecture of the GRU-based prediction model

The nonlinear mapping relationship between the time-varying mean wind speed and the quasi-static component of the WLD is established in Task 1, where the number of input features and the number of output responses are both equal to one. Fig. 3 shows the model architecture constructed in Task 1.

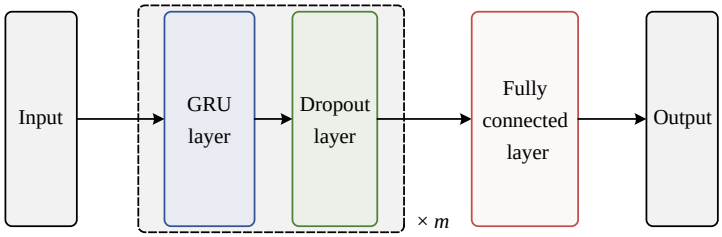

Fig. 3. Architecture of the GRU model in Task 1.

A number of m GRU layers with the same number (h) of hidden neurons are located at the model's core. m and h are model hyperparameters, and hyperparameter optimization (HPO) is required to obtain the optimal model. The final layer corresponds to the fully connected layer, which reduces the number of feature channels of the GRU layer's output from h to 1. A dropout layer follows each GRU layer, where the

proportion of dropout is p. Dropout is a regularization method for mitigating overfitting in the neural network. During the model training process, the dropout operation randomly zeros some of the outputs from the previous layer with probability p. This operation reduces complex coadaptations between the neurons, enhancing the model's generalization ability. However, the dropout operation is not applied during the model inference to ensure the stability of the model outputs.

2.3 Task 2: Predicting WLD's dynamic component using a U-Net model

2.3.1 Architecture of the U-Net model

U-Net is a type of CNN, whose basic architecture was first proposed by Ronneberger et al. (2015). It is a fully convolutional network based on an autoencoder, which features a U-shaped architecture consisting of a contracting path (encoder) and an expanding path (decoder). The input feature and the output response in Task 2 are one-dimensional (1D) time sequences. The architecture of the U-Net model built in this study is shown in Fig. 4.

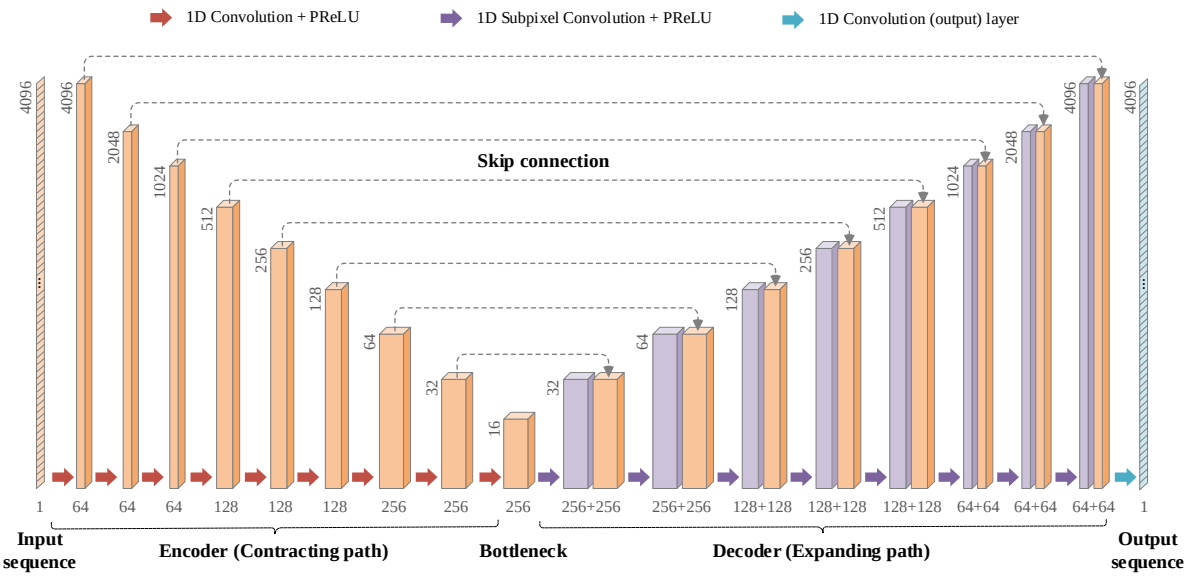

Fig. 4. Architecture of the U-Net model in Task 2; an illustrative example where the input length is 4096. The cuboids represent the feature maps. The numbers at the upper left corner of the cuboid represent the feature length, and the numbers at the bottom denote the number of feature channels. The dimension of the mini-batch size is not shown.

Specifically, the input length of the U-Net model is 2^n ($n \geq 9$) and the configuration of each layer is shown in Table 1. The contracting path in the encoder consists of nine 1D convolution layers, which are used to extract deeply hidden features known as feature maps. The length of the feature map output by each layer gradually decreases down the contracting path (i.e., downsampling), while the number of feature channels increases until reaching the bottleneck. The expanding path in the decoder consists of eight 1D subpixel

convolution layers and one output layer, which are used for upsampling the feature maps and reducing the feature channels. In this way, the output will have the same size as the input. In the decoder, the output of each subpixel convolution layer is additionally cascaded on the channel dimension to feature maps of the same size at the mirror position in the encoder, before being input to the next layer. This operation is called skip connection, which can recover some important information lost during the encoding process. Through skip connections, the model training error can skip some middle network layers and be directly backpropagated to layers in the encoder. This mechanism can help mitigate the vanishing gradient problem, which is common with deep networks. In addition, the dropout operation is implemented once every three layers in the U-Net. The parametric rectified linear unit (PReLU) is used as the activation function.

Table 1. Parameter configuration of the U-Net model (taking the input length 4096 and kernel size 11 as an example).

No.	Layer	Input shape $L_{in} \times C_{in}$	Output shape $L_{out} \times C_{out}$	Kernel size L_k ($\times C_{in}$)	Kernel number	Activation
1	Conv	4096 × 1	4096 × 64	11 (× 1)	64	PReLU
2	Conv	4096 × 64	2048 × 64	11 (× 64)	64	PReLU
3	Conv	2048 × 64	1024 × 64	11 (× 64)	64	PReLU
4	Conv	1024 × 64	512 × 128	11 (× 64)	128	PReLU
5	Conv	512 × 128	256 × 128	11 (× 128)	128	PReLU
6	Conv	256 × 128	128 × 128	11 (× 128)	128	PReLU
7	Conv	128 × 128	64 × 256	11 (× 128)	256	PReLU
8	Conv	64 × 256	32 × 256	11 (× 256)	256	PReLU
9	Conv	32 × 256	16 × 256	11 (× 256)	256	PReLU
10	Subp Conv	16 × 256	32 × (256 + 256)	11 (× 256)	256 × 2	PReLU
11	Subp Conv	32 × 512	64 × (256 + 256)	11 (× 512)	256 × 2	PReLU
12	Subp Conv	64 × 512	128 × (128 + 128)	11 (× 512)	128 × 2	PReLU
13	Subp Conv	128 × 256	256 × (128 + 128)	11 (× 256)	128 × 2	PReLU
14	Subp Conv	256 × 256	512 × (128 + 128)	11 (× 256)	128 × 2	PReLU
15	Subp Conv	512 × 256	1024 × (64 + 64)	11 (× 256)	64 × 2	PReLU
16	Subp Conv	1024 × 128	2048 × (64 + 64)	11 (× 128)	64 × 2	PReLU
17	Subp Conv	2048 × 128	4096 × (64 + 64)	11 (× 128)	64 × 2	PReLU
18	Conv (output)	4096 × 128	4096 × 1	11 (× 128)	1	—

Note: 'Conv' denotes the 1D convolution layer, whereas 'Subp Conv' denotes the 1D subpixel convolution layer; L_{in} and L_{out} are the lengths of the input and output features, respectively; C_{in} and C_{out} correspond to the channel numbers of the input and output features, respectively; and L_k is the length of the kernel. The mini-batch dimensions are omitted for brevity.

2.3.2 1D convolution and 1D subpixel convolution

An intuitive illustration of 1D convolution and 1D subpixel convolution is shown in Fig. 5. In the context of CNN, convolution denotes the operation where the convolution kernel **k** slides over the input sequence **x** (with length L_{in} and channel number C_{in}) for the dot product operation. The sliding size of the

kernel at each stride is represented by s. The length of the kernel L_k determines its receptive field size, while the number of channels of the kernel is automatically aligned with those of $\mathbf{x}$. The convolution output per stride is the sum of convolution results in all channels. To obtain an output sequence of target length, zero-padding at the two ends of the input is needed, with padding length $L_p = (L_k - 1)/2$, where L_k is an odd number. The feature sequence $\mathbf{y}$ extracted by each convolution kernel is given by:

$$\mathbf{y}(n) = \sum_{c=0}^{C_{in}-1} \sum_{i=0}^{L_k-1} \mathbf{x}_c(s \cdot n + i)\mathbf{k}_c(i) + b, \quad n = 0, 1, 2, ..., L_{out} - 1 \tag{5}$$

where b is the offset term; $\mathbf{x}_c$ and $\mathbf{k}_c$ are the c-th channels of $\mathbf{x}$ and $\mathbf{k}$, respectively; and L_{out} is the length of the convolution output $\mathbf{y}$.

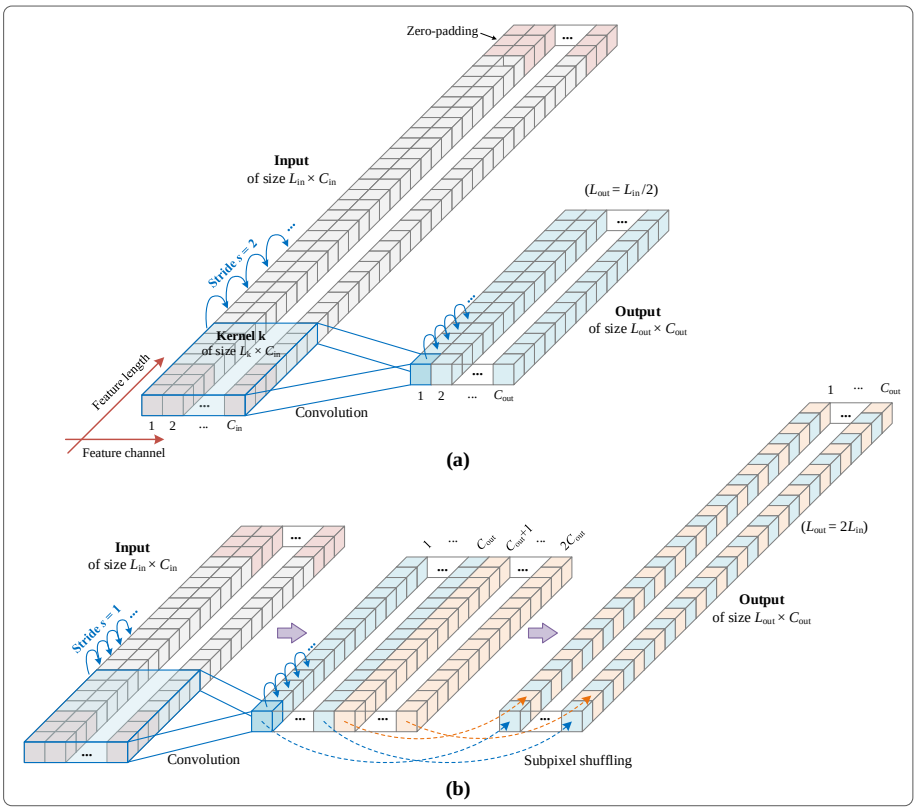

Fig. 5. Schematic diagram of (a) 1D convolution and (b) 1D subpixel convolution. The dimension of the mini-batch size is not shown herein.

In the encoder, the stride size of the convolution kernel is $s = 2$ for each convolution layer, so that $L_{out} = L_{in}/2$. This ensures a layer-by-layer halving of the feature-length. The number of kernels for each convolution layer is consistent with the target channel number C_{out} of the output feature map. The convolution result between each kernel $\mathbf{k}$ and the input $\mathbf{x}$ (i.e., Eq. (5)) constitutes one channel of the output feature map. In the decoder, the stride size is $s = 1$, and thus $L_{out} = L_{in}$. For each subpixel convolution layer, $2C_{out}$ kernels are used to output $2C_{out}$ feature sequences. These feature sequences are first divided into two

groups and then subjected to cross-stacking using subpixel shuffling to generate a feature map with length $2L_{\text{in}}$ and channel number equal to C_{out}. In this way, the feature-length is doubled in a layer-wise manner.

2.4 Loss functions for model training

2.4.1 Loss functions for Task 1

In general, the loss function is the optimal objective function for model training. In Task 1, attention should be directed to the prediction accuracy for extreme quasi-static WLD under extreme/rare wind conditions, since this relates to whether the displacement of the bridge deck lies within the permissible range or not. However, there is a scarcity of observations of extreme wind events, as opposed to those of normal wind events. In other words, the training samples have an unbalanced probability distribution. In this context, Task 1 becomes an imbalanced regression task (Ribeiro and Moniz, 2020). Suppose that the conventional regression strategy is still employed for Task 1, that is, assume that different types of wind event samples are of equal importance. In that case, obtaining a high prediction accuracy for extreme quasi-static WLD will be challenging. Therefore, an extreme value-weighted loss function was proposed for Task 1. Specifically, the weights of the prediction errors of extreme events were increased. Based on the mean squared error (MSE) loss $\mathcal{L}_{\text{MSE}}$ derived by:

$$\mathcal{L}_{\text{MSE}}(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{N}\sum_{n=1}^{N}\left(\mathbf{y}(n) - \hat{\mathbf{y}}(n)\right)^2 \tag{6}$$

the extreme value-weighted loss $\mathcal{L}_{\text{W-MSE}}$ was defined as:

$$\mathcal{L}_{\text{W-MSE}}(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{N}\sum_{n=1}^{N}\mathbf{w}(n)\left(\mathbf{y}(n) - \hat{\mathbf{y}}(n)\right)^2 \tag{7}$$

where $\mathbf{y}$ and $\hat{\mathbf{y}}$ are the true and predicted values of the quasi-static WLD response, respectively; $\mathbf{w}$ is the weight factor; and N is the sequence length.

The weight factor $\mathbf{w}(n)$ of $\mathcal{L}_{\text{W-MSE}}$ in Eq. (7) increases as the extent to which $\mathbf{y}(n)$ belongs to an extreme event increases. Such an extent can be measured by $F(\mathbf{y}(n))$, namely, the value of the cumulative distribution function (CDF) corresponding to $\mathbf{y}(n)$. It represents the probability that the data in the samples are smaller than $\mathbf{y}(n)$. In engineering practice, $\mathbf{y}(n)$ approximately follows a normal distribution with zero mean. The sign of its value is only used to differentiate between different displacement directions. Therefore, the CDF of the absolute value of $\mathbf{y}(n)$, i.e., $F(|\mathbf{y}(n)|)$, was adopted to determine the weight. $F(|\mathbf{y}(n)|)$ can be modeled using a parametric probability distribution model (e.g., half-normal distribution (Cooray and

Ananda, 2008)) or a nonparametric probability distribution model (e.g., kernel density estimation (Jones, 1993)). As shown in Fig. 6, when $F(|\mathbf{y}(n)|) < 0.5$, $|\mathbf{y}(n)|$ is considered a normal event, and the weight is equal to 1; when $F(|\mathbf{y}(n)|) \in [0.5, 1]$, four different extreme value weighting schemes were proposed as follows:

Weighting scheme I: Uniform weighting function

$$\mathbf{w}(n) = \begin{cases} 1, & \text{if } F(|\mathbf{y}(n)|) < 0.5 \\ a, & \text{otherwise} \end{cases} \tag{8}$$

Weighting scheme II: Linear weighting function

$$\mathbf{w}(n) = \begin{cases} 1, & \text{if } F(|\mathbf{y}(n)|) < 0.5 \\ 2(a-1)\left(F(|\mathbf{y}(n)|) - 0.5\right) + 1, & \text{otherwise} \end{cases} \tag{9}$$

Weighting scheme III: Cubic weighting function

$$\mathbf{w}(n) = \begin{cases} 1, & \text{if } F(|\mathbf{y}(n)|) < 0.5 \\ 8(a-1)\left(F(|\mathbf{y}(n)|) - 0.5\right)^3 + 1, & \text{otherwise} \end{cases} \tag{10}$$

Weighting scheme IV: Inverse proportional weighting function (Scheepens et al., 2023)

$$\mathbf{w}(n) = \begin{cases} 1, & \text{if } F(|\mathbf{y}(n)|) < 0.5 \\ \dfrac{a}{100} \cdot \dfrac{1}{1 - F(|\mathbf{y}(n)|) + \varepsilon} - \dfrac{a}{50} + 1, & \text{if } 0.5 \le F(|\mathbf{y}(n)|) \le 1 \\ a, & \text{if } \mathbf{w}(n) > a \end{cases} \tag{11}$$

In Eqs. (8)-(11), a is the maximum weight factor, which is a hyperparameter that needs to be optimized. In Eq. (11), ε is an infinitesimal value used to ensure the stability of the numerical calculation.

2.4.2 Loss functions for Task 2

In Task 2, predicting the dynamic components of the WLD response requires attention to accuracy in both the time and frequency domains. Two baseline loss functions $\mathcal{L}_{\text{MAE}}^{\text{T}}$ and $\mathcal{L}_{\text{Mag}}^{\text{F}}$ were built for comparison, considering only the errors in the time and frequency domains, respectively. $\mathcal{L}_{\text{MAE}}^{\text{T}}$ was defined as the mean absolute error (MAE) between the true and predicted values in the time domain, as:

$$\mathcal{L}_{\text{MAE}}^{\text{T}}(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{N} \sum_{n=1}^{N} |\mathbf{y}(n) - \hat{\mathbf{y}}(n)| \tag{12}$$

where $\mathbf{y}$ and $\hat{\mathbf{y}}$ are the true and predicted values of the dynamic component of the WLD in the time domain, respectively, and N is the sequence length. $\mathcal{L}_{\text{Mag}}^{\text{F}}$ was defined as the MAE between the time-dependent magnitude spectrum $\hat{\mathbf{Y}}_{\text{Mag}}$ of the time-domain predicted result $\hat{\mathbf{y}}$ and the true time-dependent magnitude

spectrum $\mathbf{Y}_{\text{Mag}}$, as:

$$\mathcal{L}_{\text{Mag}}^{\text{F}}(\mathbf{Y}, \hat{\mathbf{Y}}) = \frac{1}{UV}\sum_{u=1}^{U}\sum_{v=1}^{V}\left|\mathbf{Y}_{\text{Mag}}(u,v) - \hat{\mathbf{Y}}_{\text{Mag}}(u,v)\right| \tag{13}$$

$$\mathbf{Y}_{\text{Mag}}(u,v) = \sqrt{\mathbf{Y}_{\text{Re}}^2(u,v) + \mathbf{Y}_{\text{Im}}^2(u,v)} \tag{14}$$

where $\mathbf{Y} = \mathbf{Y}_{\text{Re}} + \text{i}\mathbf{Y}_{\text{Im}}$ is the time-frequency representation of $\mathbf{y}$, as obtained by the short-time Fourier transform (STFT) with $\mathbf{Y}_{\text{Re}}$ and $\mathbf{Y}_{\text{Im}}$ denoting its real and imaginary parts; $\mathbf{Y}(u,v)$ is a 2D $U \times V$ matrix defined in the complex domain, and u and v are indices in the time and frequency dimensions, respectively; i represents the imaginary unit.

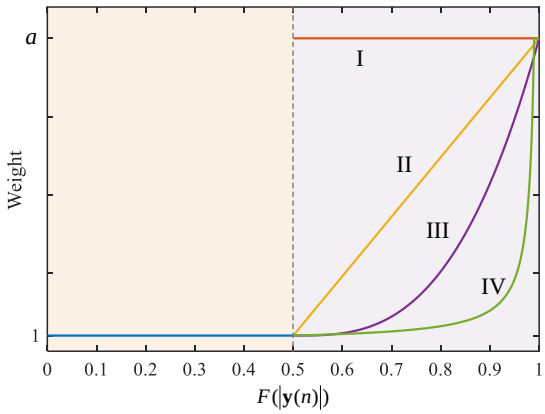

Fig. 6. Weighting curves of different extreme value weighting schemes. $F(|\mathbf{y}(n)|)$ denotes the CDF values of the absolute values of the measured WLD quasi-static component.

Furthermore, two novel time-frequency cross-domain loss functions $\mathcal{L}_{\text{MAE-Mag}}^{\text{TF}}$ and $\mathcal{L}_{\text{RI-Mag}}^{\text{TF}}$ were proposed to take into account both time- and frequency-domain information in the loss function design. $\mathcal{L}_{\text{MAE-Mag}}^{\text{TF}}$ was defined as the combination of the above time- and frequency-domain loss functions, as:

$$\mathcal{L}_{\text{MAE-Mag}}^{\text{TF}}(\mathbf{y}, \hat{\mathbf{y}}, \mathbf{Y}, \hat{\mathbf{Y}}) = \mathcal{L}_{\text{Mag}}^{\text{F}}(\mathbf{Y}, \hat{\mathbf{Y}}) + \alpha\mathcal{L}_{\text{MAE}}^{\text{T}}(\mathbf{y}, \hat{\mathbf{y}}) \tag{15}$$

where α is the time-frequency combination factor. $\mathcal{L}_{\text{RI-Mag}}^{\text{TF}}$ was defined so that it comprises a real part loss component $\mathcal{L}_{\text{Re}}$ and an imaginary part loss component $\mathcal{L}_{\text{Im}}$ of $\hat{\mathbf{Y}}$, in addition to the frequency-domain loss function $\mathcal{L}_{\text{Mag}}^{\text{F}}$, as follows:

$$\mathcal{L}_{\text{RI-Mag}}^{\text{TF}}(\mathbf{Y}, \hat{\mathbf{Y}}) = \mathcal{L}_{\text{Mag}}^{\text{F}}(\mathbf{Y}, \hat{\mathbf{Y}}) + \beta\left[\mathcal{L}_{\text{Re}}(\mathbf{Y}, \hat{\mathbf{Y}}) + \mathcal{L}_{\text{Im}}(\mathbf{Y}, \hat{\mathbf{Y}})\right] \tag{16}$$

$$\mathcal{L}_{\text{Re}}(\mathbf{Y}, \hat{\mathbf{Y}}) = \frac{1}{UV}\sum_{u=1}^{U}\sum_{v=1}^{V}\left|\mathbf{Y}_{\text{Re}}(u,v) - \hat{\mathbf{Y}}_{\text{Re}}(u,v)\right| \tag{17}$$

$$\mathcal{L}_{\text{Im}}(\mathbf{Y}, \hat{\mathbf{Y}}) = \frac{1}{UV}\sum_{u=1}^{U}\sum_{v=1}^{V}\left|\mathbf{Y}_{\text{Im}}(u,v) - \hat{\mathbf{Y}}_{\text{Im}}(u,v)\right| \tag{18}$$

where β is the time-frequency combination factor, a hyperparameter like α that needs to be optimized. It is

noted that the signal phase accuracy of the predicted result $\hat{\mathbf{y}}$ significantly impacts the model's time-domain accuracy. As shown in Fig. 7, by controlling errors in the real and imaginary parts of $\hat{\mathbf{Y}}$, the phase constraint can be imposed on the time-domain predicted result $\hat{\mathbf{y}}$, and further, the time-domain loss can be complemented.

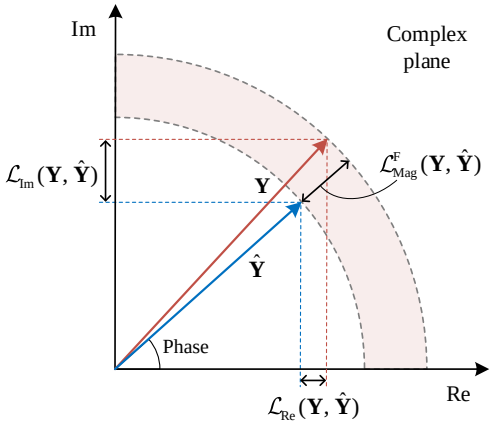

Fig. 7. Schematic diagram of the phase constraint on the complex vector in the complex plane.

Notably, the L2 regularization (a.k.a., weight decay) operation was employed in both tasks. This operation penalizes excessive model parameters by adding a penalty term to the loss function, i.e., the L2-norm of the model weight vector, aiming to reduce model complexity and thus alleviate overfitting.

2.5 Model performance evaluation metrics

A critical aspect of the proposed model relates to its predictive performance on a new dataset. This, in turn, necessitates the selection of appropriate evaluation metrics. In this regard, the root mean square error (RMSE) was chosen as the evaluation metric in the time domain for Tasks 1 and 2. RMSE is a measure of the mean deviation between the predicted values and the true values in units consistent with the evaluated object. The smaller the RMSE, the better the model's predictive performance. It is defined as:

$$\mathrm{RMSE}(\mathbf{y}, \hat{\mathbf{y}}) = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(\mathbf{y}(n) - \hat{\mathbf{y}}(n)\right)^2} \tag{19}$$

In Task 2, one should pay equal attention to frequency domain prediction accuracy. For this purpose, the log spectral distance (LSD) is chosen as the evaluation metric in the frequency domain. The LSD is the logarithmic distance between the predicted and the measured STFT magnitude spectra (i.e., $\hat{\mathbf{Y}}_{\mathrm{Mag}}$ and $\mathbf{Y}_{\mathrm{Mag}}$) and has the unit of dB. The smaller the value of the LSD is, the closer the predicted and measured STFT magnitude spectra and the smaller the frequency-domain error. It is given by:

$$\mathrm{LSD}(\mathbf{Y},\hat{\mathbf{Y}}) = \frac{1}{U}\sum_{u=1}^{U}\sqrt{\frac{1}{V}\sum_{v=1}^{V}\left[\log_{10}\frac{\mathbf{Y}^2_{\mathrm{Mag}}(u,v)}{\hat{\mathbf{Y}}^2_{\mathrm{Mag}}(u,v)}\right]^2} \tag{20}$$

3. Case study

3.1 Sample collection

The case study was based on the Jiangyin Yangtze River Bridge, a suspension bridge with a span of 1385 m, located in Jiangsu Province, China. The included angle measured clockwise between the longitudinal axis of the bridge and the north direction is 24°. Fig. 8 shows the arrangement of GPS sensors and ultrasonic anemometers in the SHM system of the bridge. The GPS sensors were deployed on the top of the south and north pylons and at 1/4, 1/2, and 3/4 of the length of the bridge deck in the main span. The ultrasonic anemometers were installed upstream of the north pylon and at the mid-span of the bridge deck. The GPS sensors record the three-way displacements of the bridge components. The anemometers have two output channels, one for wind speed and the other for wind direction. The sampling frequencies of the GPS sensors and anemometers are 1 Hz.

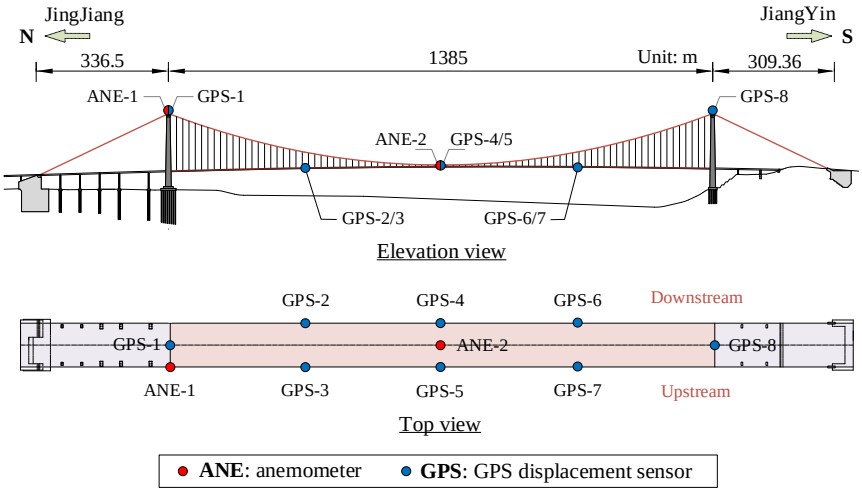

Fig. 8. Sensor arrangement of the SHM system of the Jiangyin Yangtze River Bridge, China.

The wind speed in the lateral direction and the WLD of the bridge deck at the mid-span were used as data samples. Wind samples were collected from the anemometer ANE-2, and wind speed in the lateral direction was obtained by performing wind vector decomposition. The average lateral displacements recorded by the GPS-4 and GPS-5 sensors were used as WLD samples. The positive direction of the lateral wind vector and the WLD vector was defined as pointing from upstream to downstream. Among the SHM data recorded from 2013 to 2014, 7848 hours of data were manually selected as raw samples. These selected

samples exhibited minimal instances of missing or abnormal data, ensuring their reliability for further analysis.

3.2 Sample preprocessing

3.2.1 Downsampling and decomposition

The sample preprocessing was initiated by performing a necessary data-cleaning process. Specifically, a few discrete missing points within the samples were filled by resorting to a shape-preserving piecewise cubic spline interpolation. The outliers within the samples were detected by the generalized extreme Studentized deviate test for outliers and replaced by the abovementioned interpolation method.

Fig. 9 shows the power spectral density (PSD) estimate of the dynamic component of the raw WLD signals based on one-day samples. The signal energy was mostly distributed in the 0-0.1 Hz frequency range. The sampling frequency of the raw dynamic WLD signals was too high, resulting in an increased amount of redundant information. Therefore, the sampling frequency of the WLD and wind signals was appropriately reduced from 1 to 0.5 Hz by resampling, so that the training burden in Task 2 is lessened. Namely, the signals' Nyquist frequency was reduced from 0.5 to 0.25 Hz. In addition, Fig. 9 shows that the frequency of the first-order lateral vibration of the bridge is 0.055 Hz, while the modes of higher-order lateral vibration are less significant, which agrees with the results of Brownjohn et al. (2018).

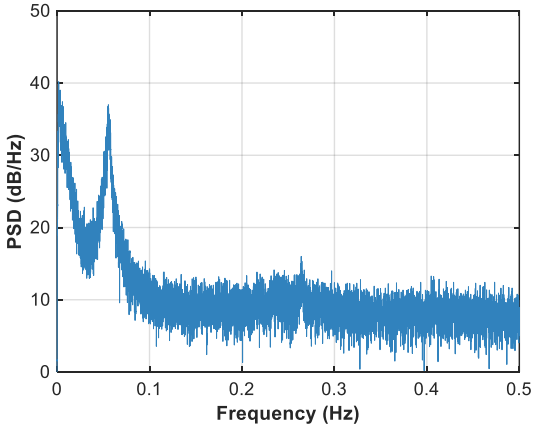

Fig. 9. PSD estimate of the dynamic component of measured WLD signals.

Fig. 10 shows the input-output relationships for the models and the sampling frequencies used for the training samples in Tasks 1 and 2. The moving average method with a 10-min-span time window was used to extract the time-varying 10-min mean wind speed from the preprocessed wind signals and to decompose the WLD signals into quasi-static and dynamic components, as shown in Fig. 11. In addition, for Task 1, the

sampling frequency of 1/600 Hz could well characterize the quasi-static trend features of the wind speed and WLD signals. In contrast, the sampling frequency of 0.5 Hz still results in redundancy. To further improve the training efficiency of the GRU model, the sampling frequency of the samples was reduced from 0.5 to 1/600 Hz through resampling.

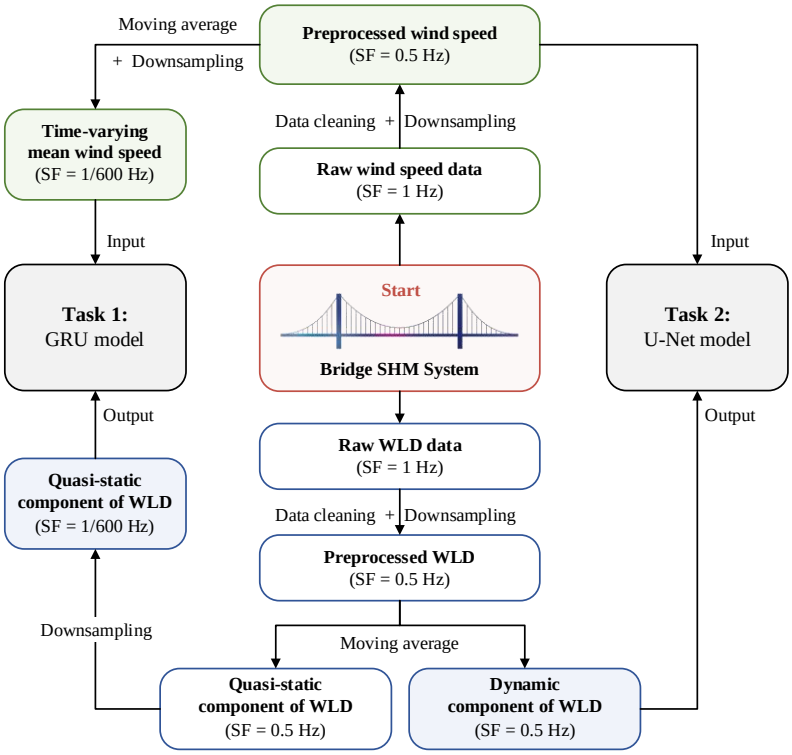

Fig. 10. Input-output relationships for the deep learning models and the sampling frequencies used for the training samples in Tasks 1 and 2. SF denotes the sampling frequency.

3.2.2 Dataset subdivision

All samples were subdivided into a training set, a validation set, and a test set, which were used for model training, applying the HPO, and conducting a generalization ability test, respectively. Dataset division was performed by interleaved sampling to obtain the representative subsets of sample points. Specifically, all samples were divided into hourly time series based on the time at which the samples were collected. These short time series were subjected to interleaved sampling (or equal-interval sampling) to obtain the validation and test sets, each accounting for 10% of the total samples. The remaining 80% constituted the training set. Table 2 summarizes the number of sample points in each dataset.

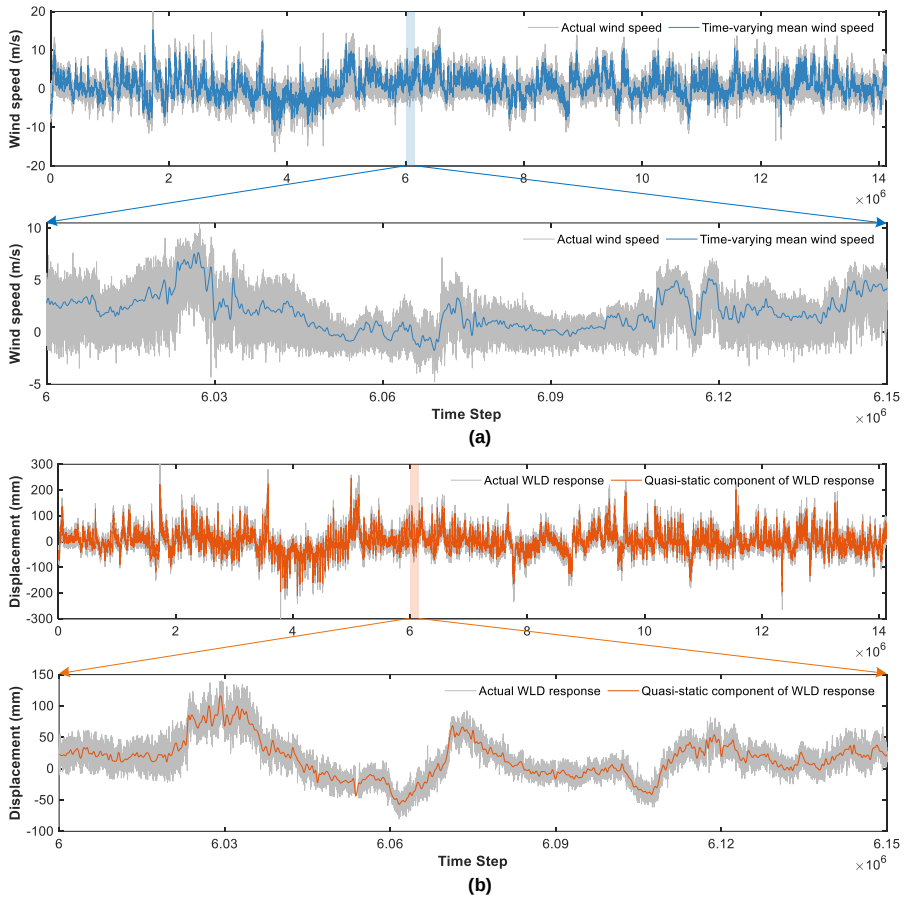

Fig. 11. Decomposition of (a) the wind signals and (b) the WLD signals using the moving average method. The results are shown with all samples.

Table 2. Dataset division and size of each dataset.

Sample dataset		Training set	Validation set	Test set
Proportion		80%	10%	10%
Time length		6278 h	785 h	785 h
Original sample size (SF = 1 Hz)		22,600,800	2,826,000	2,826,000
Sample size after downsampling	Task 1: GRU model (SF = 1/600 Hz)	37,668	4710	4710
	Task 2: U-net model (SF = 0.5 Hz)	11,300,400	1,413,000	1,413,000

Note: SF is the sampling frequency.

3.2.3 Normalization and mini-batch processing

The input and output samples of the models were subjected to Z-score normalization. Namely, data of different scales were normalized to have zero mean and unit variance in order to enhance model training stability and accelerate convergence. Z-score normalization has the following form:

$$\mathbf{Z}' = \frac{\mathbf{Z} - \mu}{\sigma} \tag{21}$$

where $\mathbf{Z}$ and $\mathbf{Z}'$ are samples before and after normalization, respectively; μ is the mean of $\mathbf{Z}$ and σ denotes

the corresponding standard deviation.

During the model training process, a mini-batch of samples was fed into the model in each iteration. As shown in Fig. 12, the initial samples were segmented into equal-length shorter sequences using the sliding window with an equal size as the model input length. The window moves forward 1/*S* of its length each time. The segmented sequences were divided into several mini-batches, and the number of sequences in each mini-batch was defined as mini-batch size *B*. In addition, the model outputs for each mini-batch were assembled by the overlap-and-add (OLA) method to restore the initial length and order. Therefore, the model inference was performed *S* times at each overlap position and the average of the *S* predictions was regarded as the final output. Adopting the above operation not only led to reducing the variance of the prediction but also to increasing its stability.

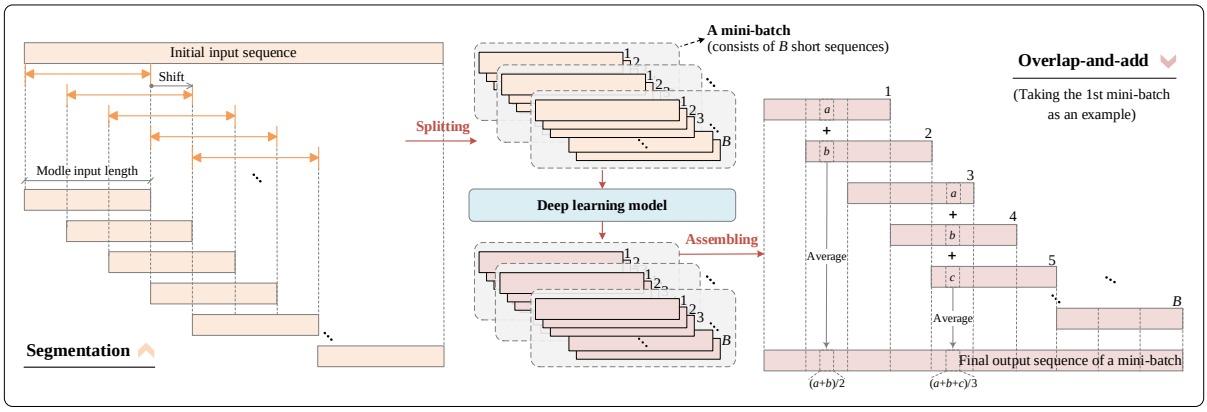

Fig. 12. Schematic diagram of sample segmentation and the overlap-and-add method (the shift ratio of the sliding window is set equal to 1/3 for illustration purposes).

3.3 Task 1: GRU model training and optimization

3.3.1 Training setup

To determine the extreme value weight factor $\mathbf{w}(n)$ in the loss function $\mathcal{L}_{\text{W-MSE}}$, the CDF of the absolute values of the quasi-static component of the WLD, $F(|\mathbf{y}(n)|)$, was established first. Fig. 13 shows the nonparametric probability distribution model built by kernel density estimation using the samples of the training set. The 50-th, 80-th, 95-th, and 99-th percentiles of the absolute quasi-static WLD were 21.8, 46.5, 84.9, and 126.7, respectively.

Table 3 provides the main hyperparameters of the GRU model in Task 1. Some hyperparameters were preset based on the pretests and related experiences to avoid combination explosion caused by a large number of hyperparameters to be optimized. Since the choice of the loss function determines the direction of model

optimization, a two-stage HPO strategy was adopted herein. Specifically, in stage I of the model HPO, the hyperparameters in the loss function $\mathcal{L}_{\text{W-MSE}}$ (i.e., weighting scheme and maximum weight factor a) were optimized. The search ranges are shown in Table 3. During this period, other hyperparameters to be optimized were set to a group of moderate values: the number of GRU layers was m = 2; the number of hidden neurons was h = 200; and the initial learning rate was $l_r = 1\times10^{-4}$. In stage II of the model HPO, the above three hyperparameters were optimized based on the optimal loss function. The grid search method determined the optimal hyperparameter combination in the multi-dimensional hyperparameter space. This optimal hyperparameter combination would result in minimum prediction error on the validation set. The model was trained for 300 epochs for each candidate combination, and the one with the minimum validation error after 250 epochs was saved.

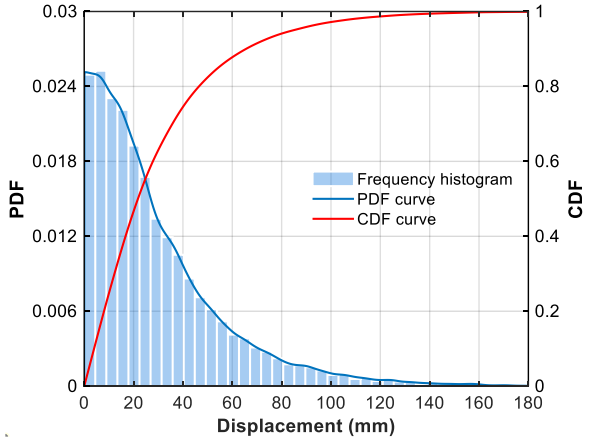

Fig. 13. Probability distribution model of the absolute values of the quasi-static component of the WLD. PDF means the probability density function, and CDF means the cumulative distribution function.

Table 3. Hyperparameters of the GRU model in Task 1.

Hyperparameters		Preset values/Candidate values
Preset ones	Learning rate schedule	Decay by 2% every five epochs
	Shift ratio of the sliding window, 1/S	1/4
	Mini-batch size, B	8
	Model input length	512
	Dropout ratio, p	0.2
	L2 regularization factor	1e-7
To be optimized in stage I	Weighting scheme in $\mathcal{L}_{\text{W-MSE}}$	I, II, III, and IV
	Maximum weight in $\mathcal{L}_{\text{W-MSE}}$, a	5, 10, 20, 50, and 100
To be optimized in stage II	Number of GRU layers, m	1, 2, and 3
	Number of hidden units, h	50, 100, 200, and 300
	Initial learning rate, l_r	1×10^{-3}, 5×10^{-4}, 1×10^{-4}, and 5×10^{-5}

3.3.2 Model HPO and optimal results

Table 4 shows the GRU model's prediction errors (RMSE) on the validation set using different weighting schemes and maximum weight factors $\mathcal{L}_{\text{W-MSE}}$ at stage I of the model HPO. The model prediction errors using the conventional loss function $\mathcal{L}_{\text{MSE}}$ are also provided for comparison. Meanwhile, for each loss function scheme, the prediction errors at four different intervals of $F(|\mathbf{y}(n)|)$, namely, [0, 0.50], (0.50, 0.80], (0.80, 0.95], and (0.95, 1], were calculated to assess the model prediction accuracy at different intervals of the absolute quasi-static WLD. In general, compared with the model validation error using $\mathcal{L}_{\text{MSE}}$, the use of $\mathcal{L}_{\text{W-MSE}}$ reduced the prediction error for the displacement interval corresponding to $0.80 < F(|\mathbf{y}(n)|) \le 1$ (regarded as extreme displacement). In contrast, the prediction error for the displacement interval corresponding to $0 \le F(|\mathbf{y}(n)|) \le 0.80$ increased (regarded as non-extreme displacement). Thus, there is a trade-off between the prediction accuracy for extreme and non-extreme displacements. To assess the performance of each weighting scheme, Table 4 provides the average prediction error (average RMSE) at the four CDF intervals for each scheme. The GRU model had the smallest average RMSE using weighting scheme IV with the maximum weight factor $a = 10$.

To further evaluate the prediction results of quasi-static WLD from a statistical perspective, Fig. 14 shows the frequency histograms of the absolute values of the optimal prediction results using the four weighting schemes. The use of $\mathcal{L}_{\text{W-MSE}}$ resulted in a probability distribution of predicted values that came closer to the true results. Specifically, compared with the predicted values obtained using $\mathcal{L}_{\text{MSE}}$, the probability density of non-extreme displacements decreased, while that of extreme displacements increased. The tail of the probability distribution (i.e., extreme displacements) of the optimal results of weighting scheme IV had the highest degree of agreement with the actual results.

The model HPO in stage II was conducted based on the optimal loss function (weighting scheme IV, $a = 10$). The average RMSE of the prediction results of the above four CDF intervals was still chosen as the performance evaluation metric. Under different hyperparameter combinations constituted by the candidate values of the number of GRU layers m, the number of hidden neurons h, and the initial learning rate l_r, the average RMSE of the GRU model on the validation set is shown in Table 5. The model had the smallest average RMSE with $m = 2$, $h = 200$, and $l_r = 1\times10^{-4}$. Under the above optimal hyperparameter combination, the training and validation loss curves during model training are shown in Fig. 15. The comparison between

the predicted and true values of the quasi-static component of the WLD response on the validation set is shown in Fig. 16. For the sake of comparison, the figure also displays the predicted values using the conventional loss function $\mathcal{L}_{\text{MSE}}$. It is readily seen that the GRU model using $\mathcal{L}_{\text{W-MSE}}$ for training had a higher prediction accuracy for extreme quasi-static WLD.

Table 4. Validation errors of the GRU model using different weighting schemes and maximum weight factors in $\mathcal{L}_{\text{W-MSE}}$ in stage I of model HPO in Task 1.

Weighting scheme	Max. weight factor, *a*	RMSE values for different intervals of $F(\lvert\mathbf{y}(n)\rvert)$				Average RMSE
		[0, 0.50]	(0.5, 0.80]	(0.80, 0.95]	(0.95, 1]	
None	N/A	13.578	22.077	32.222	49.092	29.242
I	5	16.791	22.578	28.531	46.600	28.625
	10	17.321	22.843	28.270	46.015	28.612
	20	**17.758**	**23.020**	**27.772**	**45.713**	**28.566**
	50	17.908	23.153	27.762	45.471	28.574
	100	17.954	23.213	27.850	45.378	28.599
II	5	17.869	23.229	27.767	44.160	28.256
	10	**19.227**	**23.944**	**26.975**	**42.705**	**28.212**
	20	20.105	24.513	26.626	41.818	28.265
	50	20.666	24.946	26.534	41.304	28.362
	100	20.911	25.108	26.466	41.092	28.394
III	**5**	**17.727**	**23.470**	**28.030**	**42.656**	**27.971**
	10	19.968	24.698	27.099	40.142	27.977
	20	21.846	26.025	26.803	38.370	28.261
	50	26.323	28.846	26.985	36.922	29.769
	100	30.541	31.264	26.632	36.539	31.244
IV	5	15.887	23.254	30.045	43.110	28.074
	10	**17.594**	**24.223**	**29.321**	**39.919**	**27.764**
	20	19.642	25.598	29.106	36.985	27.833
	50	23.296	27.981	29.021	34.418	28.679
	100	25.602	29.861	29.615	33.205	29.571

Note: The weighting scheme 'None' uses the conventional loss function $\mathcal{L}_{\text{MSE}}$. $F(\lvert\mathbf{y}(n)\rvert)$ denotes the CDF values of the absolute values of the measured quasi-static WLD. The number in boldface represents the smallest average RMSE value for each weighting scheme.

3.4 Task 2: U-Net model training and optimization

3.4.1 Training setup

Table 6 provides the main hyperparameters of the U-Net model in Task 2. Similar to Task 1, a set of hyperparameters was preset based on the pretests to avoid combination explosion. The model

hyperparameters were also optimized using the two-stage HPO strategy. In stage I, the optimal loss function was chosen between two time-frequency cross-domain loss functions, namely $\mathcal{L}^{\mathrm{TF}}_{\mathrm{MAE\text{-}Mag}}$ and $\mathcal{L}^{\mathrm{TF}}_{\mathrm{RI\text{-}Mag}}$, and the optimal time-frequency combination factor was determined. Meanwhile, the remaining hyperparameters to be trained were set equal to a group of moderate values, that is, the convolution kernel size was $L_{\mathrm{k}} = 11$, and the initial learning rate was $l_r = 1\times10^{-4}$. In stage II, L_{k} and l_r were optimized using the grid search method based on the optimal loss function. The model was trained for 300 epochs for each hyperparameter combination, and the one with the minimum validation error after 200 epochs was saved.

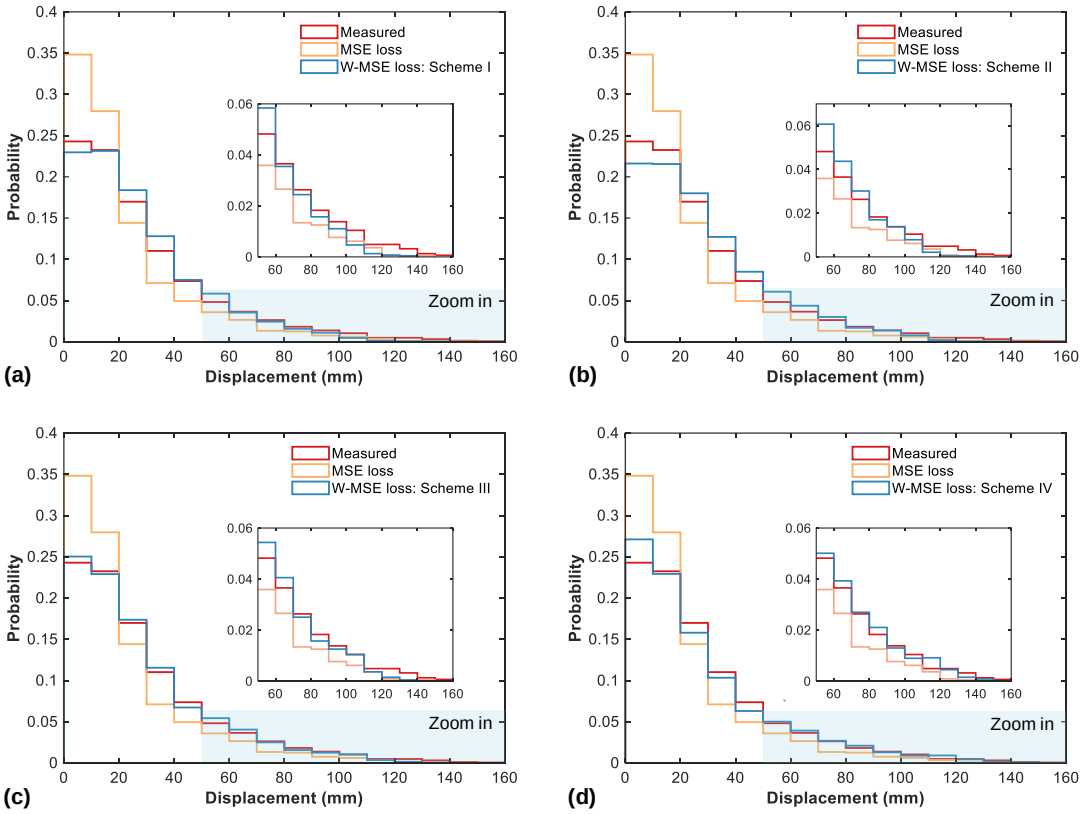

Fig. 14. Frequency histograms of the absolute values of the optimal prediction results using the four weighting schemes: (a) scheme I (a = 20); (b) scheme II (a = 10); (c) scheme III (a = 5); and (d) scheme IV (a = 10).

Table 5. Validation errors (average RMSE) of the GRU model using different hyperparameter combinations in stage II of the model HPO in Task 1.

Initial learning rate, l_r	h = 50			h = 100			h = 200			h = 300		
	m = 1	m = 2	m = 3	m = 1	m = 2	m = 3	m = 1	m = 2	m = 3	m = 1	m = 2	m = 3
1×10^{-3}	28.983	31.986	33.137	32.538	32.693	33.419	33.179	32.595	32.339	32.950	31.429	31.304
5×10^{-4}	28.103	29.659	31.846	29.316	30.997	34.924	29.474	32.785	33.436	30.065	33.138	32.509
1×10^{-4}	28.955	28.281	28.203	28.599	27.841	27.947	28.573	**27.764**	28.838	28.245	27.923	30.376
5×10^{-5}	30.553	29.058	28.780	30.643	28.268	28.088	29.450	27.800	27.812	29.204	27.774	27.766

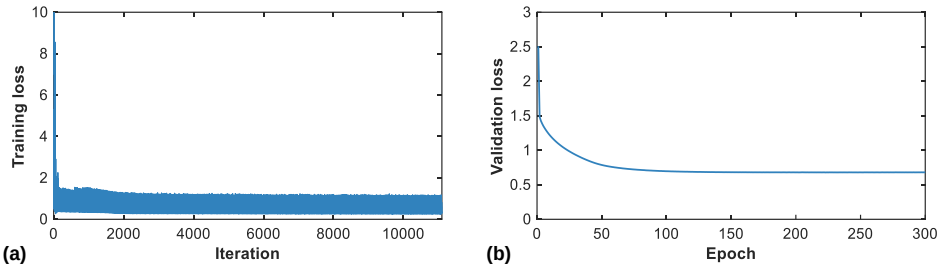

Fig. 15. Loss curves of the GRU model during training under the optimal hyperparameter combination: (a) training loss; (b) validation loss.

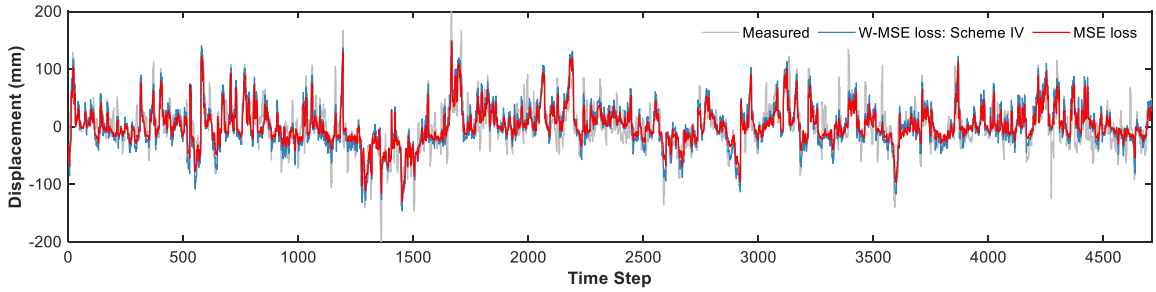

Fig. 16. Comparison between the predicted and true values of the quasi-static WLD using the GRU model under the optimal hyperparameter combination (validation set).

Table 6. Hyperparameters of the U-Net model in Task 2.

Hyperparameters		Preset values/Candidate values
Preset ones	Learning rate schedule	Decay by 2% every five epochs
	Shift ratio of the sliding window, $1/S$	1/2
	Mini-batch size, B	32
	Model input length	32768
	Dropout ratio, p	0.2
	L2 regularization factor	1e-7
	STFT analysis window	Hamming window
	STFT window length	4096
	STFT window overlap length	2048
To be optimized in stage I	Time-frequency combination factor in $\mathcal{L}^{\mathrm{TF}}_{\mathrm{MAE\text{-}Mag}}$ and $\mathcal{L}^{\mathrm{TF}}_{\mathrm{RI\text{-}Mag}}$, α and β	0.01, 0.1, 0.2, 1, 5, 10, 100
To be optimized in stage II	Initial learning rate, l_r	1×10^{-3}, 5×10^{-4}, 1×10^{-4}, and 5×10^{-5}
	Kernel size, L_k	7, 9, 11, and 13

3.4.2 Model HPO and optimal results

Table 7 shows the prediction errors (i.e., RMSE and LSD) of the U-Net model on the validation set using different loss functions in stage I of the model HPO. The prediction errors using the two baseline loss functions $\mathcal{L}^{\mathrm{T}}_{\mathrm{MAE}}$ and $\mathcal{L}^{\mathrm{F}}_{\mathrm{Mag}}$ are also provided for comparison. It is noted that only focusing on time-domain loss, i.e., using $\mathcal{L}^{\mathrm{T}}_{\mathrm{MAE}}$, misled the predictions of the high-frequency dynamic component of the WLD

response. More precisely, although the time-domain error, RMSE, was the smallest at this moment, the error in the frequency domain, LSD, was the largest. This phenomenon was manifested as the low amplitude of the dynamic component of the WLD signal in the time domain. In addition, focusing on frequency-domain loss alone, i.e., using $\mathcal{L}_{\text{Mag}}^{\text{F}}$, resulted in higher prediction accuracy in the frequency domain but also brought the problem of insufficient accuracy in the time domain.

For the two new loss functions $\mathcal{L}_{\text{MAE-Mag}}^{\text{TF}}$ and $\mathcal{L}_{\text{RI-Mag}}^{\text{TF}}$, Table 7 indicates that the time-frequency cross-domain loss considered errors in both time and frequency domains. Overall, as the time-frequency combination factors α and β increased, the RMSE and the LSD decreased and increased, respectively. Therefore, there is also a trade-off between prediction errors in the time and frequency domains. To comprehensively evaluate the performance of different loss functions, Table 7 also provides the sum of the results of the min-max normalization of the RMSE and the LSD for each loss function, i.e., $\text{RMSE}_{\text{Norm}} + \text{LSD}_{\text{Norm}}$, named integrated normalized error here, considering the different dimensions of the two error metrics. It is noted that the model had the smallest integrated normalized error when the phase-constrained cross-domain loss function $\mathcal{L}_{\text{RI-Mag}}^{\text{TF}}$ was used, with $\beta = 0.2$.

Table 7. Validation errors of the U-Net model using different loss functions at stage I of the model HPO in Task 2.

Loss function	Time-frequency combination factor	RMSE ($\text{RMSE}_{\text{Norm}}$)	LSD (LSD_{Norm})	$\text{RMSE}_{\text{Norm}} + \text{LSD}_{\text{Norm}}$
$\mathcal{L}_{\text{MAE}}^{\text{T}}$	N/A	14.3818 (0.000)	3.1139 (1.000)	1.000
$\mathcal{L}_{\text{Mag}}^{\text{F}}$	N/A	20.3680(1.000)	0.8611 (0.016)	1.016
$\mathcal{L}_{\text{MAE-Mag}}^{\text{TF}}$	$\alpha = 0.01$	15.9264 (0.258)	0.8502 (0.011)	0.269
	$\alpha = 0.1$	16.2131 (0.306)	0.8248 (0.000)	0.306
	$\alpha = 0.2$	15.7041 (0.221)	0.8690 (0.019)	0.240
	$\alpha = 1$	15.7642 (0.231)	0.8341 (0.004)	0.235
	$\alpha = 5$	15.0183 (0.106)	1.3399 (0.225)	0.331
	$\alpha = 10$	15.5196 (0.190)	1.1803 (0.155)	0.345
	$\alpha = 100$	15.7330 (0.226)	1.0792 (0.111)	0.337
$\mathcal{L}_{\text{RI-Mag}}^{\text{TF}}$	$\beta = 0.01$	15.7714 (0.232)	0.8520 (0.012)	0.244
	$\beta = 0.1$	15.2865 (0.151)	0.9081 (0.036)	0.187
	$\beta = 0.2$	**15.3573 (0.163)**	**0.8640 (0.017)**	**0.180**
	$\beta = 1$	15.7270 (0.225)	0.9809 (0.068)	0.293
	$\beta = 5$	14.8561 (0.079)	1.4509 (0.274)	0.353
	$\beta = 10$	14.8301 (0.075)	1.7683 (0.412)	0.487
	$\beta = 100$	14.7965 (0.069)	2.6041 (0.777)	0.846

Note: The $\text{RMSE}_{\text{Norm}}$ and LSD_{Norm} in parentheses denote the min-max normalization results of the data in the RMSE and LSD columns, respectively, with the range [0, 1].

The model HPO in stage II was conducted based on the optimal loss function $\mathcal{L}_{\text{RI-Mag}}^{\text{TF}}$ ($\beta = 0.2$). Under different hyperparameter combinations constituted by the candidate values of L_k and l_r, the prediction errors of the U-Net model on the validation set are shown in Table 8, where apart from the RMSE and the LSD, the integrated normalized error is also provided. The model had the smallest integrated normalized error at $L_k = 11$ and $l_r = 1\times10^{-4}$. Under the above optimal hyperparameter combination, the training and validation loss curves during the model training are shown in Fig. 17.

Table 8. Validation errors (RMSE/LSD/$\text{RMSE}_{\text{Norm}}$ + LSD_{Norm}) of the U-Net model using different hyperparameter combinations at stage II of the model HPO in Task 2.

Initial learning rate	$L_k = 7$	$L_k = 9$	$L_k = 11$	$L_k = 13$
1×10^{-3}	15.1099/1.4637/1.000	15.1818/1.0848/0.413	16.0240/0.8549/0.264	16.1420/0.9865/0.505
5×10^{-4}	16.4139/0.8570/0.369	18.9032/0.8530/1.008	16.1590/0.8492/0.290	16.2433/0.8379/0.294
1×10^{-4}	16.3000/0.8488/0.326	16.2287/0.8769/0.353	**15.3573/0.8640/0.106**	16.7481/0.8570/0.456
5×10^{-5}	15.5623/1.0471/0.452	15.8167/0.9339/0.337	18.9637/0.8853/1.076	16.2406/0.9445/0.464

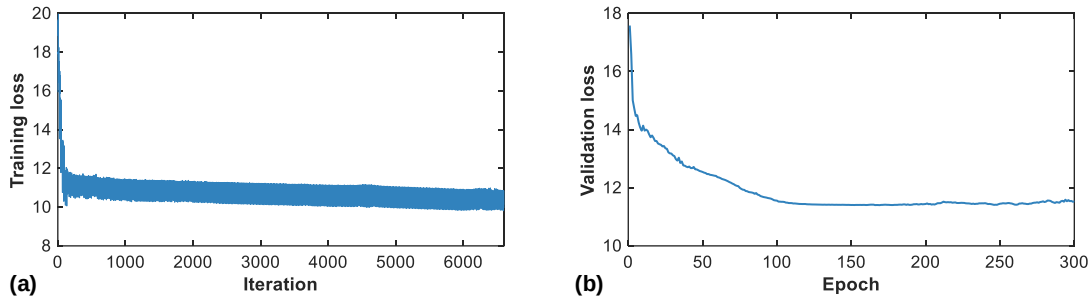

Fig. 17. Loss curves of the U-Net model during training under the optimal hyperparameter combination: (a) training loss; (b) validation loss.

During the first 250 epochs of the model training, the changes in the time- and frequency-domain information of the predicted results with the training epochs are shown in Fig. 18. Note that as the number of epochs increased, the time- and frequency-domain information of the predicted dynamic component became increasingly abundant, and the predicted signal features gradually came closer to their true values. Finally, the model had the smallest validation loss in the 261st training epoch, and thus, the model trained in this epoch was adopted. The comparison between the predicted and measured dynamic components of the WLD in the time domain is shown in Fig. 19, whereas the comparison in the frequency domain is shown in Fig. 20. The optimized U-net model had excellent predictive performance in both time and frequency domains.

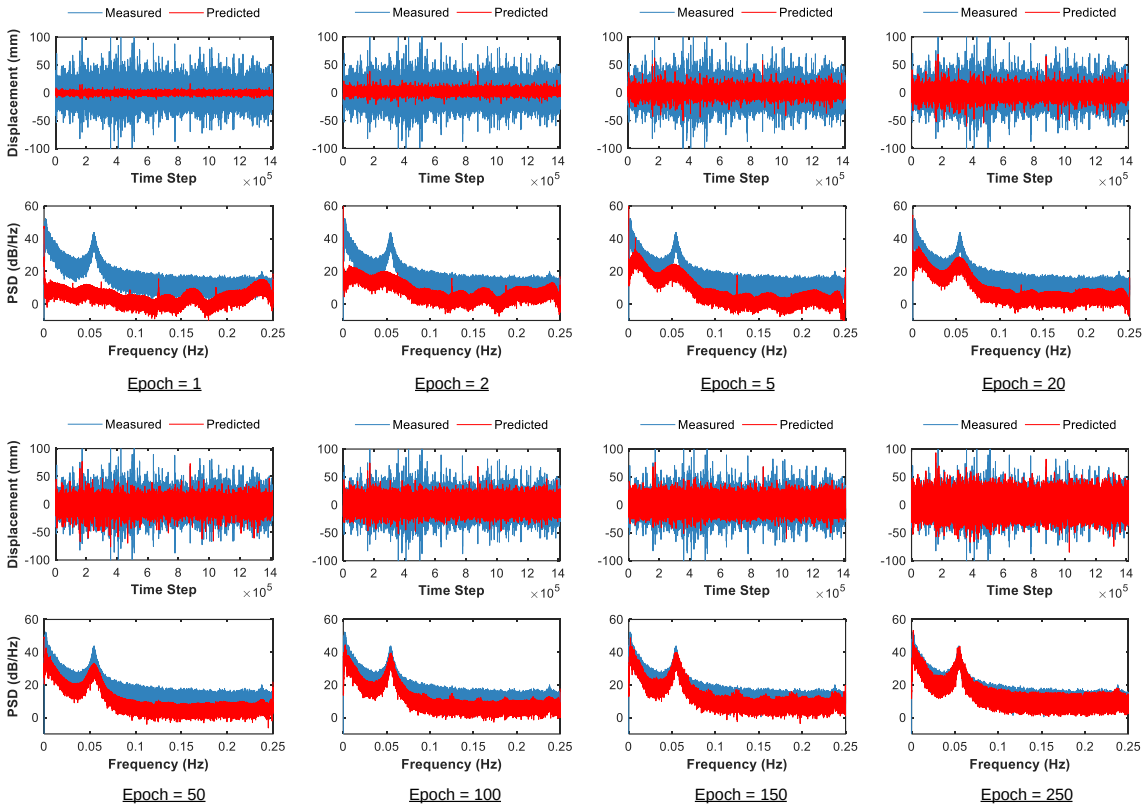

Fig. 18. Changes in the time- and frequency-domain information of the predicted results with the training epochs during the first 250 epochs of training of the U-Net model under the optimal hyperparameter combination (validation set).

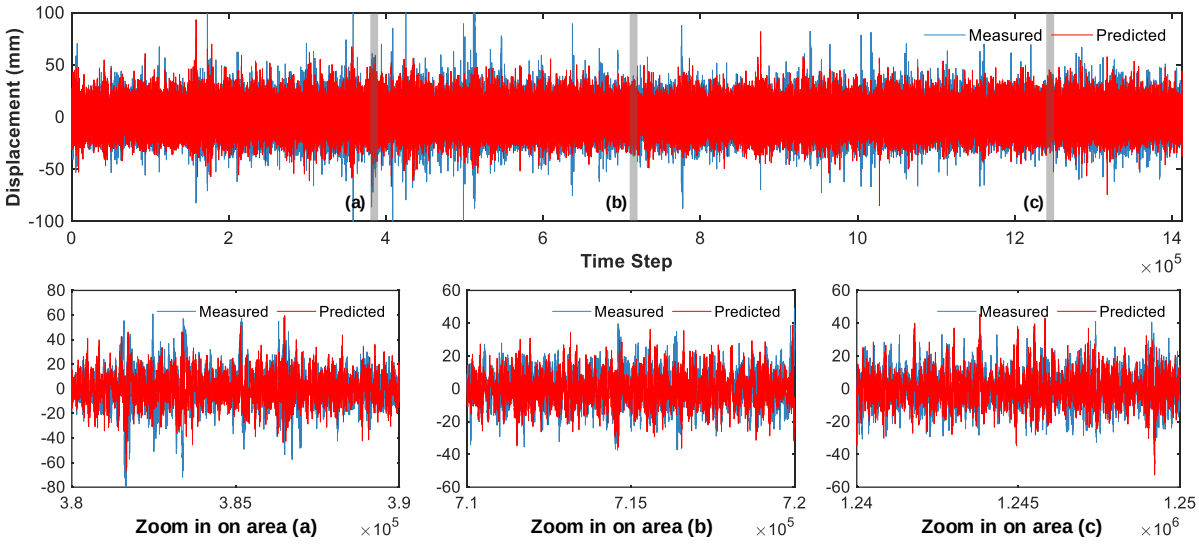

Fig. 19. Time-domain comparison between the predicted and measured dynamic components of the WLD under the optimal hyperparameter combination (validation set).

3.5 Testing the optimized models in Tasks 1 and 2

The GRU model in Task 1 and the U-Net model in Task 2 were trained and optimized independently. Here, the generalization ability and the collaborative performance of the optimized GRU and U-Net models

were assessed using the test set. The outputs of the two models were superimposed to obtain the final prediction of the WLD response. Before the superimposition operation, the output of the GRU model, i.e., the quasi-static WLD, was subjected to cubic spline interpolation so that it has the same time resolution as the dynamic WLD.

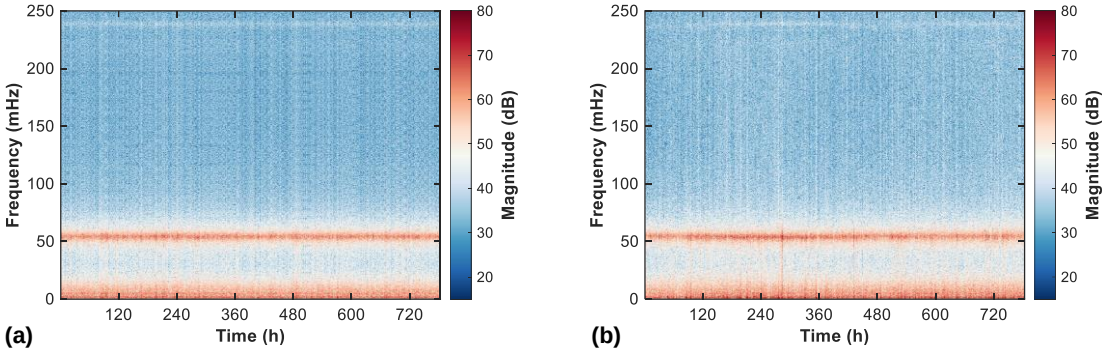

Fig. 20. Frequency-domain comparison between (a) the predicted and (b) measured dynamic components of the WLD under the optimal hyperparameter combination (validation set). It is noted that the time-varying power spectra were used to display the frequency-domain characteristics varying with time.

The comparison between the final predicted values and the measured values of the WLD response on the test set is shown in Fig. 21. The test RMSE between the final predicted values and the true values of the WLD response was 27.3098. More specifically, Fig. 22(a) gives the comparison between the predicted and the true values of the quasi-static component of the WLD response on the test set; Figs. 22(b and c) depict the time- and frequency-domain comparisons between the predicted and the measured values of the dynamic component, respectively. The average RMSE metric of the predicted quasi-static component at the aforementioned four CDF intervals was 27.4667, while the RMSE and LSD metrics of the predicted dynamic component were 15.4525 and 0.8632, respectively. The above results indicated satisfactory collaborative performance between the GRU and U-Net models and demonstrated the generalization ability of the proposed predictive framework for the WLD response.

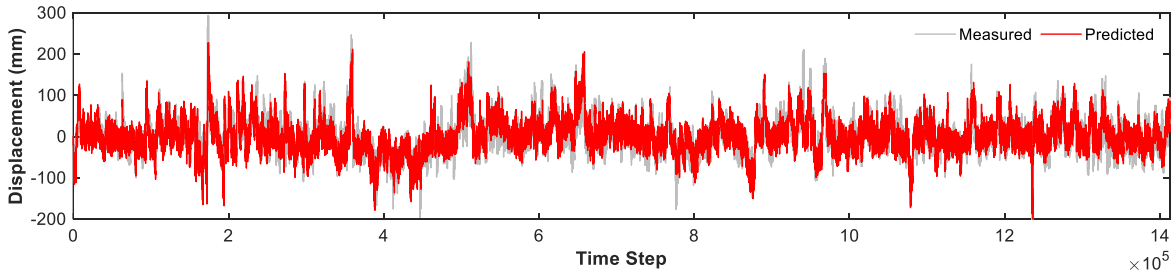

Fig. 21. Comparison between the measured WLD response and the final collaborative predicted results using GRU and U-Net models (test set).

3.6 Discussion

The deck's WLD response at the mid-span of the suspension bridge determines the lateral swing amplitude at the hanger end. The short hanger has a larger hanger-end rotation angle than the long hanger under the same end swing amplitude, and frequent lateral swing of the short hanger will cause fatigue damage to the rigid hanger anchor. Therefore, this section aimed to evaluate the accuracy of the predicted WLD response of the suspension bridge deck from the perspective of fatigue analysis for bridge hangers undergoing lateral swing.

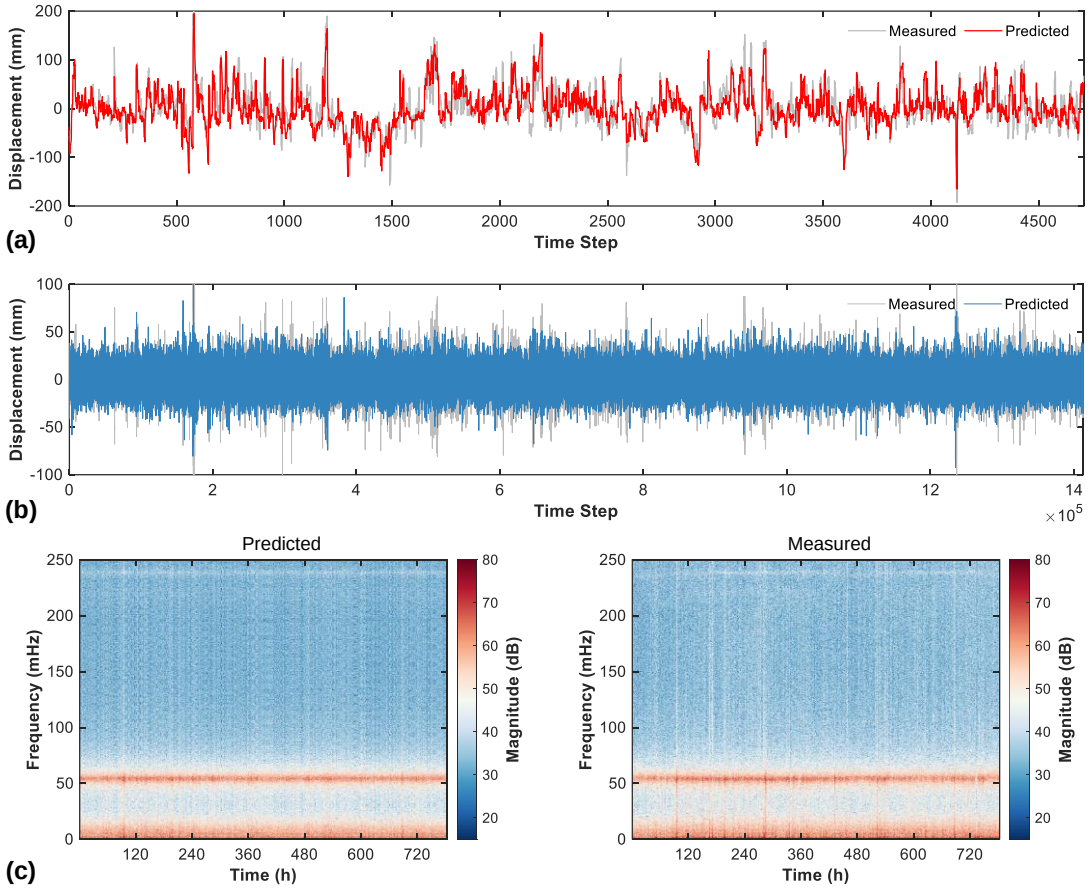

Fig. 22. Comparison between the predicted and true values of the WLD response on the test set: (a) quasi-static component; (b) dynamic component in the time domain; and (c) dynamic component in the frequency domain.

Specifically, the stress level in the short hangers' rigid anchors is positively correlated to the swing amplitude of the bridge deck. Based on this, the concept of the 'swing amplitude-cycle count' histogram of the bridge deck was designed to evaluate the model accuracy by referring to the 'stress range-cycle count' histogram obtained based on the rain flow counting algorithm in fatigue analysis. Fig. 23 compares the 'swing amplitude-cycle count' histograms corresponding to the predicted and measured time histories of the

WLD response using the test set (total duration: 785 h). The swing amplitude histogram of the bridge deck obtained from the predicted WLD response agreed well with the experimental one. Hence, the effectiveness and accuracy of the predictive framework proposed in this study were verified. At the same time, the feasibility of using the predicted WLD response for fatigue analysis of bridge hangers was proven.

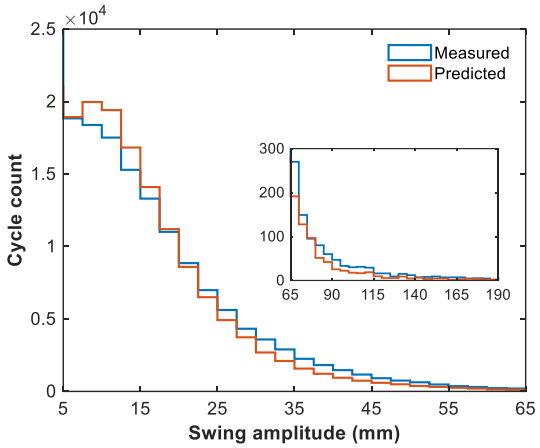

Fig. 23. Comparison of the 'swing amplitude-cycle count' histograms corresponding to the predicted and measured time histories of the WLD response (test set). The predicted and measured values of the cycle counts for the minor swings with amplitudes below 5 mm are 100,070 and 111,939, respectively; these are not shown in the figure since they are significantly greater than other values.

4. Summary and conclusions

In this study, a deep learning-based framework was proposed for predicting the wind-induced lateral displacement (WLD) response of the suspension bridge deck for structural health monitoring (SHM). The proposed framework consisted of two deep-learning tasks. In Task 1, a GRU model was built and used for predicting the quasi-static component of the WLD response of the bridge deck based on the lateral time-varying 10-min mean wind speed. In Task 2, a U-Net model was built and used for predicting the dynamic component of the WLD response based on the actual lateral wind speed. Finally, the proposed method was verified by a case study of the Jiangyin Yangtze River Bridge in China. The following conclusions were drawn:

(1) The WLD response of the suspension bridge deck can be decomposed into quasi-static and dynamic components, for which the corresponding prediction models can be established using the GRU model and U-Net model, respectively.

(2) Based on the spectral analysis of the raw WLD records, applying appropriate downsampling to the sample data can help remove redundancy and reduce the burden of model training.

(3) In Task 1, increasing the weight of the extreme displacement prediction error in the loss function can improve the prediction accuracy for the extreme quasi-static WLD response. The weight factor for the extreme value error can be determined based on the CDF of the absolute value of the WLD. The model performed optimally when using the inverse proportional weighting scheme (i.e., scheme IV) in the extreme value-weighted loss function. However, there was a trade-off between the prediction accuracy for extreme and non-extreme displacements.

(4) In Task 2, using the time-frequency cross-domain loss function can ensure both the time- and frequency-domain prediction accuracy for the dynamic component of the WLD. Compared with the two baseline loss functions $\mathcal{L}_{\mathrm{MAE}}^{\mathrm{T}}$ and $\mathcal{L}_{\mathrm{Mag}}^{\mathrm{F}}$ related only to time- or frequency-domain errors, using the phase-constrained cross-domain loss function $\mathcal{L}_{\mathrm{RI\text{-}Mag}}^{\mathrm{TF}}$ provided a better comprehensive predictive performance in both time and frequency domains. A trade-off was also identified between the prediction errors of the dynamic component of the WLD in the time and frequency domains.

(5) Leveraging the proposed 'swing amplitude-cycle count' histogram of the suspension bridge deck obtained by the rain flow counting method, the WLD response prediction accuracy can be evaluated from the perspective of fatigue analysis for bridge hangers undergoing lateral swing.

While the proposed framework has demonstrated effectiveness on the selected SHM dataset of the Jiangyin Yangtze River Bridge, it is crucial to note the limitations involving the model's generalization and transferability. For Task 1, challenges may arise for the current model to achieve the same level of prediction accuracy for rare extreme events wherein the mean wind speeds significantly exceed the maximum values in the selected training set. Future work will attempt to address this issue by increasing the sample scale to encompass more extreme events. Additionally, the physics-enhanced method, which integrates physical principles into neural networks, will be explored, and it is expected to exhibit robust generalization ability facing rare extreme wind speeds. Furthermore, the direct transferability of the optimized models here to other bridges may be challenging, as they were specifically trained as surrogates for the case bridge using bridge-specific data. However, the proposed framework offers a valuable reference for other bridges to build their WLD response prediction models with their own data.

CRediT authorship contribution statement

Zhi-wei Wang: Conceptualization, Methodology, Software, Formal analysis, Visualization, Writing - Original Draft. **Xiao-fan Lu:** Methodology, Software, Formal analysis. **Wen-ming Zhang:** Supervision, Validation, Writing - Review & Editing, Funding acquisition. **Vasileios C. Fragkoulis:** Writing – review & editing. **Yu-feng Zhang:** Resources. **Michael Beer:** Supervision, Validation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The authors gratefully acknowledge the support by the National Key R&D Program of China (No. 2022YFB3706703), the National Natural Science Foundation of China under Grant 52078134, the Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX21_0118), and the Scientific Research Foundation of Graduate School of Southeast University (YBPY2129).

References

Brownjohn, J.M.W., Au, S.K., Zhu, Y.C., Sun, Z., Li, B.B., Bassitt, J., . . . Sun, H.B., 2018. Bayesian operational modal analysis of Jiangyin Yangtze River Bridge. Mech. Syst. Signal Process. 110, 210-230.

Castellon, D.F., Fenerci, A., Øiseth, O., 2021. A comparative study of wind-induced dynamic response models of long-span bridges using artificial neural networks, support vector regression and buffeting theory. J. Wind Eng. Ind. Aerodyn. 209, 104484.

Cheng, J., Xiao, R.C., 2006. A simplified method for lateral response analysis of suspension bridges under wind loads. Commun. Numer. Methods Eng. 22 (8), 861-874.

Chung, J., Gulcehre, C., Cho, K., Bengio, Y., 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555.

Cooray, K., Ananda, M.M.A., 2008. A generalization of the half-normal distribution with applications to lifetime data. Commun. Stat. Theory Methods 37 (9), 1323-1337.

Cuomo, S., Di Cola, V.S., Giampaolo, F., Rozza, G., Raissi, M., Piccialli, F., 2022. Scientific machine learning through physics–informed neural networks: Where we are and what's next. J. Sci. Comput. 92 (3), 88.

Deng, Y., Ju, H.W., Zhai, W.Q., Li, A.Q., Ding, Y.L., 2022. Correlation model of deflection, vehicle load, and

698 temperature for in-service bridge using deep learning and structural health monitoring. Struct. Control Health
699 Monit. 29 (12), e3113.
700 Faroughi, S.A., Pawar, N.M., Fernandes, C., Raissi, M., Das, S., Kalantari, N.K., Mahjour, S.K., 2024. Physics-
701 guided, physics-informed, and physics-encoded neural networks and operators in scientific computing: Fluid
702 and solid mechanics. J. Comput. Inf. Sci. Eng. doi: https://doi.org/10.1115/1.4064449.
703 Fenerci, A., Lystad, T.M., Øiseth, O., 2023. Full-scale monitored wind and response characteristics of a suspension
704 bridge compared with wind tunnel investigations at the design stage. J. Wind Eng. Ind. Aerodyn. 242, 105583.
705 Fenerci, A., Øiseth, O., 2018. Strong wind characteristics and dynamic response of a long-span suspension bridge
706 during a storm. J. Wind Eng. Ind. Aerodyn. 172, 116-138.
707 Jiang, F., Ding, Y., Song, Y., Geng, F., Wang, Z., 2022. Automatic pixel-level detection and measurement of
708 corrosion-related damages in dim steel box girders using Fusion-Attention-U-net. J. Civ. Struct. Health Monit.
709 13 (1), 199-217.
710 Jiang, K.J., Han, Q., Du, X.L., Ni, P.H., 2021. Structural dynamic response reconstruction and virtual sensing
711 using a sequence to sequence modeling with attention mechanism. Autom. Constr. 131, 103895.
712 Jones, M.C., 1993. Simple boundary correction for kernel density-estimation. Stat. Comput. 3 (3), 135-146.
713 Lei, X.M., Siringoringo, D.M., Sun, Z., Fujino, Y., 2022. Displacement response estimation of a cable-stayed
714 bridge subjected to various loading conditions with one-dimensional residual convolutional autoencoder
715 method. Struct. Health Monit. 22 (3), 1790-1806.
716 Li, X., Zhang, W., 2022. Physics-informed deep learning model in wind turbine response prediction. Renew.
717 Energy 185, 932-944.
718 Liu, Z.X., Guo, T., Huang, L.Y., Pan, Z.H., 2017. Fatigue life evaluation on short suspenders of long-span
719 suspension bridge with central clamps. J. Bridge Eng. 22 (10), 04017074.
720 Minh, N.N., Miyata, T., Yamada, H., Sanada, Y., 1999. Numerical simulation of wind turbulence and buffeting
721 analysis of long-span bridges. J. Wind Eng. Ind. Aerodyn. 83 (1), 301-315.
722 Ni, P., Li, Y.X., Sun, L.M., Wang, A., 2022. Traffic-induced bridge displacement reconstruction using a physics-
723 informed convolutional neural network. Comput. Struct. 271, 106863.
724 Oh, B.K., Kim, J., 2021. Optimal architecture of a convolutional neural network to estimate structural responses
725 for safety evaluation of the structures. Measurement 177, 109313.
726 Rahman, A., Wu, Z.Y., Kalfarisi, R., 2021. Semantic deep learning integrated with RGB feature-based rule
727 optimization for facility surface corrosion detection and evaluation. J. Comput. Civ. Eng. 35 (6), 04021018.
728 Ribeiro, R.P., Moniz, N., 2020. Imbalanced regression and extreme value prediction. Mach. Learn. 109, 1803-
729 1835.
730 Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: convolutional networks for biomedical image segmentation.
731 In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (Eds.), Medical Image Computing and Computer-Assisted
732 Intervention – MICCAI 2015, Part III, Lecture Notes in Computer Science. Springer, Switzerland, pp. 234-
733 241.
734 Scheepens, D.R., Schicker, I., Hlavackova-Schindler, K., Plant, C., 2023. Adapting a deep convolutional RNN
735 model with imbalanced regression loss for improved spatio-temporal forecasting of extreme wind speed events
736 in the short to medium range. Geosci. Model Dev. 16 (1), 251-270.
737 Tang, H., Kang, J., Li, Y., 2023. Aerostatic and aerodynamic stability of a suspension bridge during early erection
738 stages. J. Wind Eng. Ind. Aerodyn. 233, 105297.
739 Tang, Z.Y., Bao, Y.Q., Li, H., 2021. Group sparsity-aware convolutional neural network for continuous missing
740 data recovery of structural health monitoring. Struct. Health Monit. 20 (4), 1738-1759.
741 Tao, T.Y., Xu, Y.L., Huang, Z.F., Zhan, S., Wang, H., 2020. Buffeting analysis of long-span bridges under typhoon

742 winds with time-varying spectra and coherence. J. Struct. Eng. 146 (12), 04020255.
743 Vincent, G.S., 1958. Golden Gate Bridge vibration studies. J. Struct. Div. 84 (6), 1817-1-1817-40.
744 Wang, C.W., Ansari, F., Wu, B., Li, S.J., Morgese, M., Zhou, J.T., 2022a. LSTM approach for condition assessment
745 of suspension bridges based on time-series deflection and temperature data. Adv. Struct. Eng. 25 (16), 3450-
746 3463.
747 Wang, G.X., Ding, Y.L., 2014. Mathematical modeling for lateral displacement induced by wind velocity using
748 monitoring data obtained from main girder of Sutong Cable-Stayed Bridge. Math. Probl. Eng. 2014, 723152.
749 Wang, Y., Yang, D.H., Yi, T.H., 2021a. Accurate correlation modeling between wind speed and bridge girder
750 displacement based on a multi-rate fusion method. Sensors (Basel) 21 (6), 1967.
751 Wang, Y.W., Ni, Y.Q., Zhang, Q.H., Zhang, C., 2021b. Bayesian approaches for evaluating wind-resistant
752 performance of long-span bridges using structural health monitoring data. Struct. Control Health Monit. 28
753 (4), e2699.
754 Wang, Z.W., Li, A.D., Zhang, W.M., Zhang, Y.F., 2022b. Long-term missing wind data recovery using free access
755 databases and deep learning for bridge health monitoring. J. Wind Eng. Ind. Aerodyn. 230, 105201.
756 Wang, Z.W., Lu, X.F., Zhang, W.M., Fragkoulis, V.C., Beer, M., Zhang, Y.F., 2023. Deep learning-based
757 reconstruction of missing long-term girder-end displacement data for suspension bridge health monitoring.
758 Comput. Struct. 284, 107070.
759 Xu, F.Y., Chen, A.R., Zhang, Z., 2013. Aerostatic wind effects on the Sutong Bridge. In: Proceedings of the Third
760 International Conference on Intelligent System Design and Engineering Applications (ISDEA 2013), 247-256.
761 IEEE, Hong Kong, China.
762 Xu, X., Xu, D.H., Caballero, A., Ren, Y., Huang, Q., Chang, W.J., Forde, M.C., 2023a. Vehicle-induced deflection
763 prediction using long short-term memory networks. Structures 54, 596-606.
764 Xu, Z., Wang, H., Xing, C., Tao, T., Mao, J., Liu, Y., 2023b. Physics guided wavelet convolutional neural network
765 for wind-induced vibration modeling with application to structural dynamic reliability analysis. Eng. Struct.
766 297, 117027.
767 Yang, Y., Zhang, J., Cao, F., Ge, Y., Zhao, L., 2022. Evaluation and improvement of wind environment and vehicle
768 safety on long-span bridge deck under strong crosswind. J. Wind Eng. Ind. Aerodyn. 228, 105089.
769 Yuan, P., Adhikari, S., Dong, Y., 2023. A partitioned combined computational method for multi-scale dynamic
770 systems. Int. J. Numer. Methods Eng. 124 (16), 3494-3523.
771 Zhang, J., Zhu, C., Ma, C., 2024. Driving safety analysis of wind–vehicle–bridge system considering aerodynamic
772 interference. J. Wind Eng. Ind. Aerodyn. 245, 105649.
773 Zhang, R., Liu, Y., Sun, H., 2020. Physics-informed multi-LSTM networks for metamodeling of nonlinear
774 structures. Comput. Methods Appl. Mech. Eng. 369, 113226.
775 Zhang, W.M., Chen, J., Tian, G.M., Chang, J.Q., 2022. Static response assessment of the entire suspension bridge
776 under horizontal transverse live load: an analytical calculation method. J. Bridge Eng. 27 (7), 04022052.
777 Zhu, J., Xiong, Z., Xiang, H., Huang, X., Li, Y., 2021. Ride comfort evaluation of stochastic traffic flow crossing
778 long-span suspension bridge experiencing vortex-induced vibration. J. Wind Eng. Ind. Aerodyn. 219, 104794.