Online meta-learning approach for sensor fault diagnosis using limited data

Online Meta-Learning for Sensor Fault Diagnosis Using Limited Data

Lei Wang[1], Dukang Huang[1, 2], Ke Huang[1]*, Marco Civera[2]

[1] School of Civil Engineering, Changsha University of Science and Technology, Changsha 410114, China
[2] Department of Structural, Building and Geotechnical Engineering, Politecnico di Torino, Turin 10129, Italy)

*Corresponding author: kehuang@csust.edu.cn

ABSTRACT

The accurate and timely diagnosis of sensor faults plays a critical role in ensuring the reliability and performance of structural health monitoring systems. However, the challenge is detecting, locating, and estimating sensor faults in an online manner using limited training data. To resolve this problem, a novel approach for online sensor fault diagnosis is proposed for structural health monitoring. The proposed approach is based on meta-learning, which enables superior model generalization capabilities using limited data. The detection, localization, and estimation of typical sensor faults in an online manner can be achieved efficiently by the proposed approach. First, a one-dimensional convolutional neural network (1-D CNN) is designed to detect and locate faulty sensors. The initial model parameters of the 1-D CNN are optimized using a model-agnostic meta-learning training strategy. This strategy allows the acquisition of transferable prior knowledge, which can speed up the learning process on new sensor fault detection and localization tasks. The meta-learning strategy also enables efficient and accurate detection and localization of potential faulty sensors with limited data. After detecting and locating the faulty sensors, an online updating algorithm based on a dual Kalman filter is used to estimate the severity of sensor faults and structural states simultaneously. The proposed approach is demonstrated with simulated sensor faults that cover a numerical example and field measurements from the Canton Tower. The results show that the proposed approach is applicable for online sensor fault diagnosis in structural health monitoring.

Keywords:

Structural health monitoring, meta-learning, online estimation, fault diagnosis, Kalman filter

1 Introduction

Civil infrastructures are susceptible to aging and deterioration caused by severe environmental conditions and excessive loads, posing a significant threat to public safety [1-2]. Structural health monitoring (SHM) has emerged as a crucial solution to mitigate these challenges and plays a critical role in evaluating structural integrity [3-6]. SHM encompasses the monitoring of variations in structural materials and geometric properties by utilizing measurements from sensors installed on structures.

However, the sensors mounted on structures are susceptible to malfunction due to adverse environmental conditions and aging. A sensor is considered faulty when it exhibits an unacceptable deviation between its readings and the true values of the observed quantities [7]. A faulty sensor can lead to false alarms or missed detections regarding the structural health condition [8, 9]. Therefore, it is essential to ensure the reliability of the sensor system before evaluating the structural health. A typical sensor fault diagnosis method comprises three essential components, namely, sensor fault detection, localization, and estimation [10]. Specifically, sensor fault detection aims to determine the presence of malfunctioning sensors within the sensor system. Sensor fault localization refers to identifying the locations of malfunctioning sensors. Sensor fault estimation is focused on quantifying the severity of the faults, which is an essential component for calibrating the distorted sensing data.

Typically, sensor fault detection methods are categorized into three primary classes, statistical-based methods [11-13], residual-based methods [14-16], and classification-based methods [17-19]. Statistical-based methods establish sensor fault detection indices and derive their control limits from data obtained under fault-free conditions. When the indices of the current sensing data exceed the prescribed limits, a sensor fault alarm should be triggered. However, the effectiveness of the statistical-based methods heavily relies on the suitable definition of control limits, often requiring subjective manual adjustments. Residual-based methods generate estimated outputs by establishing a predictive model, and then residuals are calculated by comparing the actual outputs with the estimated ones. Significant deviations of residuals from zero serve as an indicator of sensor faults within the network. Classification-based methods treat sensor fault detection as a classification problem. The measurements are transformed into low-dimensional feature vectors, and then the feature vectors are fed into a classifier for sensor fault detection.

Several researchers have developed sensor fault detection approaches based on machine learning, such as support vector machines [17], artificial neural networks [18], and extreme learning machines [19]. Nonetheless, in these machine learning methods, effort is required to find the optimal combination of feature extraction and classification to ensure the accuracy of sensor fault detection. Deep learning has been extensively adopted for its powerful nonlinear feature extraction and integration capabilities. In recent decades, one-dimensional convolutional neural networks (1-D CNNs) have achieved state-of-the-art performance in SHM applications [20, 21]. However, achieving the desired performance of these deep neural networks relies heavily on training with extensive labeled data. Therefore, it often occurs that 1-D CNNs with deep architectures may not achieve satisfactory results in applications with limited labeled data [22]. On the other hand, meta-learning has emerged as a popular research topic in recent years, spanning image classification [23], video object detection [24], and fault diagnosis [25]. Meta-learning is a methodology that leverages prior learning knowledge to expedite the learning process for new tasks. In contrast to traditional deep learning methods designed for specific tasks, meta-learning focuses on a set of similar tasks to extract hidden features, which enhances its knowledge extraction capabilities. By training on a set of relevant learning tasks, transferable prior knowledge can be acquired. Based on the transferable prior knowledge, meta-learning can achieve efficient learning with limited data in new learning tasks. Model-agnostic meta-learning (MAML) [26] is one of the most representative optimization-based meta-learning methods. The core idea of MAML is to obtain the optimal initial model parameters for a deep neural network by using gradient descent during the meta-training phase. Then the deep neural network achieves its optimal performance by updating these initial model parameters when confronted with new learning tasks featuring limited data.

If the sensor system is found to be faulty, it becomes imperative to perform fault localization to pinpoint the specific faulty sensor(s). Conventional methods, including missing variable methods [27], contribution analysis methods [28], and probability quantification methods [29] have been proposed to locate the faulty sensor(s). However, missing variable methods often lead to a serious combination explosion issue when dealing with a large number of sensors in the network.

Furthermore, contribution analysis methods and probability quantification methods perform effectively under the assumption that, at most, a single sensor within the sensor system experiences a malfunction.

After detecting and locating the specific faulty sensors, it is necessary to quantify the severity of the fault for correcting the distorted sensing data. The most commonly employed methods for this purpose are regression-based approaches [30, 31]. These approaches quantify sensor faults by estimating the extent of corruption in sensing data from pristine fault-free observations.

Despite extensive research efforts dedicated to sensor fault diagnosis methods, critical challenges remain unresolved. Existing sensor fault diagnosis methods based on deep learning require a substantial amount of labeled sensor failure data for training a satisfactory classification model. However, such failure data are inherently challenging to obtain due to their scarcity and difficulty, particularly in comparison to the more readily available normal data [32]. Moreover, it is time-consuming to label samples with sensor faults, particularly in cases where a substantial number of data points require manual annotations. This process demands a tremendous allocation of human resources. On the other hand, the majority of existing sensor fault diagnosis methods are developed by analyzing historical offline data. However, these approaches encounter challenges in the context of online sensor fault diagnosis applications.

To resolve the above problems, we propose an online sensor fault diagnosis method based on meta-learning. First, a 1-D CNN is designed to detect and locate potential faulty sensors. By optimizing the initial model parameters of the 1-D CNN using the MAML training strategy, prior knowledge can be acquired to expedite the learning process on new sensor fault detection and localization tasks based on a limited number of samples. This results in efficient and accurate detection and localization of potential faulty sensors. After locating all possible faulty sensors, an online updating algorithm based on a dual Kalman filter (KF) is used to simultaneously update both the severity of sensor faults and the structural states.

The main innovations in this study are summarized as follows: (1) the proposed approach avoids using massive data points to train the model for sensor fault detection and localization, and it only requires a limited number of samples to obtain accurate and reliable estimation results; (2) this novel method is established based on a meta-learning strategy that enables superior model generalization capabilities and adapts to changing environmental conditions; and (3) the proposed method provides efficient detection, localization, and estimation of typical sensor faults in an online manner, which can minimize downtime by allowing for rapid response and preventive maintenance.

The remainder of this paper is organized as follows. In Section 2, the formulation of the sensor fault diagnosis problem is introduced. In Section 3, an online sensor fault diagnosis approach is proposed. First, a 1-D CNN based on meta-learning is designed to detect and locate faulty sensors online. Then, an online updating algorithm based on a dual KF is used to simultaneously estimate the severity of sensor faults and structural states. In Section 4, the proposed method is validated using a bridge example. In Section 5, the proposed approach is demonstrated by integrating field measurements and simulations from the Canton Tower. Finally, Section 6 gives the conclusion of the paper.

2 Problem Formulation

In this section, we present the sensor fault diagnosis problem for fault-corrupted sensor networks based on time-invariant systems.

Consider a dynamical system with N_d degrees-of-freedom (DOFs), and it is formulated by the following equation of motion:

$$\mathbf{M}\ddot{\mathbf{u}}(t) + \mathbf{L}\dot{\mathbf{u}}(t) + \mathbf{K}\mathbf{u}(t) = \mathbf{T}\mathbf{f}(t) \tag{1}$$

where $\mathbf{M} \in \mathbb{R}^{N_d \times N_d}$, $\mathbf{L} \in \mathbb{R}^{N_d \times N_d}$, $\mathbf{K} \in \mathbb{R}^{N_d \times N_d}$ are the mass, damping, and stiffness matrix of the dynamical system, respectively; $\mathbf{u}(t) \in \mathbb{R}^{N_d}$, $\dot{\mathbf{u}}(t) \in \mathbb{R}^{N_d}$, and $\ddot{\mathbf{u}}(t) \in \mathbb{R}^{N_d}$ are the displacement, velocity, and acceleration vector at time t, respectively; $\mathbf{f}(t) \in \mathbb{R}^{N_f}$ is the unmeasured excitation applied to the system at time t, and $\mathbf{f}$ is modeled as a Gaussian process with zero mean and covariance matrix $\mathbf{Q}^f \in \mathbb{R}^{N_f \times N_f}$; $\mathbf{T} \in \mathbb{R}^{N_d \times N_f}$ is the excitation distributing matrix.

Eq. (1) can be converted to the state-space form as follows:

$$\dot{\mathbf{y}}(t) = \mathbf{A}_r\mathbf{y}(t) + \mathbf{B}_r\mathbf{f}(t) \tag{2}$$

where $\mathbf{y}(t) = [\mathbf{u}(t)^{\mathrm{T}}, \ \dot{\mathbf{u}}(t)^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{2N_d}$ is the state vector which includes the structural displacement and velocity; the state transitional matrix $\mathbf{A}_r$ and input distributing matrix $\mathbf{B}_r$ are given by:

$$\mathbf{A}_r = \begin{bmatrix} \mathbf{0}_{N_d \times N_d} & \mathbf{I}_{N_d} \\ -\mathbf{M}^{-1}\mathbf{K} & -\mathbf{M}^{-1}\mathbf{L} \end{bmatrix}$$

$$\mathbf{B}_r = \begin{bmatrix} \mathbf{0}_{N_d \times N_f} \\ \mathbf{M}^{-1}\mathbf{T} \end{bmatrix} \tag{3}$$

where $\mathbf{0}_{N_d \times N_d}$ and $\mathbf{I}_{N_d}$ are the $N_d \times N_d$ zero matrix and N_d identity matrix, respectively.

Then, Eq. (2) can be discretized using a sampling time interval Δt:

$$\mathbf{y}_k = \overline{\mathbf{A}}_r \mathbf{y}_{k-1} + \overline{\mathbf{B}}_r \mathbf{f}_{k-1}, \quad k = 1,2, \ldots, \tag{4}$$

where $\mathbf{y}_k \equiv \mathbf{y}(k\Delta t)$, $\overline{\mathbf{A}}_r \equiv \exp(\mathbf{A}_r \Delta t)$, $\overline{\mathbf{B}}_r \equiv {\mathbf{A}_r}^{-1}(\overline{\mathbf{A}}_r - \mathbf{I}_{2N_d})\mathbf{B}_r$, $\mathbf{f}_{k-1} \equiv \mathbf{f}[(k-1)\Delta t]$.

Assuming that structural responses are measured by N_o sensors with sampling time interval Δt, the regular noise-corrupted measurements $\mathbf{z}_k \in \mathbb{R}^{N_o}$ at the k^{th} time step are given by:

$$\mathbf{z}_k = \mathbf{C}_d \mathbf{y}_k + \mathbf{n}_k \tag{5}$$

where $\mathbf{z}_k \equiv \mathbf{z}(k\Delta t)$; $\mathbf{C}_d \in \mathbb{R}^{N_o \times 2N_d}$ is the observation matrix; $\mathbf{n}_k$ is the measurement noise at the k^{th} time step, and $\mathbf{n}$ is modeled as a zero-mean Gaussian stochastic process with covariance $\mathbf{R} \in \mathbb{R}^{N_o \times N_o}$.

It is unavoidable for sensors in the network to be faulty due to harsh environments and aging. A sensor fault is considered an unacceptable deviation between the sensor readings and the actual values of observed quantities 7. The fault-corrupted measurements $\tilde{\mathbf{z}}_k \in \mathbb{R}^{N_o}$ at the k^{th} time step are given by:

$$\tilde{\mathbf{z}}_k = \mathbf{z}_k + \mathbf{H}\boldsymbol{\xi}_k \tag{6}$$

where $\mathbf{H} \equiv [\mathbf{H}^{(1)}, \mathbf{H}^{(2)}, \ldots, \mathbf{H}^{(N_p)}] \in \mathbb{R}^{N_o \times N_p}$ is the localization matrix of the N_p faulty sensor(s) and $\mathbf{H}^{(i)} \in \mathbb{R}^{N_o}$, $i = 1,2, \ldots, N_p$ indicates the localization vector of the i^{th} faulty sensor. In particular, $\mathbf{H}^{(i)}$, $i = 1,2, \ldots, N_p$ consists of one unity element and $(N_o - 1)$ zero elements. The position of the unity element in $\mathbf{H}^{(i)}$ is associated with the i^{th} faulty sensor. $\boldsymbol{\xi}_k \equiv$

$\left[\xi_k^{(1)}, \xi_k^{(2)}, \dots, \xi_k^{(N_p)}\right] \in \mathbb{R}^{N_p}$ is the sensor fault vector at the k^{th} time step, and $\xi_k^{(i)}$, $i = 1,2,\dots,N_p$ denotes the sensor fault value of the i^{th} faulty sensor at the k^{th} time step.

The unknown sensor fault $\boldsymbol{\xi}_k$ can be modeled as a random walk process to indicate its variation [33]:

$$\boldsymbol{\xi}_k = \boldsymbol{\xi}_{k-1} + \mathbf{v}_k \tag{7}$$

where $\mathbf{v}_k$ is the process noise related to the random walk model, and $\mathbf{v}$ is assumed to be an N_p-variate Gaussian vector with zero mean and covariance matrix $\mathbf{Q}^{\mathrm{v}} \in \mathbb{R}^{N_p \times N_p}$.

The goal of sensor fault diagnosis is to detect, locate, and quantify the severity of the sensor faults in the sensor network. Conventional methods for the detection and localization of faulty sensors involve training a classification model based on large amounts of labeled data from historical offline sources. As a result, these methods are not applicable to online sensor fault diagnosis. To resolve this problem, a novel online sensor fault diagnosis method based on meta-learning is proposed for SHM. The localization matrix of faulty sensors can be obtained by using a 1-D CNN based on meta-learning. The proposed method only requires a limited number of samples to train the 1-D CNN model for sensor fault detection and localization. After obtaining the localization matrix of faulty sensors, the proposed method can quantify the unknown sensor faults and can update the structural states vector simultaneously online.

3 Online Sensor Fault Diagnosis

In this section, an online sensor fault diagnosis approach is presented. First, a 1-D CNN is used to build the model for sensor fault detection and localization, and the initial model parameters are optimized by using model-agnostic meta-learning training strategy. Then, based on the trained model, faulty sensors can be detected and located online using a moving time window strategy. Finally, after determining the specific locations of the faulty sensors, the sensor faults can be reconstructed, and the structural states can be estimated simultaneously in an online manner by using a dual Kalman filter approach.

3.1 Online Sensor Fault Detection and Localization

3.1.1 Architecture of sensor fault detection and localization model

The proposed method for sensor fault detection and localization requires training a 1-D CNN for all sensors in the sensor network. A typical 1-D CNN architecture is shown in Figure 1. A 1-D CNN typically contains one feature extraction stage and one classification stage.

In the feature extraction stage, the input goes through convolutional layer, batch normalization (BN) layer, and max-pooling layer in sequence for feature extraction. Convolutional layer extracts features from input data by convolving the input local regions with convolutional kernels. The parameters of the convolution operation comprise kernel size and stride. The kernel size refers to the length of the convolutional filter. The stride refers to the step size at which the filter moves along the input data during convolution. Before performing the convolution operation, a spatial dimension control technique padding can be used to preserve the spatial dimension of the output. It involves adding additional zero elements around the input. There are two types of padding, namely, valid padding and same padding. The valid padding means no zero elements are added to the input. The same padding refers to zero elements are added symmetrically to the input. Each neuron in convolutional layer is connected to a local region of neurons in the previous layer. This connection enables the neuron to learn and capture local hidden features of the data. The BN layer is specifically designed to expedite the training process of deep neural networks by minimizing internal covariance shift [34]. The BN layer follows the convolutional layer. Pooling operation is performed after convolutional layer to reduce the spatial size of features and the number of parameters for the network. The max-pooling layer conducts local maximum operation over the input features. Activation functions play a crucial role in enabling the network to acquire a nonlinear expression ability. Rectified Linear Unit (ReLU) is chosen as the activation function for convolutional layer. This is because ReLU efficiently captures the non-linearity effect to accelerate the convergence of the CNN.

Several fully-connected layers compose a multi-layer perceptron in the classification stage. The softmax function is used to transform the original output vector of the multi-layer perceptron into a probability distribution. It produces a set of probabilities with a unity sum. Output value of the softmax function can be interpreted as the probability that the input belongs to a particular class.

In this study, we slightly modify the 1-D CNN model proposed in [35, 36] to detect and locate the faulty sensors in the sensor system. The input of the sensor fault detection and localization model is the observed data. The feature extraction stage contains several convolutional modules. Each module is composed of a convolutional layer, BN layer, max-pooling layer, and ReLU nonlinearity in series. By stacking more convolutional modules, the network becomes deeper. This deeper network helps to acquire suitable representations of the observed data [35]. However, a deeper network will also lead to a longer training duration. The number of convolutional modules is set to five by striking a balance between feature extraction and training efficiency. The classification stage is composed of a fully-connected layer with two neurons. Then, the softmax function is used to transform the output of the two neurons into the faulty probability of the sensor. As a result, the main differences between the original 1-D CNN model and the proposed one rely on the number of data points in the input and the number of artificial neurons in the fully-connected layer. Hyperparameters of the 1-D CNN include the number of data points for the input, the kernel number, and the kernel size in convolutional layers. The hyperparameters are case-dependent, and they rely on specific datasets. To find the optimal hyperparameters, the random search hyperparameter optimization method [37] is used to conduct a random search on the parameter space. Specifically, the hyperparameter optimization process is composed of a series of iterations based on the training dataset, which is divided into three sets with one for training and the remaining two for cross-validation. In each iteration, the random search optimization method randomly selects a configuration of values for hyperparameters from a specified set of possible values. The 1-D CNN model with the selected hyperparameters is then trained based on the training dataset. The average value of the classification accuracy by using the cross-validation strategy is then served as average validation classification accuracy. The classification accuracy refers to the measure of how correctly a classification model predicts the class labels of the input data. It is calculated by dividing the number of correctly classified instances by the total number of instances in the dataset. The hyperparameter configuration providing the highest average validation classification accuracy is selected as the optimal hyperparameter for the model. A predefined threshold of 90% is set in the optimization process for acceptable accuracy. When the highest accuracy is lower than the predefined threshold, it is suggested to improve the maximum number of the hyperparameter optimization iterations. The hyperparameters of the 1-D CNN used in this study are listed in Table 1.

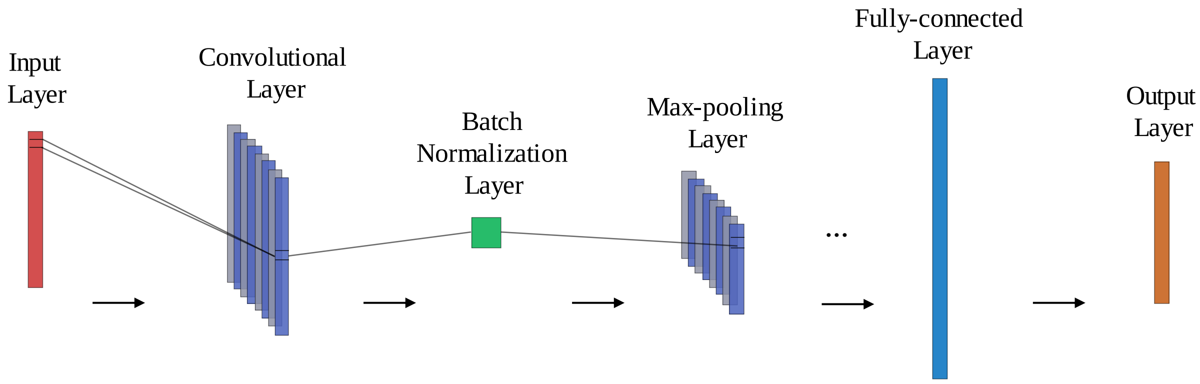

Figure 1. Architecture of a typical 1-D CNN

Table 1 Hyperparameters of the 1-D CNN

No.	Layer	Kernel number	Kernel size	Stride	Activation	Padding
1	Input	1	1000	-	-	-
2	Convolution1	16	(64,1)	(16,1)	-	Same
3	BN1	-	-	-	-	-
4	Max-pooling1	16	(2,1)	(2,1)	ReLU	Valid
5	Convolution2	32	(3,3)	(1,1)	-	Same
6	BN2	-	-	-	-	-
7	Max-pooling2	32	(2,1)	(2,1)	ReLU	Valid
8	Convolution3	64	(3,3)	(1,1)	-	Same
9	BN3	-	-	-	-	-
10	Max-pooling3	64	(2,1)	(2,1)	ReLU	Valid
11	Convolution4	64	(3,3)	(1,1)	-	Same
12	BN4	-	-	-	-	-
13	Max-pooling4	64	(2,1)	(2,1)	ReLU	Valid
14	Convolution5	64	(3,3)	(1,1)	-	Same
15	BN5	-	-	-	-	-
16	Max-pooling5	64	(2,1)	(2,1)	ReLU	Valid
17	Fully-connected	1	2	-	Softmax	-
18	Output	1	2	-	-	-

3.1.2 MAML training strategy

Training a reliable sensor fault detection and localization model with limited samples can be considered as a few-shot learning problem [38]. To achieve satisfactory performance using limited data, 1-D CNN based on meta-learning is proposed to online detect and locate faulty sensors. By optimizing the initial model parameters of the 1-D CNN with the MAML training strategy, transferable prior knowledge can be acquired to expedite the learning process on new sensor fault detection and localization tasks based on a limited number of samples. The training strategy of MAML is based on these tasks. In this study, the sensor fault detection and localization task is defined as a 2-way, $N_{\mathcal{S}}$-shot, $N_{\mathcal{Q}}$-query classification problem. The term 2-way indicates that each task contains 2 types of samples, i.e., regular and faulty. $N_{\mathcal{S}}$ and $N_{\mathcal{Q}}$ represent the number of samples of each type for training and testing in a task, respectively. $\mathcal{T}^{tr}$ denotes a sensor fault detection and localization task from training dataset $\mathcal{D}_{\text{train}}$, and it consists of one support set $\mathcal{S}^{tr}$ and one query set $\mathcal{Q}^{tr}$. Both $\mathcal{S}^{tr}$ and $\mathcal{Q}^{tr}$ contain regular and faulty samples. In particular, $\mathcal{S}^{tr}$ and $\mathcal{Q}^{tr}$ contain $N_{\mathcal{S}}$ and $N_{\mathcal{Q}}$ samples per class, respectively. Samples from $\mathcal{S}^{tr}$ and $\mathcal{Q}^{tr}$ are denoted as $\left(x_i^{tr,\mathcal{S}}, e_i^{tr,\mathcal{S}}\right), i = 1,2, \dots, 2N_{\mathcal{S}}$ and $\left(x_i^{tr,\mathcal{Q}}, e_i^{tr,\mathcal{Q}}\right), i = 1,2, \dots, 2N_{\mathcal{Q}}$, respectively. For the sensor fault detection and localization task, x denotes observations and e $(e = 0,1)$ denotes the label of x. Specifically, the regular and faulty observations are labeled as 0 and 1, respectively.

The meta-learning based CNN is denoted as meta-model. The meta-model is represented by a parametrized function $F_{\boldsymbol{\theta}}$ with parameters $\boldsymbol{\theta}$. The meta-model maps inputs x to outputs $F_{\boldsymbol{\theta}}(x)$. According to the MAML training strategy, sensor fault detection and localization tasks $\mathcal{T}_j^{tr}, j = 1,2, \dots, N_{ba}$, are generated by random sampling from $\mathcal{D}_{\text{train}}$, and N_{ba} denotes the number of tasks. Then, based on all $\mathcal{S}^{tr}$ and $\mathcal{Q}^{tr}$ in the N_{ba} sensor fault detection and localization tasks, the inner and outer loop optimizations are implemented, respectively. The optimization of the inner loop refers to sub-model updates of each task $\mathcal{T}_j^{tr}, j = 1,2, \dots, N_{ba}$. The purpose of the inner loop optimization is to obtain the updated parameters of the sub-model. For each task $\mathcal{T}_j^{tr}$, the initial parameters $\boldsymbol{\theta}_j$ of the sub-model will be updated by gradient descent algorithm [26]:

$$\boldsymbol{\theta}_j^{(l+1)} = \boldsymbol{\theta}_j^{(l)} - \alpha \nabla J_{\mathcal{S}^{tr}}\left(\boldsymbol{\theta}_j^{(l)}\right), j = 1,2, \dots, N_{ba} \tag{8}$$

where $\boldsymbol{\theta}_j^{(l)}$, $l = 1,2, \dots, N_{up}$, is the sub-model parameters of $\mathcal{T}_j^{tr}$ after the l^{th} updating, and N_{up} denotes the number of gradient updates; α is the inner loop learning rate; $\nabla J_{\mathcal{S}^{tr}}\left(\boldsymbol{\theta}_j^{(l)}\right)$ is the gradient of the objective function $J_{\mathcal{S}^{tr}}\left(\boldsymbol{\theta}_j^{(l)}\right)$ for the inner loop optimization based on $\mathcal{S}^{tr}$ and it is given as follows:

$$J_{\mathcal{S}^{tr}}\left(\boldsymbol{\theta}_j^{(l)}\right) = \sum_{i=1}^{2N_{\mathcal{S}}}\{e_{i,j}^{tr,\mathcal{S}}\ln F_{\boldsymbol{\theta}_j^{(l)}}(x_{i,j}^{tr,\mathcal{S}}) \\ +(1 - e_{i,j}^{tr,\mathcal{S}})\ln\,[1 - F_{\boldsymbol{\theta}_j^{(l)}}(x_{i,j}^{tr,\mathcal{S}})]\} \\ /(2N_{\mathcal{S}}) \tag{9}$$

Moreover, it requires only a few gradient updates to obtain the optimal initial parameters of the sub-model [26].

The outer loop optimization refers to the meta-model updates. The purpose of the outer loop is to optimize the initial parameters of the meta-model. Using all $\mathcal{Q}^{tr}$ in the tasks $\mathcal{T}_j^{tr}, j = 1,2,\dots,N_{ba}$, the initial parameters $\boldsymbol{\theta}$ of the meta-model are updated by:

$$\boldsymbol{\theta}^* = \boldsymbol{\theta} - \beta\sum_{j=1}^{N_{ba}}\nabla J_{\mathcal{Q}^{tr}}\left(\boldsymbol{\theta}_j^{(l+1)}\right) \tag{10}$$

where $\boldsymbol{\theta}^*$ is the optimal initial parameters of the meta-model; β is the outer loop learning rate; $\nabla J_{\mathcal{Q}^{tr}}\left(\boldsymbol{\theta}_j^{(l+1)}\right)$ is the gradient of the objective function $J_{\mathcal{Q}^{tr}}\left(\boldsymbol{\theta}_j^{(l+1)}\right)$ for the outer loop optimization based on $\mathcal{Q}^{tr}$, and it is defined as follows:

$$J_{\mathcal{Q}^{tr}}\left(\boldsymbol{\theta}_j^{(l+1)}\right) = \sum_{i=1}^{2N_{\mathcal{Q}}}\{e_{i,j}^{tr,\mathcal{Q}}\ln F_{\boldsymbol{\theta}_j^{(l+1)}}(x_{i,j}^{tr,\mathcal{Q}}) \\ +(1 - e_{i,j}^{tr,\mathcal{Q}})\ln\,[1 - F_{\boldsymbol{\theta}_j^{(l+1)}}(x_{i,j}^{tr,\mathcal{Q}})]\} \\ /(2N_{\mathcal{Q}}) \tag{11}$$

In general, MAML trains the model parameters for fast learning with few gradient updates [26].

The training process of the MAML is shown in Figure 2.

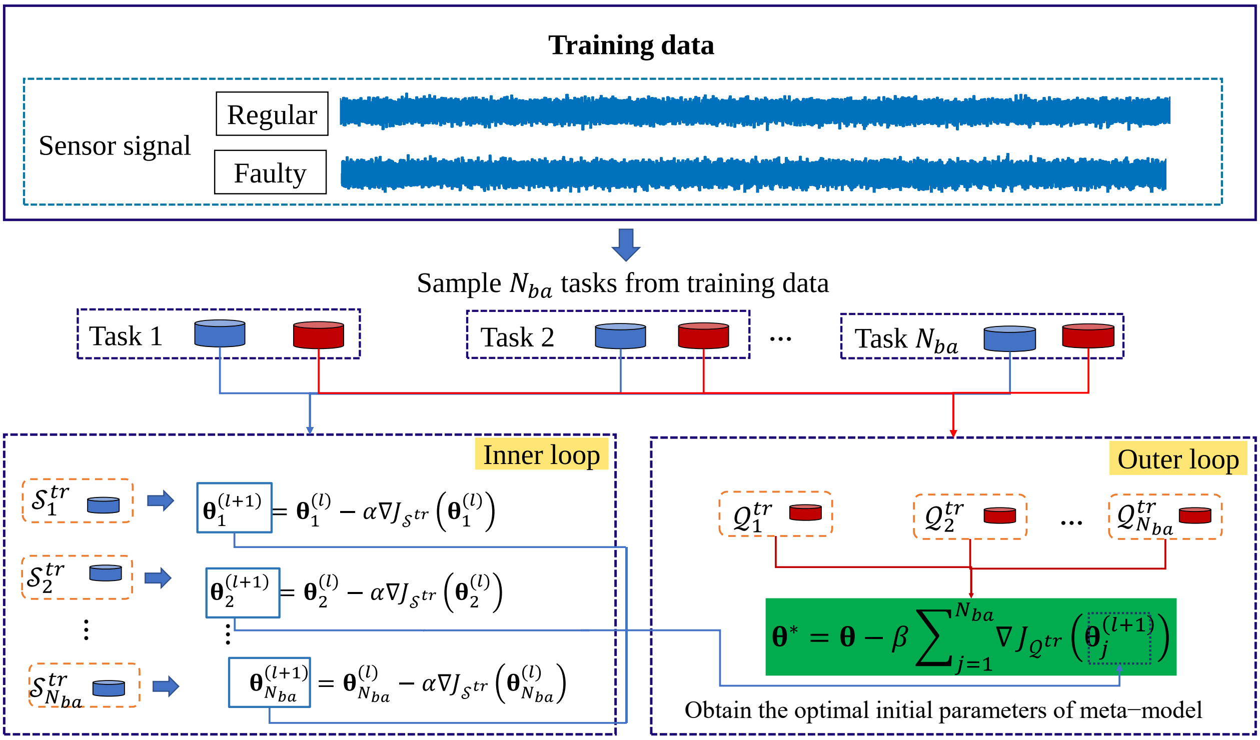

Figure 2. Training process of the MAML

3.1.3. Online testing

After the MAML training is completed, the optimal initial parameters $\boldsymbol{\theta}^*$ of the meta-model are used to expedite the learning process on new sensor fault detection and localization tasks. $\mathcal{T}^{te}$ denotes a sensor fault detection and localization task from testing dataset $\mathcal{D}_{\text{test}}$. $\mathcal{T}^{te}$ consists of one support set $\mathcal{S}^{te}$ and one query set $\mathcal{Q}^{te}$. $\mathcal{S}^{te}$ contains $N_{\mathcal{S}}$ samples per class, and the samples from $\mathcal{S}^{te}$ are denoted as $\left(x_i^{te,\mathcal{S}}, e_i^{te,\mathcal{S}}\right), i = 1, \ldots, 2N_{\mathcal{S}}$. For the few-shot problem, $N_{\mathcal{S}}$ is a small positive integer. Similarly, $\mathcal{Q}^{te}$ contains $N_{\mathcal{Q}}$ samples per class, and the samples from $\mathcal{Q}^{te}$ are denoted as $\left(x_i^{te,\mathcal{Q}}, e_i^{te,\mathcal{Q}}\right), i = 1,2,\ldots,2N_{\mathcal{Q}}$. The goal of the new sensor fault detection and localization task $\mathcal{T}^{te}$ is to identify the label of $\mathcal{Q}^{te}$ using a limited number of samples in $\mathcal{S}^{te}$. The optimal initial parameters $\boldsymbol{\theta}^*$ of meta-model are fine-tuned by gradient descent algorithm [26]:

$$\boldsymbol{\theta}^{*(l+1)} = \boldsymbol{\theta}^{*(l)} - \alpha\nabla J\left(\boldsymbol{\theta}^{*(l)}\right) \tag{12}$$

where $\boldsymbol{\theta}^{*(l)}, l = 1,2, \ldots, N_{up}$, is the parameters of the meta-model after the l^{th} updating; $\nabla J\left(\boldsymbol{\theta}^{*(l)}\right)$

is the gradient of the objective function $J(\boldsymbol{\theta}^{*(l)})$ for $\mathcal{T}^{te}$, and it is given by:

$$J(\boldsymbol{\theta}^{*(l)}) = \sum_{i=1}^{2N_{\mathcal{S}}}\{e_i^{te,\mathcal{S}}\ln F_{\boldsymbol{\theta}^{*(l)}}(x_i^{te,\mathcal{S}}) \\ +(1-e_i^{te,\mathcal{S}})\ln[1-F_{\boldsymbol{\theta}^{*(l)}}(x_i^{te,\mathcal{S}})]\} \\ /(2N_{\mathcal{S}}) \tag{13}$$

The fine-tuning process in Eq. (12) can be interpreted as searching for the optimal parameters that are the most suitable for $\mathcal{D}_{\text{test}}$ based on the prior knowledge of $\mathcal{D}_{\text{train}}$. Then, the fine-tuned meta-model $F_{\boldsymbol{\theta}^{*(l+1)}}$ is used to predict the label of $\mathcal{Q}^{te}$. The testing classification accuracy ACC is calculated by checking the consistency between the actual labels and predicted labels of $\mathcal{Q}^{te}$, and it is given by:

$$ACC = \sum_{i=1}^{2N_{\mathcal{Q}}} f_B\left(F_{\boldsymbol{\theta}^{*(l+1)}}(x_i^{te,\mathcal{Q}}),\quad e_i^{te,\mathcal{Q}}\right) \\ /(2N_{\mathcal{Q}}) \tag{14}$$

where $f_B(\cdot)$ is the Boolean function, it outputs 1 when the predicted label $F_{\boldsymbol{\theta}^{*(l+1)}}(x_i^{te,\mathcal{Q}})$ is equal to the actual label $e_i^{te,\mathcal{Q}}$, otherwise, it outputs 0. The fine-tuned meta-model $F_{\boldsymbol{\theta}^{*(l+1)}}$ with the highest testing classification accuracy is recognized as the best meta-model when training iteration reaches to the maximum number of training iterations $\overline{N}_{train}$. During training, an early stopping strategy [39] is adopted to avoid overfitting. The best meta-model is then used to online detect and locate the faulty sensors in the sensor system.

The online nature of the proposed method for sensor fault detection and localization implementation is interpreted as follows. Owing to the fast-learning ability of the MAML, the MAML provides a computationally efficient alternative to detect and locate sensor faults in an online manner. The entire monitoring duration can be discretized as a series of continuous small-time windows, and each time window contains N_{win} measured data points for all N_o measurement channels. A single round of sensor fault detection and localization implementation is described as follows. For each sensor, a time window is used to aggregate N_{win} data points. Then, the best meta-model is used to classify the N_{win} aggregated data points in each time window. If the output label remains 1 (i.e., faulty) for N_{con} ($N_{con} \geq 2$) consecutive time windows, the location of the specific faulty sensor can be determined. N_{con} is suggested to approximately cover a fundamental period of the monitored structure to obtain sufficient information, and it can be determined using a simple Fast Fourier Transform analysis of raw measurements. After one round of the implementation is completed, the meta-model is updated by using a limited number of samples. And possible faulty sensors could be detected and located online by the updated meta-model in the next round of implementation. The process of online testing is shown in Figure 3.

The proposed method leverages the fast-learning capability of model-agnostic meta-learning to acquire prior knowledge. This allows for efficient updating of the sensor fault detection and localization model with limited data. Although the proposed method still needs the training phase, only a limited number of samples is required in the online testing phase owing to the fast-learning capability of model-agnostic meta-learning, which is computationally very efficient for achieving online detection and localization. As a result, the proposed method is capable of detecting and locating sensor faults in an online manner.

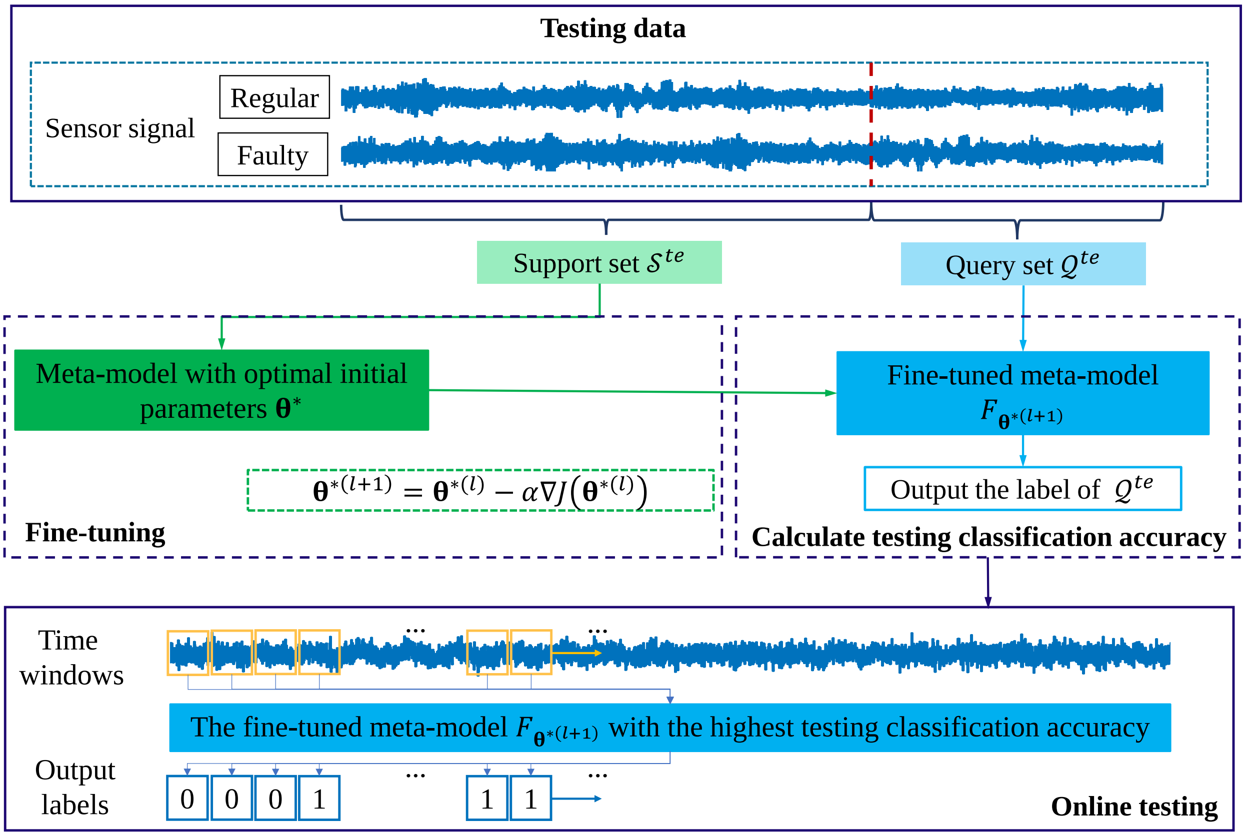

Figure 3. Online testing

3.2 Online Sensor Fault Quantification

After locating the possible faulty sensors in the sensor system, a dual KF is proposed to update the unknown sensor faults and structural states simultaneously in an online manner. Owning to the online nature of the proposed method, sensor fault types can be identified as well by analyzing the variations in the estimated sensor faults over time.

Given the measurement set $\mathcal{D}_{k-1} \equiv \{\tilde{\mathbf{z}}_1, \tilde{\mathbf{z}}_2, \dots, \tilde{\mathbf{z}}_{k-1}\}$, the one-step-ahead predicted state vector, and its associated covariance matrix can be calculated by taking the conditional expected value based on Eq. (4):

$$\mathbf{y}_{k|k-1} \equiv E[\mathbf{y}_k | \mathcal{D}_{k-1}] = \overline{\mathbf{A}}_r \mathbf{y}_{k-1|k-1} \tag{15}$$

$$\begin{aligned} \mathbf{P}^{\mathbf{y}}_{k|k-1} &\equiv E\left[\left(\mathbf{y}_k - \mathbf{y}_{k|k-1}\right)\left(\mathbf{y}_k - \mathbf{y}_{k|k-1}\right)^{\mathrm{T}} | \mathcal{D}_{k-1}\right] \\ &= \overline{\mathbf{A}}_r \mathbf{P}^{\mathbf{y}}_{k-1|k-1} \overline{\mathbf{A}}_r^{\mathrm{T}} + \overline{\mathbf{B}}_r \mathbf{Q}^f \overline{\mathbf{B}}_r^{\mathrm{T}} \end{aligned} \tag{16}$$

where the notation $|k-1$ denotes the conditional expectation on all the past data points up to the $(k-1)^{th}$ time step, and it will be used hereafter. On the other hand, the one-step-ahead predicted sensor fault vector and its associated covariance matrix can be calculated by taking the conditional expected value based on Eq. (7):

$$\boldsymbol{\xi}_{k|k-1} \equiv E[\boldsymbol{\xi}_k | \mathcal{D}_{k-1}] = \boldsymbol{\xi}_{k-1|k-1} \tag{17}$$

$$\begin{aligned} \mathbf{P}^{\boldsymbol{\xi}}_{k|k-1} &\equiv E\left[\left(\boldsymbol{\xi}_k - \boldsymbol{\xi}_{k|k-1}\right)\left(\boldsymbol{\xi}_k - \boldsymbol{\xi}_{k|k-1}\right)^{\mathrm{T}} | \mathcal{D}_{k-1}\right] \\ &= \mathbf{P}^{\boldsymbol{\xi}}_{k-1|k-1} + \mathbf{Q}^{\mathbf{v}} \end{aligned} \tag{18}$$

Given a new measurement $\tilde{\mathbf{z}}_k$ at the k^{th} time step, the structural state vector, and its covariance matrix can be updated by using the KF:

$$\mathbf{y}_{k|k} = \mathbf{y}_{k|k-1} + \mathbf{G}^{\mathbf{y}}_k\left(\tilde{\mathbf{z}}_k - \mathbf{C}_d \mathbf{y}_{k|k-1} - \mathbf{H}\boldsymbol{\xi}_{k|k}\right) \tag{19}$$

$$\mathbf{P}^{\mathbf{y}}_{k|k} = \mathbf{P}^{\mathbf{y}}_{k|k-1} - \mathbf{G}^{\mathbf{y}}_k \mathbf{C}_d \mathbf{P}^{\mathbf{y}}_{k|k-1} \tag{20}$$

where $\mathbf{G}^{\mathbf{y}}_k$ is the Kalman gain of the k^{th} time step for the state vector, and it is given by:

$$\mathbf{G}_k^{\mathbf{y}} = \mathbf{P}_{k|k-1}^{\mathbf{y}} \mathbf{C}_d^{\mathrm{T}} \left(\mathbf{C}_d \mathbf{P}_{k|k-1}^{\mathbf{y}} \mathbf{C}_d^{\mathrm{T}} + \mathbf{R} \right)^{-1} \tag{21}$$

It is implied in Eq. (19) that the sensor fault vector is assumed to be known, and it is equal to the value at the current time step, when calculating the updated structural state vector. The sensor fault vector $\boldsymbol{\xi}_{k|k}$ in Eq. (19) can also be obtained by using KF:

$$\boldsymbol{\xi}_{k|k} = \boldsymbol{\xi}_{k|k-1} + \mathbf{G}_k^{\boldsymbol{\xi}} \left(\tilde{\mathbf{z}}_k - \mathbf{C}_d \boldsymbol{y}_{k-1|k-1} - \mathbf{H} \boldsymbol{\xi}_{k|k-1} \right) \tag{22}$$

In addition, the covariance matrix of the updated sensor fault vector can be obtained as follows:

$$\mathbf{P}_{k|k}^{\boldsymbol{\xi}} = \mathbf{P}_{k|k-1}^{\boldsymbol{\xi}} - \mathbf{G}_k^{\boldsymbol{\xi}} \mathbf{H} \mathbf{P}_{k|k-1}^{\boldsymbol{\xi}} \tag{23}$$

where $\mathbf{G}_k^{\boldsymbol{\xi}}$ is the Kalman gain of the k^{th} time step for the sensor fault vector, and it is given by:

$$\mathbf{G}_k^{\boldsymbol{\xi}} = \mathbf{P}_{k|k-1}^{\boldsymbol{\xi}} \mathbf{H}^{\mathrm{T}} \left(\mathbf{H} \mathbf{P}_{k|k-1}^{\boldsymbol{\xi}} \mathbf{H}^{\mathrm{T}} + \mathbf{R} \right)^{-1} \tag{24}$$

Eq. (15) – Eq. (24) formulate the essential propagation formulae of a dual KF algorithm for updating the unknown sensor faults and the structural states vector online.

3.3 Summary

The procedure of the proposed online sensor fault diagnosis approach is summarized as follows:

Sensor fault detection and localization

(1) Construct the meta-model for sensor fault detection and localization based on 1-D CNN.

(2) For each training iteration,

(2.1) Sample a batch of sensor fault detection and localization tasks $\mathcal{T}_j^{tr}, j = 1,2, \dots, N_{ba}$ from training dataset $\mathcal{D}_{\text{train}}$.

(2.2) For each task $\mathcal{T}_j^{tr}, j = 1,2, \dots, N_{ba}$, calculate the updated initial parameters of the sub-model using Eqs. (8) and (9).

(2.3) Calculate the optimal initial parameters of the meta-model using Eqs. (10) and (11).

(2.4) Sample a sensor fault detection and localization task $\mathcal{T}^{te}$ from testing dataset $\mathcal{D}_{\text{test}}$.

(2.5) Fine-tune the optimal initial parameters of the meta-model using Eqs. (12) and (13).

(2.6) Calculate the testing classification accuracy using Eq. (14).

(2.7) Save the fine-tuned meta-model with the highest testing classification accuracy as the best meta-model. Check the following criteria:

a If the testing classification accuracy decreases for five consecutive training iterations or the training iteration reaches the maximum number of training iterations $\overline{N}_{train}$, it implies that the training process is completed. Then go to Step (3).

b Otherwise, repeat Step (2).

(3) Classify the N_{win} aggregated data points in time windows for all N_o measurement channels. Check the following criteria:

a If the predicted labels remain 1 for N_{con} consecutive time windows, the specific location of the faulty sensors can be determined. Then go to Step (8).

b Otherwise, go to Step (4).

(4) Sample a sensor fault detection and localization task $\mathcal{T}^{te}$ from testing dataset $\mathcal{D}_{\text{test}}$.

(5) Fine-tune the optimal initial parameters of the meta-model using Eqs. (12) and (13).

(6) Calculate the testing classification accuracy using Eq. (14).

(7) Save the fine-tuned meta-model with the highest testing classification accuracy as the best meta-model. Then go back to Step (3) for the next round of sensor fault detection and localization.

Sensor fault and structural state vector estimations

(8) Quantify the unknown sensor faults and update the structural states vector using the proposed dual KF algorithm.

4 Illustrative Numerical Example: Three-Span Bridge

A three-span bridge (shown in Figure 4) is utilized to demonstrate the proposed approach. The deck of the bridge has a uniform box cross-section with area A_d=1.65 m^2 and moment of inertia

I_d=4.28 m^4. The two piers have uniform circular cross-section with radius r=0.6 m. The length between two consecutive nodes on the deck is 10 m, and the length between two consecutive nodes on the pier is 6 m. The mass density is ρ=3500 kg/m^3, and the modulus of elasticity is E=10^9 Pa. As a result, the first five modal frequencies of the bridge are 0.331, 0.833, 0.986, 1.028, and 1.085 Hz. The Rayleigh damping model is utilized, so the damping matrix is given by $\mathbf{L} = \alpha_1\mathbf{M} + \alpha_2\mathbf{K}$, where $\alpha_1 = 0.0603\ \mathrm{s}^{-1}$ and $\alpha_2 = 0.0053$ s. As a result, the modal damping ratio is 2% for the first two modes.

Vertical acceleration responses of the 2nd, 3rd, 6th, 8th, 9th, 10th, 15th, and 16th nodes, and horizontal acceleration responses of the 18th, 19th, 21st, and 22nd nodes are measured. The sampling rate is taken as 500 Hz. The measurement noise level is taken to be 10% root mean square (RMS) of the corresponding noise-free response quantities. The bridge is subjected to horizontal and vertical ground excitations

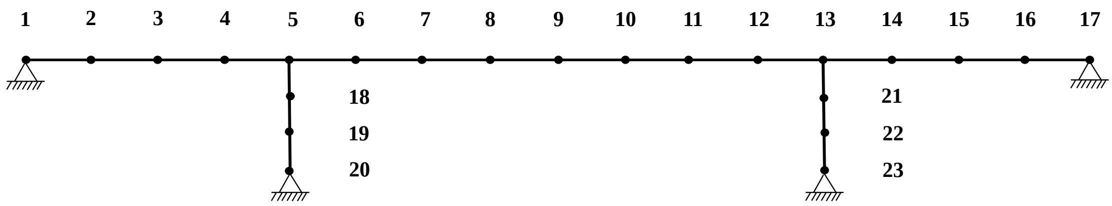

Figure 4. Three-span bridge system

4.1 Validation of the proposed method

In this subsection, the ground excitations in both horizontal and vertical directions are modeled as zero-mean Gaussian white noise with spectral intensity 1.5×10^{-8} m^2/s^3.

4.1.1 Generation of training and testing dataset

Both the length of the training and that of the testing dataset are taken as 40 s, and the total number of data points for each sensor is 20000. In this study, the regular and faulty samples are generated using 1000 data points, and both the number of regular samples and that of faulty samples are 20. Four typical sensor faults are considered: (1) bias, (2) drift, (3) gain, and (4) precision degradation (PD). For each sensor, the 20 faulty samples contain the four types of sensor faults, and each faulty type corresponds to five samples. In this section, measurements with bias faults are given by $\tilde{z}(t) = z(t) + 0.2RMS\big(z^*(t)\big)$, where $z(t)$ is the regular noisy measurement, and $RMS\big(z^*(t)\big)$ represents the RMS of the noise-free response quantity $z^*(t)$. Measurements with drift faults are given by $\tilde{z}(t) = z(t) + 0.2RMS\big(z^*(t)\big) + 0.0005t$. Measurements with gain faults are given by $\tilde{z}(t) = 1.3z(t)$. Measurements with PD faults are given by $\tilde{z}(t) = z(t) + \varrho(t)$, where $\varrho(t)$ is an additional zero-mean Gaussian noise with variance δ_ϱ. δ_ϱ is assumed to be 80% RMS of the corresponding noise-free response quantity.

4.1.2 Performance evaluation

In this study, N_{up}, N_{con}, N_{ba}, N_S, and N_Q are taken as 2, 2, 20, 5, and 5, respectively. The inner and outer loop learning rates are taken as 0.001 and 0.01, respectively. The maximum number of training iterations is taken as 200. To demonstrate the superiority of the proposed method, four conventional deep learning methods, namely, long-short term memory (LSTM) network, gated recurrent unit (GRU) network, convolutional neural network (CNN), and residual neural network (ResNet) are utilized for comparison. The proposed method and these conventional methods are executed on the same computing system, comprising an AMD R7-5800H @3.2 GHz CPU, NVIDIA RTX3060 GPU, Windows 11, CUDA 11.3, and Torch 1.14. The training and testing processes of the five methods are performed 50 times independently. The comparison results of average testing classification accuracy are shown in Table 2. It is seen that the proposed method outperforms the conventional methods with a large margin. The dataset used for sensor fault diagnosis is described as follows. The entire monitoring period is 100 s. In this subsection, the 1st, 5th, and 10th sensors are assumed to have bias faults at the beginning of the monitoring period, and the bias levels of the 1st, 5th, and 10th sensors are set to be 20%, 25%, and 30% RMS of the corresponding noise-free response quantities, respectively.

Figure 5 shows a typical set of measurements from the 1st, 5th, and 10th sensors. Figure 6 shows the output labels of the 2nd, 3rd, 4th, 5th, and 10th sensors using the proposed method. It is observed that the proposed method provides accurate results for detecting and locating the faulty sensors in the network

Table 2. Comparison of average testing classification accuracy

Method		LSTM	GRU	CNN	ResNet	Proposed
Accuracy (%)	Section 4.1.2	63.1	66.5	72.9	74.3	93.6
	Section 4.2.2	68.9	68.2	72.6	75.6	92.1

Two indicators, namely, false alarm rate and miss detection rate, are utilized to evaluate the performance of faulty sensor detection and localization over the 50 independent runs. Table 3 shows the comparison results of outlier detection for the conventional methods (i.e., LSTM, GRU, CNN, and ResNet). The first, second, and third rows indicate the measurement channel, location, and fault type of the sensor, respectively. The fourth, fifth, sixth, seventh, and eighth rows show the false alarm rate using LSTM, GRU, CNN, ResNet, and the proposed approach, respectively. The ninth, tenth, eleventh, twelfth, and thirteenth rows show the miss detection rate using LSTM, GRU, CNN, ResNet, and the proposed approach, respectively. It is observed that both the false alarm rate and the miss detection rate are close to 0 by using the proposed method. This implies that the proposed method provides satisfactory results for the detection and localization of faulty sensors. However, the conventional methods are unable to detect and locate the faulty sensors, since the miss detection rates are equal to 1. This is attributed to the challenges of training a deep learning model with limited data. The comparison results of the false alarm rate and miss detection rate demonstrate the superior effectiveness of the proposed method over the conventional methods.

During the period from t=0 s to t=4 s, the output labels of the 1st, 5th, and 10th sensors remained 1 for two consecutive time windows. It indicated that the detection and localization stage is completed at t=4 s. Then the algorithm shifts to the stage of sensor fault estimation. In the estimation stage, the severity of the sensor faults and structural states can be updated simultaneously by using a dual KF in an online manner. Our study chose the 99.7% credible interval, which is widely used in statistical analysis [40] and parameter estimation [41]. By striking a balance between being informative and not overly conservative, this interval provides a reasonable level of confidence in parameter estimation and enables a comprehensive assessment of the uncertainty associated with results. Figure 7 shows the time-history estimation results of the sensor faults, and it is observed that the estimated sensor faults are accurate within the 99.7% credible intervals. Besides, the fault type of the sensors can be determined to be bias fault as the estimated sensor faults are time-invariant and remain at specific values.

Figure 8 shows the estimated vertical displacement response at the 7th node and the estimated horizontal velocity response at the 14th node versus their actual values. The 45° line of perfect match is drawn for reference. It can be observed that the proposed method provided a satisfactory prediction of the structural states as all points are distributed closely around the 45° line.

In conclusion, the proposed approach can not only provide satisfactory detection and localization results for faulty sensors but also estimate the severity of the sensor faults and update the structural system in an online manner.

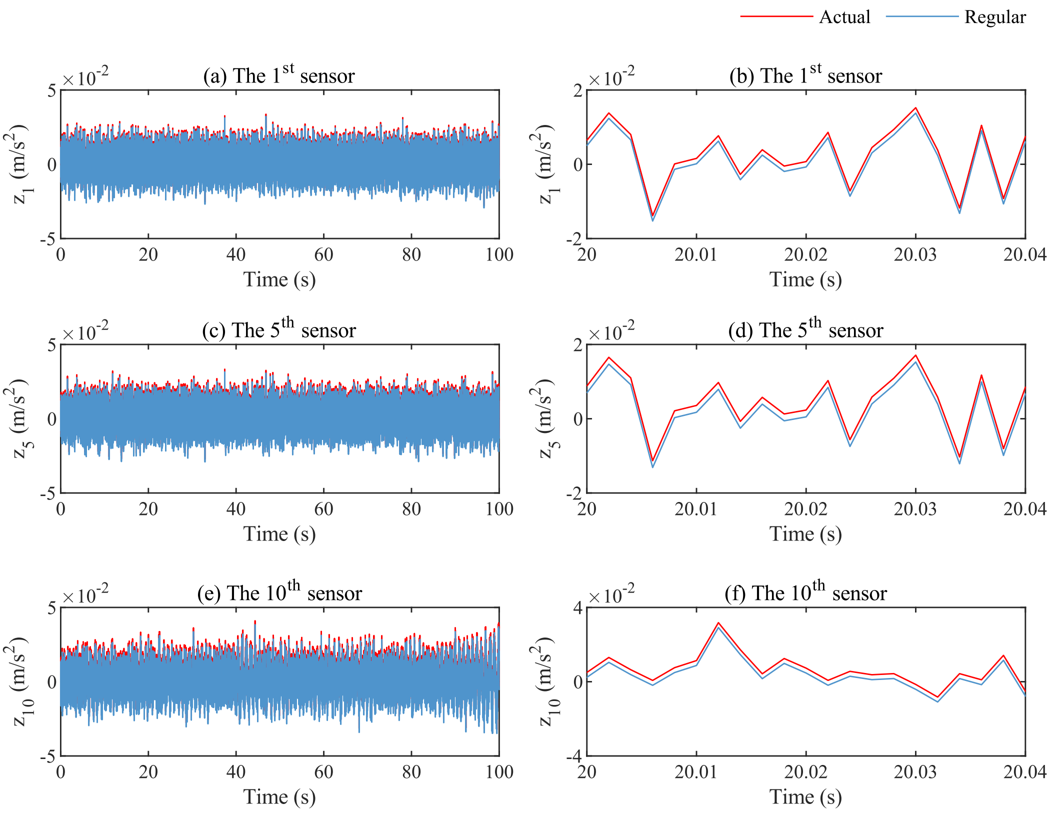

Figure 5. Typical measurements from the 1st, 5th, and 10th sensors (Section 4.1.2.)

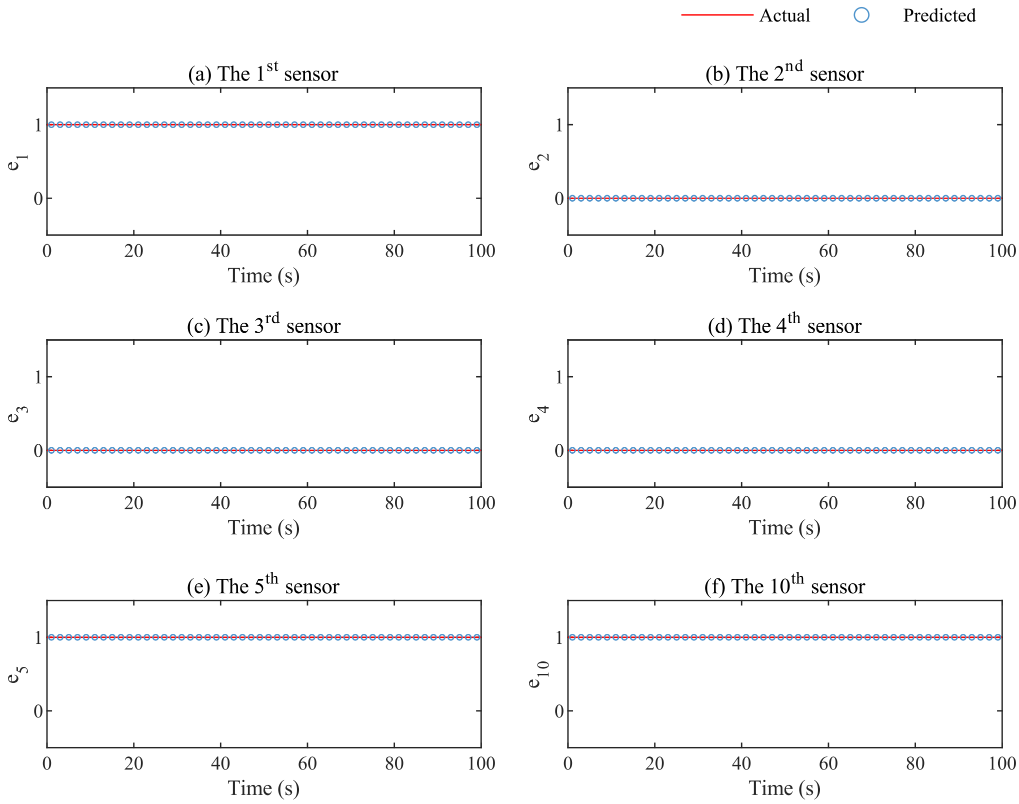

Figure 6. Output labels of the 1st, 2nd, 3rd, 4th, 5th, 8th, 10th, and 11th sensors (Section 4.1.2.)

Table 3. Comparison of outlier detection results (Section 4.1.2.)

Sensor number		1	2	3	4	5	6	7	8	9	10	11	12
Location		2nd	3rd	6th	8th	9th	10th	15th	16th	18th	19th	21st	22nd
Fault		Bias	-	-	-	Bias	-	-	-	-	Bias	-	-
False alarm rate	LSTM	-	0.00	0.02	0.00	-	0.00	0.00	0.02	0.00	-	0.00	0.02
	GRU	-	0.00	0.00	0.00	-	0.00	0.00	0.00	0.00	-	0.00	0.02
	CNN	-	0.00	0.00	0.00	-	0.00	0.00	0.02	0.00	-	0.00	0.00
	ResNet	-	0.00	0.02	0.00	-	0.00	0.00	0.02	0.00	-	0.00	0.00
	Proposed	-	0.00	0.00	0.00	-	0.00	0.00	0.00	0.00	-	0.02	0.00
Miss detection rate	LSTM	1.00	-	-	-	1.00	-	-	-	-	1.00	-	-
	GRU	1.00	-	-	-	1.00	-	-	-	-	1.00	-	-
	CNN	1.00	-	-	-	1.00	-	-	-	-	1.00	-	-
	ResNet	1.00	-	-	-	1.00	-	-	-	-	1.00	-	-
	Proposed	0.00	-	-	-	0.00	-	-	-	-	0.00	-	-

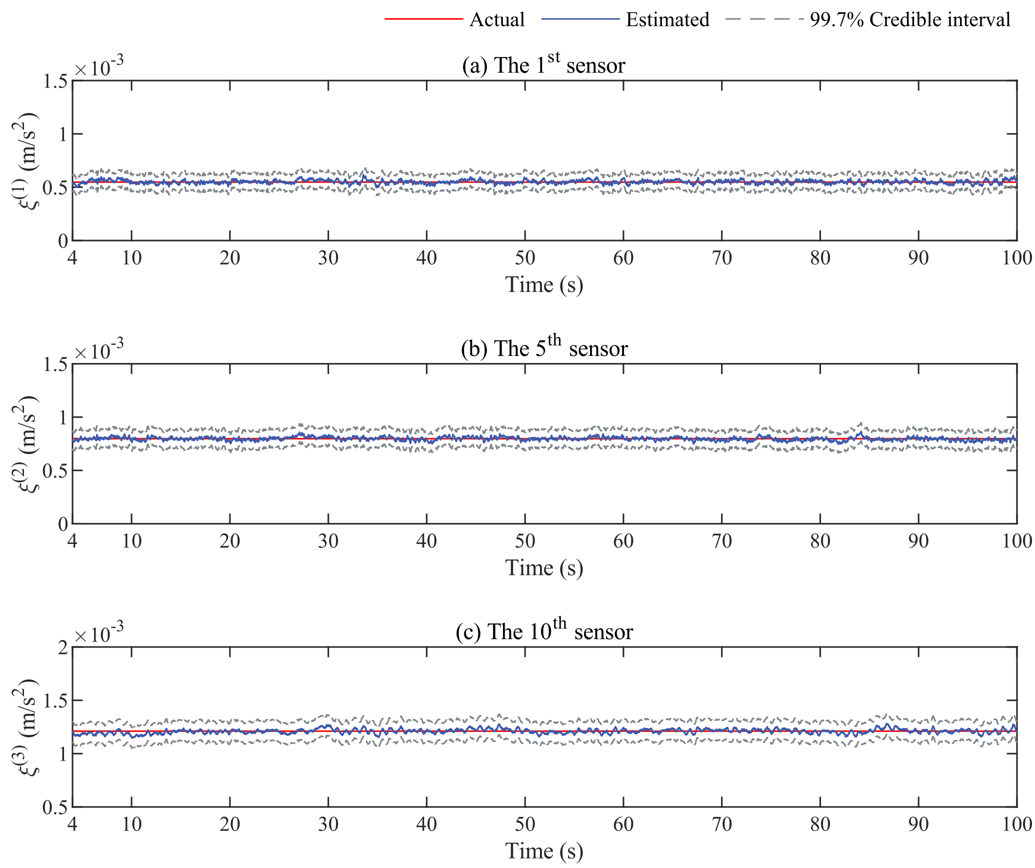

Figure 7. Time history estimation results of the sensor faults (Section 4.1.2.)

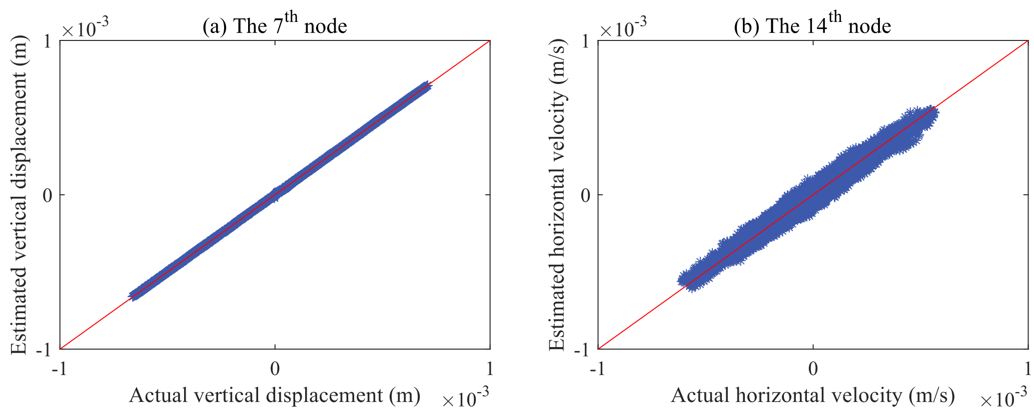

Figure 8. Estimation results of structural responses for the 7th and 14th nodes (Section 4.1.2.)

4.2 Meta-learning under nonstationary ground excitation

In this subsection, the ground excitations in both horizontal and vertical directions are modeled as modulated zero-mean Gaussian white noise process $\mathbf{f}(t) = A_f(t)\boldsymbol{g}(t)$. The amplitude modulating function $A_f(t)$ is given by:

$$A_f(t) = \begin{cases} 1, & 0 \leq t < t_2 \\ 1 + \frac{(t-t_2)}{4t_1} \exp\left(1 - \frac{t-t_2}{4t_1}\right), & t \geq t_2 \end{cases} \tag{25}$$

where $t_1 = 10s$ and $t_2 = 20s$. The spectral intensity of $\boldsymbol{g}(t)$ is taken as 5.5×10^{-8} m^2/s^3.

4.2.1 Generation of training and testing dataset

To verify the generalization ability of the proposed method, the training dataset is taken as the same training dataset as Section 4.1.1., and the testing dataset is generated under the above nonstationary ground excitation. As a result, the main difference between the testing dataset of Section 4.2.1. and that of Section 4.1.1. relies on unmeasured excitations applied to the system.

4.2.2 Performance evaluation

The comparison results of the average testing classification accuracy over 50 runs are shown in Table 2. It can be observed that the proposed method outperforms the conventional methods since it possesses the largest average testing classification accuracy.

The dataset used for sensor fault diagnosis is described as follows. The diagnosis dataset is generated under the nonstationary ground excitation. The monitoring period is 100 s. In this subsection, the 1^{st} sensor is assumed to have bias fault, and its measured data are given by $\tilde{z}(t) = z(t) + 0.3RMS\big(z^*(t)\big)$. The 4^{th} sensor is assumed to have drift fault, and its measured data are given by $\tilde{z}(t) = z(t) + 0.15RMS\big(z^*(t)\big) + 0.002t$. The 8^{th} sensor is assumed to have gain fault, and its measured data are given by $\tilde{z}(t) = 1.6z(t)$. The 11^{th} sensor is assumed to have PD fault, and its measured data are given by $\tilde{z}(t) = z(t) + \varrho(k)$. The variance of $\varrho(k)$ is assumed to be 120% RMS of the corresponding noise-free response quantity. Moreover, the 1^{st} and 4^{th} sensors are assumed to be faulty at t=10 s, and the 8^{th} and 11^{th} sensors are assumed to be faulty at t=20 s. It is noted that the aforementioned information is unknown throughout the entire process of diagnosis.

Figure 9 shows a typical set of measurements from the 1^{st}, 4^{th}, 8^{th}, and 11^{th} sensors. Figure 10 shows the output labels of the 1^{st}, 4^{th}, 8^{th}, 9^{th}, 10^{th}, and 11^{th} sensors using the proposed method. The results show that the proposed method can precisely detect and locate faulty sensors in the network.

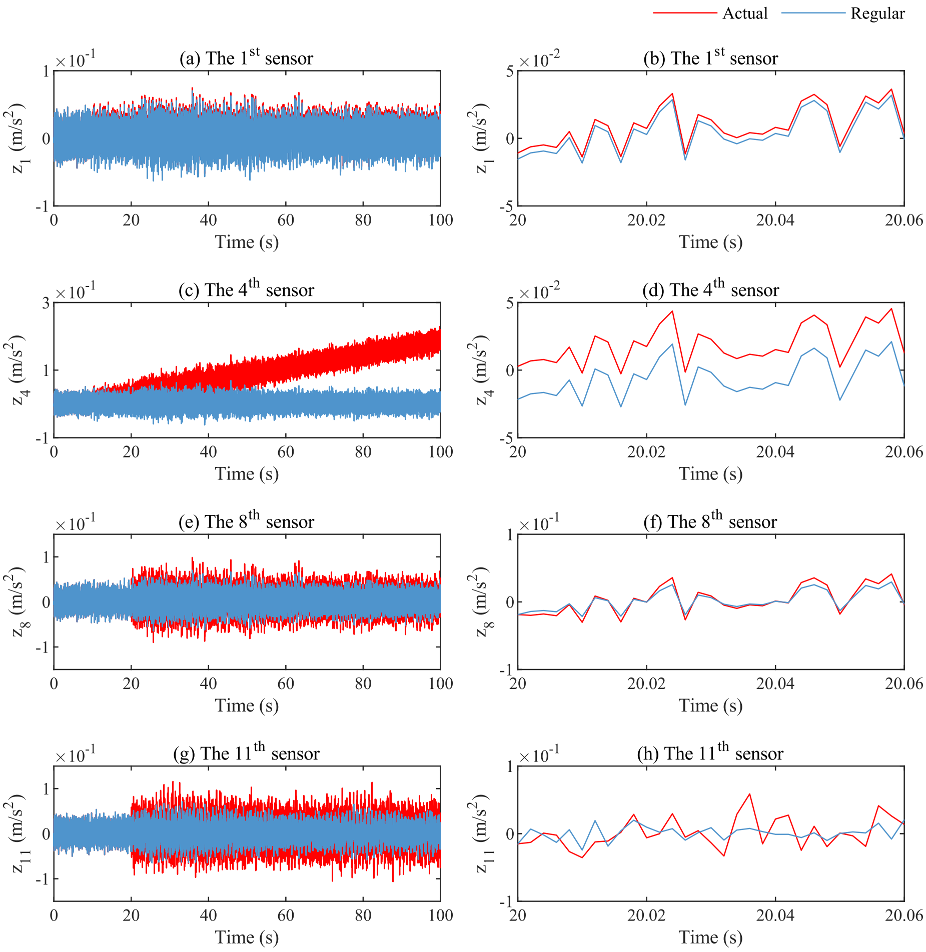

Figure 9. Typical measurements from the 1st, 4th, 8th, and 11th sensors (Section 4.2.2.)

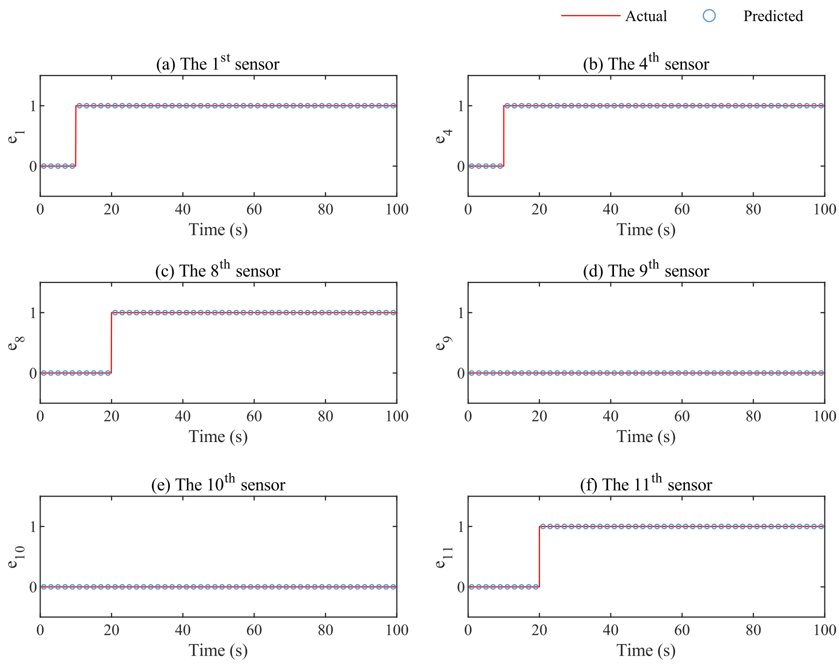

Figure 10. Output labels of the 1st, 4th, 8th, 9th, 10th, and 11th sensors (Section 4.2.2.)

Table 4 presents the comparison results of outlier detection in a similar fashion as Table 3. The proposed method yields false alarm rates and miss detection rates close to 0. This indicates that the proposed method gives satisfactory results for the detection and localization of faulty sensors. Compared with the proposed method, the conventional methods fail to recognize the faults in the 1^{st}, 8^{th}, and 11^{th} sensors. However, the 4^{th} sensor with drift fault is successfully identified by using the conventional methods. The reason is elaborated as follows. Drift faults typically involve a gradual change in the baseline behavior of the sensing data over time. They are easily detectable because they result in noticeable deviations from the historical behavior for long-term monitoring. On the other hand, bias, gain, and PD faults are usually subtle, and they might not result in noticeable deviations in the data.

Table 4. Comparison of outlier detection results (Section 4.2.2.)

Sensor number		1	2	3	4	5	6	7	8	9	10	11	12
Location		2^{nd}	3^{rd}	6^{th}	8^{th}	9^{th}	10^{th}	15^{th}	16^{th}	18^{th}	19^{th}	21^{st}	22^{nd}
Fault		Bias	-	-	Drift	-	-	-	Gain	-	-	PD	-
False alarm rate	LSTM	-	0.00	0.00	-	0.00	0.00	0.02	-	0.00	0.00	-	0.00
	GRU	-	0.00	0.02	-	0.00	0.00	0.02	-	0.00	0.00	-	0.00
	CNN	-	0.02	0.00	-	0.02	0.00	0.02	-	0.00	0.02	-	0.00
	ResNet	-	0.00	0.00	-	0.00	0.00	0.02	-	0.00	0.00	-	0.00
	Proposed	-	0.00	0.02	-	0.02	0.00	0.00	-	0.00	0.00	-	0.00
Miss detection rate	LSTM	1.00	-	-	0.00	-	-	-	1.00	-	-	1.00	-
	GRU	1.00	-	-	0.00	-	-	-	1.00	-	-	1.00	-
	CNN	1.00	-	-	0.00	-	-	-	1.00	-	-	1.00	-
	ResNet	1.00	-	-	0.00	-	-	-	1.00	-	-	1.00	-
	Proposed	0.00	-	-	0.00	-	-	-	0.00	-	-	0.00	-

PD: precision degradation

The detection and localization stage for the 1^{st} and 4^{th} sensors is completed at t=14 s, and the detection and localization stage for the 8^{th} and 11^{th} sensors is completed at t=24 s. Figure 11 shows the time-history estimation results of the sensor faults. The estimated sensor faults are accurate within the 99.7% credible intervals. Moreover, the fault types of the 1^{st}, 4^{th}, 8^{th}, and 11^{th} sensors can be determined correctly by analyzing the variations in the estimated sensor faults over time.

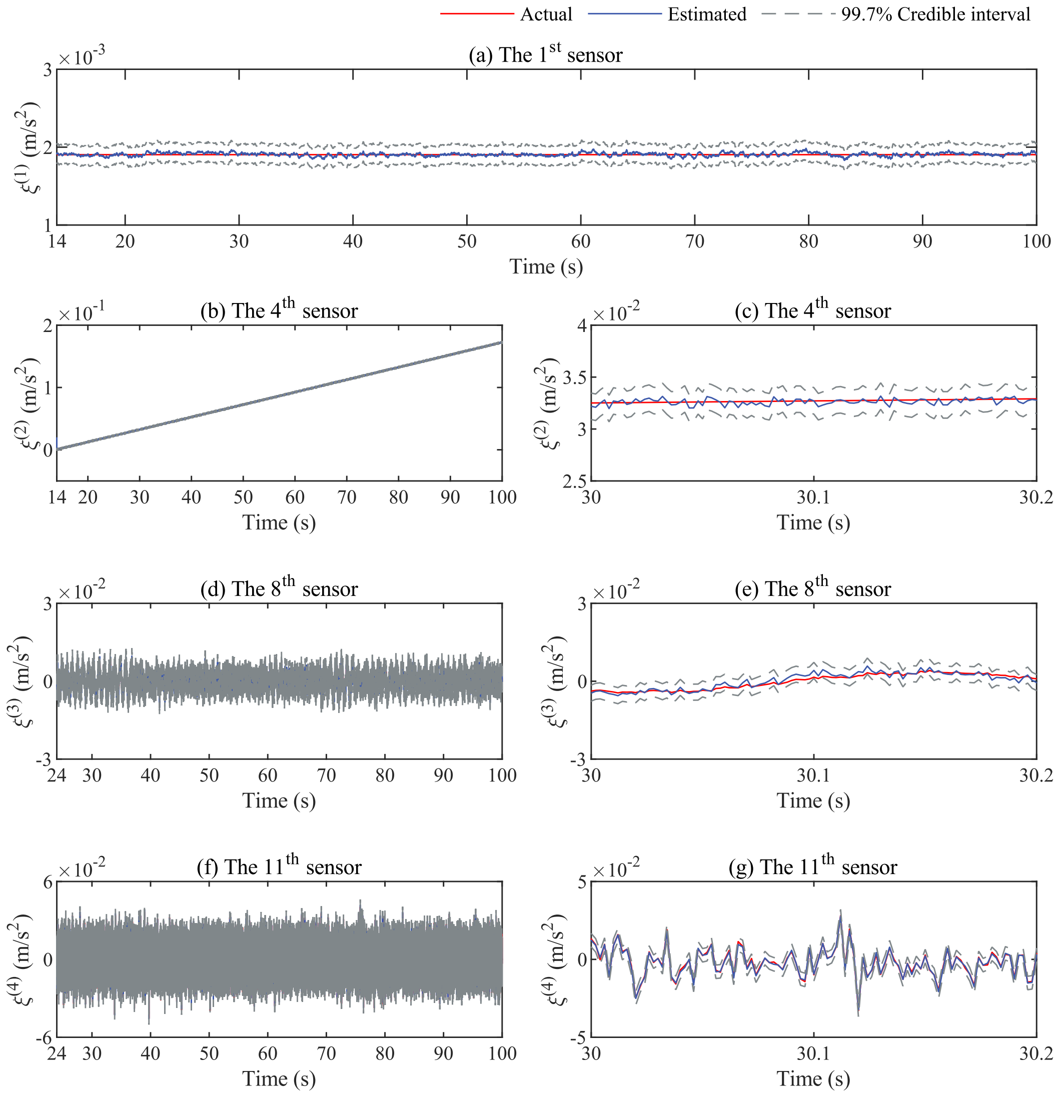

Figure 11. Time-history estimation results of the sensor faults (Section 4.2.2.)

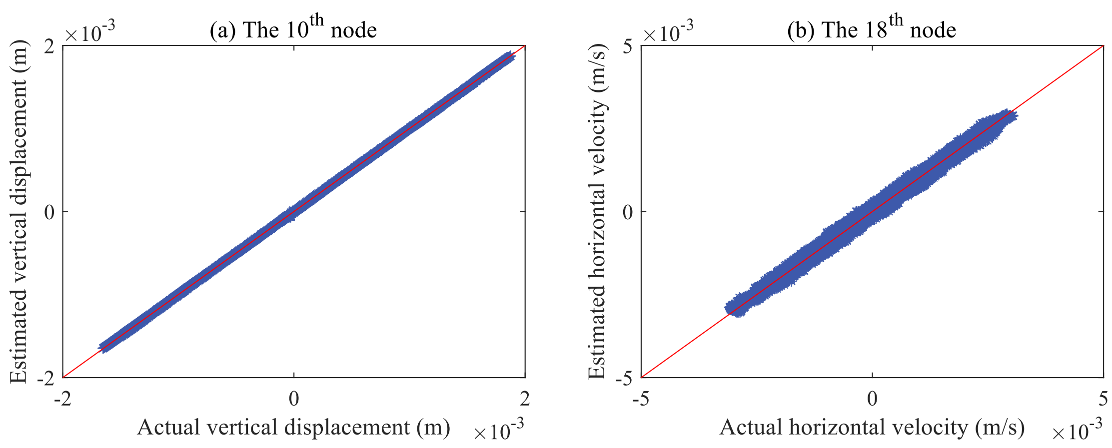

Figure 12. Estimation results of structural responses for the 10th and 18th nodes (Section 4.2.2.)

Figure 12 shows the estimated vertical displacement response at the 10th node and the estimated horizontal velocity response at the 18th node versus their actual values. It can be observed that the proposed method provides a satisfactory prediction of the structural states as all points are distributed closely around the line of perfect match.

In conclusion, the proposed approach is capable of detection, localization, and estimation for typical sensor faults. Besides, the proposed method has superior model generalization capability under nonstationary ground excitation

5 Application to Field Measurements and Simulations from The Canton Tower

In this section, a benchmark project of the Canton Tower [42] is utilized to validate the effectiveness of the proposed method. The Canton Tower is a supertall structure with a height of 610 m. In the benchmark project, 20 uniaxial accelerometers with a sampling frequency of 50 Hz are mounted at different levels and directions throughout the structure (shown in Figure 13). From the 20 accelerometers, 24 hours of field measurements are obtained. The field measurement data can be found in 31. On the other hand, a reduced-order three-dimensional beam model developed from the benchmark project is considered in this section. The Canton Tower is modeled as a cantilever beam with 37 elements and 38 nodes. The nodal number increases from 1 at the fixed base to 38 at the free top end. A specific description of the reduced-order model can be found in 42. The first five modal frequencies of the reduced-order model are 0.110, 0.159, 0.347, 0.368, and 0.399 Hz. The Rayleigh damping model is assumed. The damping matrix is given by $\mathbf{L} = \alpha_1\mathbf{M} + \alpha_2\mathbf{K}$, where the Rayleigh damping coefficients are $\alpha_1 = 0.0164\ \mathrm{s}^{-1}$ and $\alpha_2 = 0.0237\ \mathrm{s}$. As a result, the modal damping ratio is 2% for the first two modes.

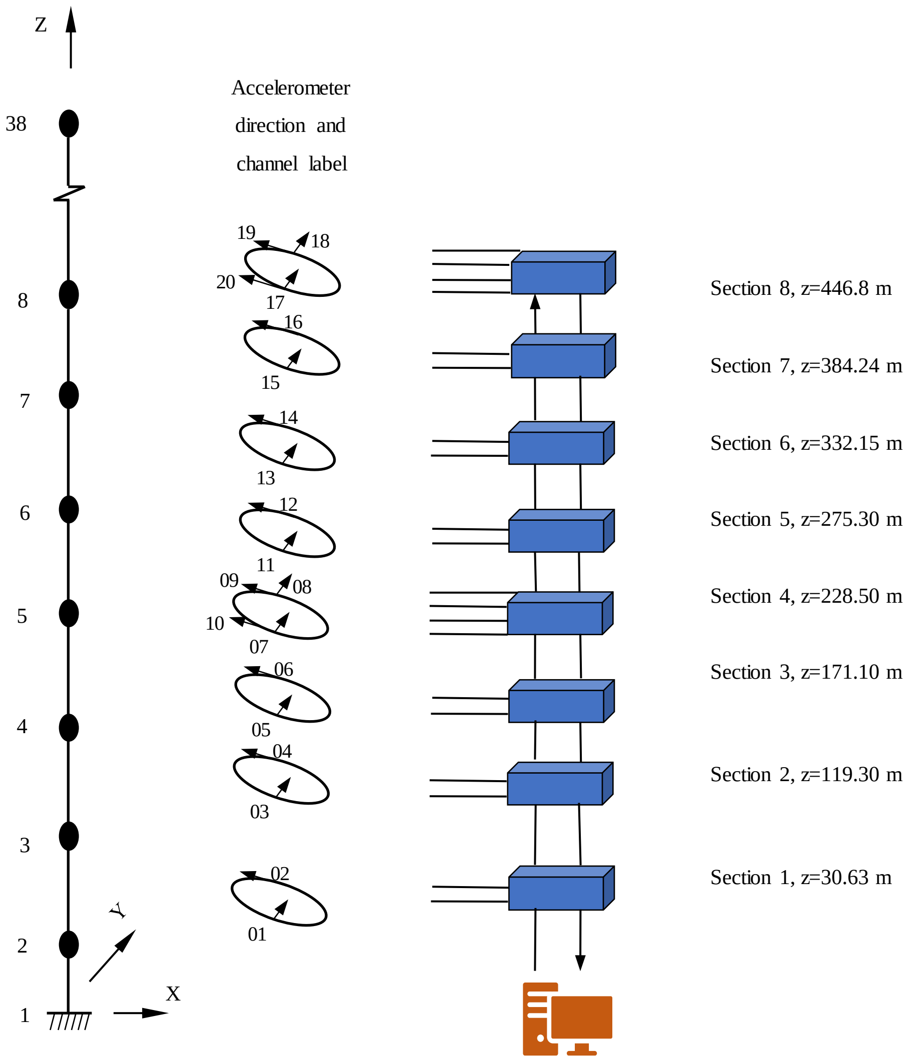

Figure 13. Deployment of accelerometers in the Canton tower [42]

5.1 Generation of training and testing dataset

The field measurements of acceleration recorded from 18:00 to 19:00 on 19th January 2010 are utilized to generate the training and testing dataset. Sensor faults are simulated in each sensor by changing the corresponding field measurements. Both the length of the training and testing dataset are taken as 400 s, and the total number of data points for each sensor is 20000. The training and testing datasets are generated in the same way as in Section 4.1.1. Specifically, the measurements with bias faults are given by $\tilde{z}(t) = z(t) + 0.3RMS\big(z(t)\big)$. The measurements with drift faults are given by $\tilde{z}(t) = z(t) + 0.15RMS\big(z(t)\big) + 1 \times 10^{-7}t$. The measurements with gain faults are given by $\tilde{z}(t) = 1.6z(t)$. The measurements with PD faults are given by $\tilde{z}(t) = z(t) + \varrho(k)$, and the variance of $\varrho(k)$ is assumed to be 150% RMS of the corresponding fault-free response quantity.

5.2 Performance evaluation

The training and testing processes of the proposed method are performed 50 times independently. The average value of the testing classification accuracy for the proposed method over the 50 runs is 95.2%. The dataset used for sensor fault diagnosis is described as follows. The diagnosis dataset is taken from the field measurements of acceleration recorded from 20:00 to 21:00 on 19th January 2010. The entire monitoring period is 500 s. Sensor faults are simulated by choosing four sensors to be faulty during the monitoring period. In particular, the 1st sensor is assumed to have bias fault, and its measured data are given by $\tilde{z}(t) = z(t) + 0.3RMS\big(z(t)\big)$. The 6th sensor is assumed to have drift fault, and its measured data are given by $\tilde{z}(t) = z(t) + 0.3RMS\big(z(t)\big) + 5 \times 10^{-7}t$. The 11th sensor is assumed to have gain fault, and its measured data are given by $\tilde{z}(t) = 1.7z(t)$. The 16th sensor is assumed to have PD fault, and its measured data are given by $\tilde{z}(t) = z(t) + \varrho(k)$, and the variance of $\varrho(k)$ is assumed to be 120% RMS of the corresponding fault-free response quantity. Moreover, the 1st and 6th sensors are assumed to be faulty at t=100 s, and the 11th and 16th sensors are assumed to be faulty at t=150 s. It is noted that the aforementioned information is unknown throughout the entire process of diagnosis.

Figure 14 shows a typical set of measurements from the 1st, 6th, 11th, and 16th sensors. Figure 15 displays the output labels of the 1st, 2nd, 6th, 9th, 11th, and 16th sensors. The results indicate that the proposed method detects and locates the faulty sensor in the network accurately.

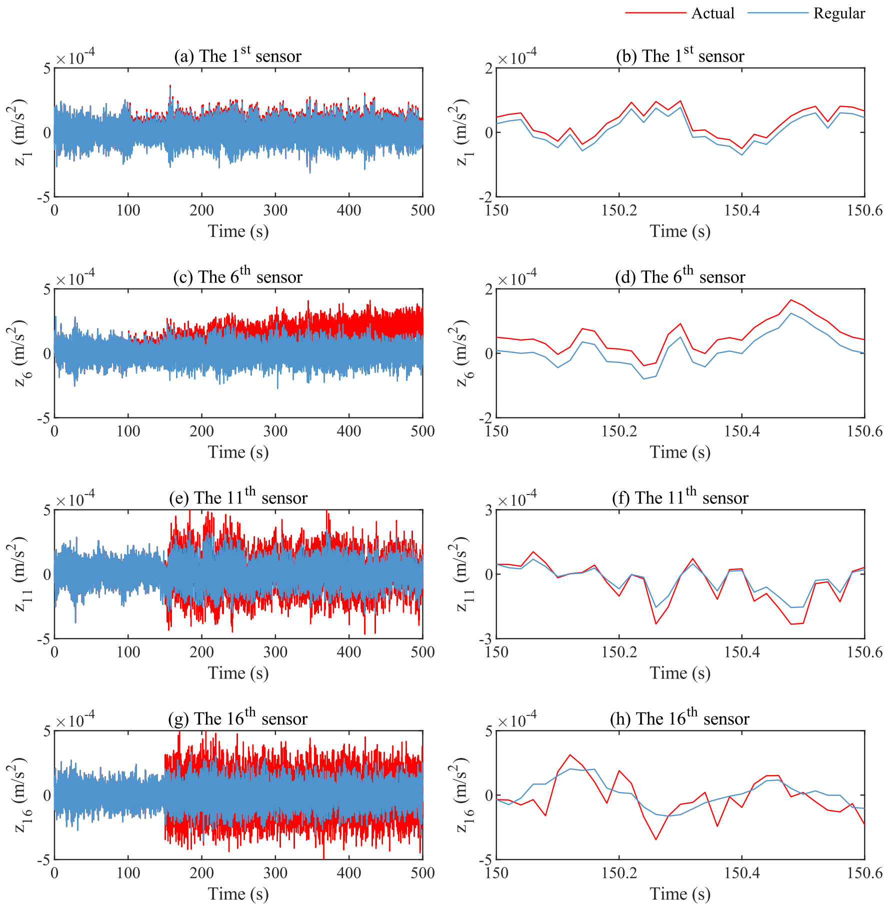

Figure 14. Typical measurements from the 1st, 6th, 11th, and 16th sensors (Section 5.2.)

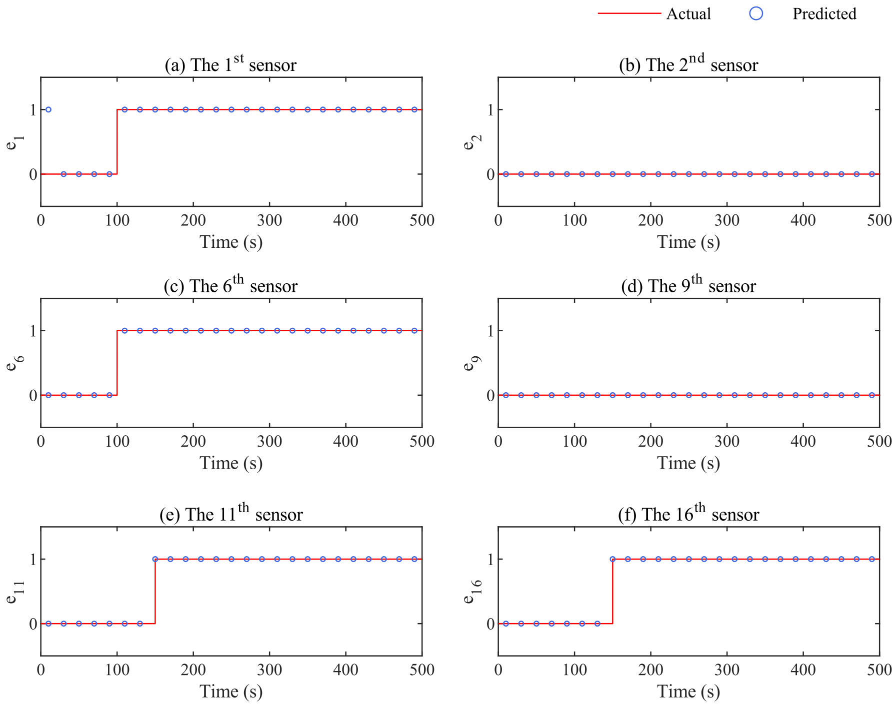

Figure 15. Output labels of the 1st, 2nd, 6th, 9th, 11th, and 16th sensors (Section 5.2.)

Table 5 presents the outlier detection results of the proposed method. The first and second columns indicate the measurement channel and fault type of the sensor, respectively. The third and fourth columns show the false alarm rate and miss detection rate over 50 runs using the proposed approach, respectively. It is observed that both the false alarm rate and miss detection rate are close to 0. This indicates that the proposed method gives satisfactory detection and localization results for faulty sensors.

Table 5. Outlier detection results of the proposed method (Section 5.2.)

Sensor number	Fault	False alarm rate	Miss detection rate
1	Bias	-	0.04
2	-	0.00	-
3	-	0.00	-
4	-	0.00	-
5	-	0.00	-
6	Drift	-	0.00
7	-	0.02	-
8	-	0.00	-
9	-	0.00	-
10	-	0.00	-
11	Gain	-	0.00
12	-	0.06	-
13	-	0.02	-
14	-	0.08	-
15	-	0.00	-
16	Precision degradation	-	0.00
17	-	0.00	-
18	-	0.00	-
19	-	0.00	-
20	-	0.00	-

The detection and localization stage for the 1st and 6th sensors is completed at t=140 s, and the detection and localization stage for the 11th and 16th sensors is completed at t=190 s. Figure 16 shows the time-history estimation results of the sensor faults. It is observed that the proposed approach can provide a reasonable estimation for different types of sensor faults.

Figure 17 shows the estimated acceleration response of the 5th node versus their actual values. It can be observed that the proposed method provides a satisfactory prediction of the structural states as all points lay closely along the line of perfect match.

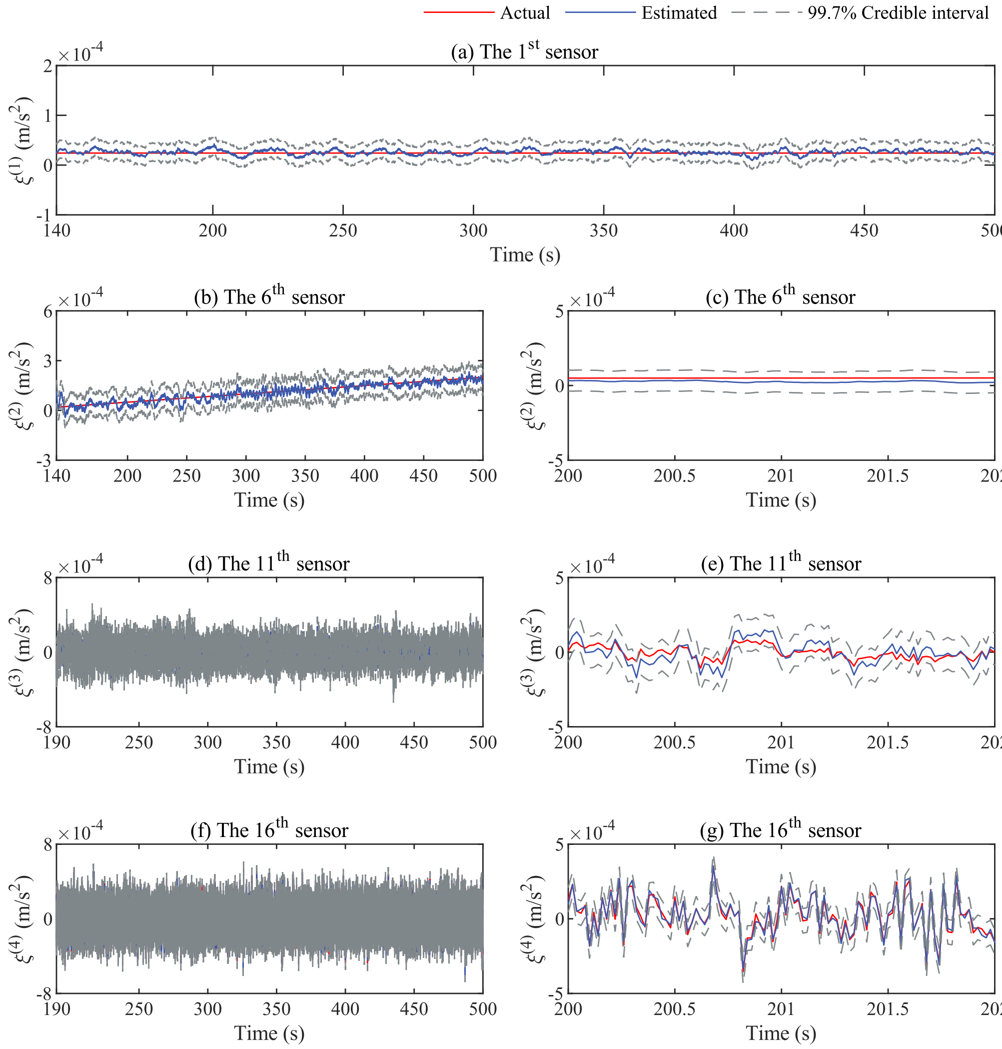

Figure 16. Time-history estimation results of the sensor faults (Section 5.2.)

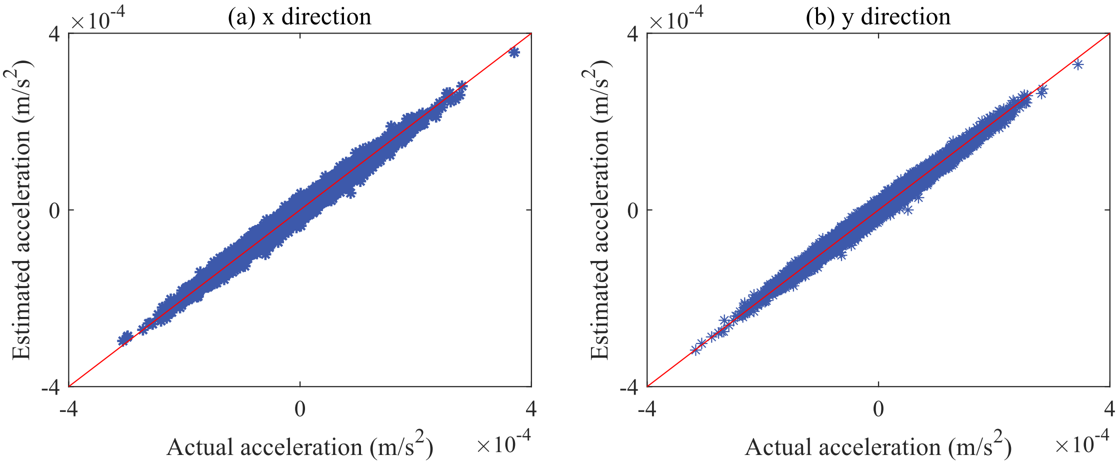

Figure 17. Estimation results of structural responses for the 5th node (Section 5.2.)

This application validates the efficacy of the proposed approach on online sensor fault diagnosis for SHM in practice. The results show that the proposed approach provides an effective and practical tool for sensor fault diagnosis.

6 Conclusion

A novel online sensor fault diagnosis approach based on meta-learning has been presented. The proposed approach deals with the typical problems in structural health monitoring for sensor fault detection, localization, and sensor fault estimation. The detection and localization of faulty sensors are realized by a one-dimensional convolutional neural network (1-D CNN) based on meta-learning. Then, the sensor faults and structural states are estimated by an online updating algorithm based on a dual KF. The proposed approach can improve the efficiency and accuracy of conventional 1-D CNN for sensor fault detection and localization using limited data, and it can estimate the sensor faults and structural states in an online manner. Numerical validation on a bridge structural system subjected to ground excitation with various types of sensor faults has been conducted. The validation result demonstrates that the proposed approach has more competitive performance than the conventional deep learning methods. In addition, the proposed approach achieves satisfactory performance in the field measurements and simulated sensor faults from the Canton Tower. The proposed approach provides an effective and applicable tool in sensor fault diagnosis for SHM.

Future research directions for the proposed method encompass two main aspects:

1. The flexibility of the proposed method under limited unlabeled data can be further improved by allowing an alternative model for sensor fault detection and localization.
2. Since both sensor faults and structural damages can lead to anomaly data, the authors will investigate the possibility of distinguishing between anomaly data caused by sensor faults and structural damages.

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.

Acknowledgements

This work was supported by the National Key Research and Development Program of China (Grant No. 2021YFB2600900), the National Natural Science Foundation of China (Grant No. 52008037), the Natural Science Foundation of Hunan Province, China (Grant No. 2021JJ40582), the Research Foundation of Education Bureau of Hunan Province, China (Grant No. 21B0294), the Science and Technology Innovation Program of Hunan Province of China (Grant No. 2023GK2036), the Postgraduate Scientific Research Innovation Project of Hunan Province (Grant No. CX20220856), and China Scholarship Council. These supports are gratefully acknowledged.

Reference

1 Ma Y F, He Y, Wang G D, Wang L, Zhang J R and Lee D 2023 Corrosion fatigue crack growth prediction of bridge suspender wires using Bayesian Gaussian process *Int. J. Fatigue* **168** 107377

2 Luo Y, Liu X F, Chen F H, Zhang H P and Xiao X H 2023 Numerical simulation on crack–inclusion interaction for rib-to-deck welded joints in orthotropic steel deck *Metals* **13** 1402

3 Xu X, Ren Y, Huang Q, Fan Z Y, Tong Z J, Chang W J and Liu B 2020 Anomaly detection for large span bridges during operational phase using structural health monitoring data *Smart Mater. Struct.* **29** 045029

4 Huang K and Yuen K V 2020 Hierarchical outlier detection approach for online distributed structural identification *Struct. Control. Health Monit.* **27** e2623

5 Su J, Xia Y and Weng S 2020 Review on field monitoring of high-rise structures *Struct. Control. Health Monit.* **27** e2629

6 Gong X Y, Song X D, Cai C S, Li G Q and Xiong W 2023 Early warning for abnormal strains of continuous bridges using a proposed temperature–strain mapping model *Smart Mater. Struct.* **32** 125018

7 Yi T H, Huang H B and Li H N 2017 Development of sensor validation methodologies for structural health monitoring: a comprehensive review *Measurement* **109** 200-214

8 Fu Y, Peng C, Gomez F, Narazaki Y and Spencer B F 2019 Sensor fault management techniques for wireless smart sensor networks in structural health monitoring *Struct. Control. Health Monit.* **26** e2362

9 Bao Y Q, Li J, Nagayama T, Xu Y, Spencer B F and Li H 2021 The 1st international project competition for structural health monitoring (IPC-SHM, 2020): A summary and benchmark problem *Struct. Health Monit.* **20** 2229-2239

10 Kullaa J 2013 Detection, identification, and quantification of sensor fault in a sensor network *Mech. Syst. Signal Proc.* **40** 208-221

11 Yi T H, Li H N, Song G and Guo Q 2016 Detection of shifts in GPS measurements for a long-span bridge using CUSUM chart *Int. J. Struct. Stab. Dyn.* **16** 1640024

12 Huang H B, Yi T H and Li H N 2017 Bayesian combination of weighted principal-component analysis for

diagnosing sensor faults in structural monitoring systems *J. Eng. Mech.* **143** 04017088
13 Uddin Z, Qamar A and Alam F 2021 ICA based sensors fault diagnosis: an audio separation application *Wirel. Pers. Commun.* **118** 3369-3384
14 Smarsly K and Law K H 2014 Decentralized fault detection and isolation in wireless structural health monitoring systems using analytical redundancy *Adv. Eng. Softw.* **73** 1-10
15 Chang C M, Chou J Y, Tan P and Wang L 2017 A sensor fault detection strategy for structural health monitoring systems *Smart Struct. Syst.* **20** 43-52
16 Hu J, Bian X L, Wei Z B, Li J W and He H W 2022 Residual statistics-based current sensor fault diagnosis for smart battery management *IEEE J. Emerg. Sel. Top. Power Electron.* **10** 2435-2444
17 Javaid A, Javaid N, Wadud Z, Saba T, Sheta O E, Saleem M Q and Alzahrani M E 2019 Machine learning algorithms and fault detection for improved belief function based decision fusion in wireless sensor networks *Sensors* **19** 1334
18 Dragos K and Smarsly K 2016 Distributed adaptive diagnosis of sensor faults using structural response data *Smart Mater. Struct.* **25** 105019
19 Jan S U, Lee Y D and Koo I S 2021 A distributed sensor-fault detection and diagnosis framework using machine learning *Inf. Sci.* **547** 777-796
20 Kiranyaz S, Avci O, Abdeljaber O, Ince T, Gabbouj M and Inman D 2021 1D convolutional neural networks and applications: a survey *Mech. Syst. Signal Proc.* **151** 107398
21 Pandey P, Akshay R and Mira M 2022 Explainable 1-D convolutional neural network for damage detection using Lamb wave *Mech. Syst. Signal Proc.* **164** 108220
22 Zhu R, Peng W W, Wang D and Huang C G 2023 Bayesian transfer learning with active querying for intelligent cross-machine fault prognosis under limited data *Mech. Syst. Signal Proc.* **183** 109628
23 Zhu X M and Li S X 2022 Mgml: momentum group meta-learning for few-shot image classification *Neurocomputing* **514** 351-361
24 Deng J J, Pan Y W, Yao T, Zhou W G, Li H Q and Mei T 2021 Minet: meta-learning instance identifiers for video object detection *IEEE Trans. Image Process.* **30** 6879-6891
25 Su H, Xiang L, Hu A, Xu Y and Yang X 2022 A novel method based on meta-learning for bearing fault diagnosis with small sample learning under different working conditions *Mech. Syst. Signal Proc.* **169** 108765
26 Finn C, Abbeel P, Levine S 2017 Model-agnostic meta-learning for fast adaptation of deep networks *Int. Conf. Mach. Learn.* **70** 1126-1135
27 Huang H B, Yi T H and Li H N 2017 Sensor fault diagnosis for structural health monitoring based on statistical hypothesis test and missing variable approach *J. Aerosp. Eng.* **30** B4015003
28 Huang H B, Yi T H and Li H N 2016 Canonical correlation analysis based fault diagnosis method for structural monitoring sensor networks *Smart Struct. Syst.* **17** 1031-1053
29 Sharifi R and Langari R 2013 Sensor fault diagnosis with a probabilistic decision process *Mech. Syst. Signal Proc.* **34** 146-155
30 Kullaa J 2010 Sensor validation using minimum mean square error estimation *Mech. Syst. Signal Proc.* **24** 1444-1457
31 Ma Y F, He Y, Wang L and Zhang J R 2022 Probabilistic reconstruction for spatiotemporal sensor data

integrated with Gaussian process regression *Probab. Eng. Mech.* **69** 103264
32 Zhang J, Sun Y, Gou L, Gao H, Hong X and Song H 2020 A new bearing fault diagnosis method based on modified convolutional neural networks *Chin. J. Aeronaut.* **33** 439-447
33 Huang K, Yuen K V, Wang L, Jiang T Y and Dai L Z 2022 Sensor fault detection, localization, and reconstruction for online structural identification *Struct. Control. Health Monit.* e2925
34 Ioffe S and Szegedy C 2015 Batch normalization: Accelerating deep network training by reducing internal covariate shift *arXiv* 1502.03167
35 Zhang W, Peng G, Li C, Chen Y and Zang Z 2017 A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals *Sensors* **17** 425
36 Zhang W, Li C, Peng G, Chen Y and Zhang Z 2018 A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load *Mech. Syst. Signal Proc.* **100** 439-453
37 Bergstra J and Bengio Y 2012 Random search for hyper-parameter optimization *J. Mach. Learn.* **13** 281-305
38 Sun Q, Liu Y, Chua T S and Schiele B 2019 Meta-transfer learning for few-shot learning *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* 403-412
39 Prechelt L 2022 Neural networks: Tricks of the trade *Lecture Notes in Computer Science (Early Stopping - But When?)* ed G Montavon et al (Berlin: Springer) p 55-69
40 Sim J and Reid N 1999 Statistical inference by confidence intervals: issues of interpretation and utilization *Phys. Therapy.* **79** 186-195
41 Civera M and Surace C 2021 A Comparative analysis of signal decomposition techniques for structural health monitoring on an experimental benchmark *Sensors* **21** 1825
42 Xia Y, Ni Y Q, Ko J M and Chen H B 2008 Development of a structural health monitoring benchmark problem for high rise slender structures *Adv. Sci. Technol.* **56** 489-494.