Wen-Jie Jiang, C.W. Kim, Y. Goi, F. Zhang: Modeling variability in vibration-based long-term SHM of bridges, ASCE-ASME J of Risk and Reliability in Engineering Systems, Part A: Civil Engineering, 8(1), 2022 (https://doi.org/10.1061/AJRUA6.0001203).

Data normalization and anomaly detection in a steel plate-girder bridge using LSTM

Wen-Jie Jiang[1], Chul-Woo Kim[2], Yoshinao Goi[3] and Feng-Liang Zhang[4]

Abstract: Modal properties are recognized as indicators reflecting structural condition in structural health monitoring (SHM). However, changing environmental and operational variables (EOVs) cause variability in the identified modal parameters and subsequently obscure damage effects. To address the issue caused by the EOVs-related variability, this study investigated the variability of modal frequencies in long-term SHM of a steel plate-girder bridge. A Bayesian fast Fourier transform (FFT) method was used for operational modal analysis in a probabilistic viewpoint. Bayesian linear regression (BLR) and Gaussian process regression (GPR) models were utilized to capture the variability in the identified most probable values (MPVs) of modal frequencies as temperature-driven models, and the limitation of these models for data normalization with latent EOVs was discussed. To overcome the interference of latent EOVs indirectly, a long short-term memory (LSTM) network was established to trace the variability as an autocorrelated process, with a traditional seasonal autoregressive integrated moving average (SARIMA) model as a benchmark. Finally, an anomaly detection method based on residuals of one-step ahead predictions by LSTM was proposed associating with the Mann–Whitney *U* test.

Keywords: anomaly detection, EOVs, fast Bayesian FFT method, LSTM, SARIMA

Introduction

Modal parameters of a dynamic system of structures have been regarded as indicators reflecting structural integrity and widely utilized in the research of vibration-based long-term SHM (Ralbovsky et al. 2010; Fan and Qiao 2011). In modal parameter identification for bridge structures, operational modal analysis (OMA) which claims output-only system identification with ambient vibration data is a practical way of acquiring structural modal parameters. The accuracy of identified modal properties might yet be affected because of the absence of excitation information and the low amplitude of ambient vibration. A series of Bayesian OMA approaches including Bayesian spectral density approach, Bayesian time domain approach, and Bayesian fast Fourier transform (FFT) approach have been developed and widely used with a capacity of quantifying uncertainty in system identification (Yuen and Katafygiotis 2001 a, b and 2003; Au et al. 2013; Au 2017). In the framework of Bayesian OMA, uncertainties are classified into two groups: identification uncertainty, which indicates the uncertainty in a posterior estimation and can be denoted by the posterior covariance matrix of modal parameters; and variability, which mainly represents the variation of modal parameters related with the

[1] PhD candidate, Department of Civil & Earth Resources Engineering, Graduate School of Engineering, Kyoto University, Kyoto 615-8540, Japan. Email: jiang.wenjie.47n@st.kyoto-u.ac.jp
[2] Professor, Department of Civil & Earth Resources Engineering, Graduate School of Engineering, Kyoto University, Kyoto 615-8540, Japan. Email: kim.chulwoo.5u@kyoto-u.ac.jp (corresponding author)
[3] Assistant Professor, Department of Civil & Earth Resources Engineering, Graduate School of Engineering, Kyoto University, Kyoto 615-8540, Japan. Email: goi.yoshinao.2r@kyoto-u.ac.jp
[4] Professor, School of Civil and Environmental Engineering, Harbin Institute of Technology, Shenzhen, China. Email: zhangfengliang@hit.edu.cn

environmental variations. Uncertainty laws have been clarified to manage identification uncertainty in Bayesian OMA (Au 2017). However, practical application of vibration-based SHM remains restricted because of the remaining variability, which exists in long-term SHM and might be inferred as a combined effect of various EOVs, such as temperature, traffic, humidity, and wind.

Results of field studies have shown that such variability can mask changes in modal parameters caused by damage or deterioration, and further reduce accuracy of structural condition assessment (Sohn 2007). Comanducci et al. (1999) used an analytical model of a suspension bridge to study wind loading effects on modal frequencies compared with damage effects. They demonstrated that frequency variations caused by changing wind speed can be more significant than those produced by slight damage. Zhou and Sun (2019) investigated the periodical variability of modal frequencies in a sea-crossing bridge as correlated to EOVs including temperature, traffic, and wind. They reported the dominant EOVs with specific correlation patterns to modal frequencies in different time scales. Therefore, for the long-term SHM of bridge structures, consideration of the variability of modal frequency represents an important issue.

Data normalization research, as a crucial part in practical SHM, aims to normalize the data by modeling and removing the EOVs-related variability in the data, which will further improve the damage sensitivity in anomaly detection. Generally, major studies can be divided into two groups: EOVs-driven and non-EOVs methods. The EOVs-driven method is a supervised way with measured EOVs. Generally speaking, it specifically emphasizes the correlation between measured EOVs and modal frequency. Subsequently, a model connecting modal frequency and measured EOVs can be constructed to catch the correlation and used for anomaly detection. Nandan and Singh (2014) introduced a state space model-based approach to investigate correlation between the modal frequency and temperature, and proposed two kinds of filtering methods to remove seasonal trend in observations. Kim et al. (2018) proposed a Bayesian approach considering multiple factors such as temperature and vehicle weights in a long-term SHM on a Gerber-type steel plate girder bridge. Avendaño-Valencia and Chatzi (2020) combined a Gaussian process with a vector autoregressive model in different time-scales to model the variation of structural dynamics under varying wind speeds and ambient temperatures.

Nevertheless, because monitoring all significant EOVs in SHM campaign is quite difficult, it might engender a latent variable issue in the EOVs-driven methods. This issue sometimes makes the residual of model a nonstationary process, and affects the efficacy of the model-based prediction and anomaly detection. Alternatively, it has been recommended that non-EOVs methods can be used to avoid such issues (Deraemaeker 2018; Zhou and Sun 2019). Non-EOVs methods work in a different path to tackle issues of variability without involving partially measured EOVs, and can be treated in an unsupervised way, such as principal component analysis (PCA), cointegration-based methods, and autoregressive methods. Sen et al. (2019) verified the effectiveness of the PCA for decoupling structural damage and environmental effects in different bridge structures. Liang et al. (2018) used a frequency cointegration-based damage detection method to find a robust co-integration relation between modal frequencies which is insensitive to EOVs, and demonstrated the feasibility of the proposed method by case studies. Xin et al. (2018) applied a composite Kalman-ARIMA-GARCH model to simulate the variability of deformation data collected from a medium span bridge, and observed the existence of heteroscedasticity in the long-term record of deformation data under operational condition.

Compared with traditional methods, long short-term memory (LSTM) (Horchreiter and Schmidhuber 1997; Gers et al. 2000), as a deep learning technique which has been used in many fields like speech recognition (Zia and Zahid 2019), handwriting recognition (Graves et al. 2009) and time series prediction (Wierstra et al. 2005), may also possess the ability of data normalization in addressing the variability of damage-sensitive features like modal frequency in long-term SHM, while related investigations about the application to the SHM of civil structure are still scarce. Therefore, this study is intended to investigate the efficacy of LSTM in data normalization of modal frequencies in the SHM of a steel plate-girder bridge comparing with some traditional methods, and to propose a residual-based anomaly detection approach. Long-term monitoring data for an in-service steel plate-girder bridge are considered. A Bayesian FFT system identification method (Au et al. 2013; Au 2017) is introduced to identify the modal parameters of the in-service steel plate-girder bridge and quantify the identification uncertainty. First, the thermal effect to modal frequencies is investigated using temperature-driven models considering the measured temperature as an EOV. The limitation of these models for data normalization in the target bridge is discussed. Then, the performances of non-EOVs methods in an autoregressive way by the means of seasonal autoregressive integrated moving average (SARIMA) model (Box et al. 2015) and LSTM network are investigated. With the SARIMA model selected as a benchmark model, a vanilla LSTM neural network, as a deep learning model with the ability for sequence feature learning and prediction (Wu et al. 2018; Guo et al. 2020), is used to trace and remove the variability in the identified modal frequency. Finally, the validity of anomaly detection based on the normalized data, i.e. residuals of the two non-EOVs methods, is verified using the Mann-Whitney *U* test. A flowchart for the proposed anomaly detection is presented in Fig. 1.

Bayesian FFT for OMA

Most probable value

The Bayesian FFT method is a frequency domain system identification method (Au et al. 2013; Au 2017) that associates Bayesian inference and FFTs of vibration response with a basic form as

$$p\left(\boldsymbol{\theta}\middle|\left\{\hat{F}_k\right\}\right)=\frac{p\left(\left\{\hat{F}_k\right\}\middle|\boldsymbol{\theta}\right)p(\boldsymbol{\theta})}{p\left(\left\{\hat{F}_k\right\}\right)}, \tag{1}$$

where $\boldsymbol{\theta}$ denotes a set of modal parameters to be identified; $\hat{F}_k$ is the estimated FFTs at corresponding frequency f_k, while $\{\hat{F}_k\}$ denotes a set of $\hat{F}_k$ in a selected frequency band.

Without loss of generality, given non-informative prior, it can be further denoted as Eq. (2).

$$p\left(\boldsymbol{\theta}\middle|\left\{\hat{F}_k\right\}\right)\propto p\left(\left\{\hat{F}_k\right\}\middle|\boldsymbol{\theta}\right) \tag{2}$$

Then, given that the FFTs are Gaussian and independent at different frequencies, it can be written as

$$p\left(\boldsymbol{\theta}\middle|\left\{\hat{F}_k\right\}\right)\propto p\left(\left\{\hat{F}_k\right\}\middle|\boldsymbol{\theta}\right)=\frac{\pi^{-nN_f}}{\prod_k\left|E_k(\boldsymbol{\theta})\right|}\exp\left[-\sum_k\hat{F}_k^{*}E_k(\boldsymbol{\theta})^{-1}\hat{F}_k\right]=e^{-L(\boldsymbol{\theta})}, \tag{3}$$

where

$$E_k(\boldsymbol{\theta})=E\left[F_kF_k^{*}\middle|\boldsymbol{\theta}\right]+E\left[\varepsilon_k\varepsilon_k^{*}\middle|\boldsymbol{\theta}\right]=\sum_{i=1}^{m}\sum_{j=1}^{m}h_{ik}h_{jk}^{*}S_{ij}\boldsymbol{\varphi}_i\boldsymbol{\varphi}_j^{T}+S_e\boldsymbol{I}_n \tag{4}$$

is the theoretical power spectral density (PSD) matrix of data at the k^{th} FFT for given $\boldsymbol{\theta}$. Here, the set of modal parameters $\boldsymbol{\theta}$ comprises modal frequencies $\{f_i\}_{i=1}^m$ and modal damping ratios $\{\zeta_i\}_{i=1}^m$ denoted in transfer functions $\{h_{ik}\}_{i=1}^m$ corresponding to each mode, partial mode shapes $\{\boldsymbol{\varphi}_i\}_{i=1}^m$, PSD matrix of modal forces $\boldsymbol{S}=\left[S_{ij}\right]_{m\times m}$, and the PSD matrix of prediction errors $S_e\boldsymbol{I}_n$. In addition, m represents the number of dominant modes in a specified frequency band where the estimation is conducted. n is the number of sensors to collect the ambient vibration response. N_f is the number of FFT points in the specified frequency band. System parameters $\boldsymbol{\theta}$ can be estimated with an objective function shown in Eq. (5), which is the 'negative log-likelihood function' (NLLF), and the most probable value (MPV) $\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} L(\boldsymbol{\theta})$.

$$L(\boldsymbol{\theta}) = nN_f \ln \pi + \sum_k \ln\left|E_k(\boldsymbol{\theta})\right| + \sum_k \hat{F}_k^* E_k(\boldsymbol{\theta})^{-1} \hat{F}_k \tag{5}$$

Identification uncertainty

The identification uncertainly is one of the major contributions in Bayesian FFT method since it is a fully Bayesian approach. The posterior probability density function (PDF) of the system parameter of the structure $p(\boldsymbol{\theta}|\hat{F}_k)$ can also show the identification uncertainty. It has been demonstrated that the posterior PDF can be well approximated as a Gaussian form for typical data size. Au (2017) showed that a second-order Taylor approximation of the NLLF at the MPV engenders a Gaussian approximation of the posterior PDF as

$$p\left(\boldsymbol{\theta}\middle|\left\{\hat{F}_k\right\}\right) \approx (2\pi)^{-n_{\boldsymbol{\theta}}/2}\left|\hat{\boldsymbol{C}}\right|^{-1/2} \exp\left[-\frac{1}{2}\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}\right)^T \hat{\boldsymbol{C}}^{-1}\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}\right)\right] \tag{6}$$

where $\hat{\boldsymbol{\theta}}$ is the MPV, and $\hat{\boldsymbol{C}}$ denotes the posterior covariance matrix which is the inverse of the Hessian of NLLF at the MPV. Also, $n_{\boldsymbol{\theta}}$ represents the number of parameters in $\boldsymbol{\theta}$. The posterior covariance matrix $\hat{\boldsymbol{C}}$ can provide useful information reflecting the identification uncertainty.

EOVs-driven model

Bayesian linear regression

As a combination of Bayes' theorem and the multiple linear regression (MLR), Bayesian linear regression (BLR) (Marin and Robert 2007; Gelman et al. 2014) can be used as an EOVs-driven model to address variability in the long-term SHM. A basic representation of BLR can be given as

$$\boldsymbol{y} = \boldsymbol{\alpha X} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(0, \sigma_n^2 \boldsymbol{I}_N) \tag{7a}$$

$$p\left(\boldsymbol{\alpha}, \sigma_n^2 | \boldsymbol{X}, \boldsymbol{y}\right) = \frac{p\left(\boldsymbol{y}\middle|\boldsymbol{X}, \boldsymbol{\alpha}, \sigma_n^2\right) p\left(\boldsymbol{\alpha}, \sigma_n^2\right)}{p\left(\boldsymbol{y}\middle|\boldsymbol{X}\right)} \tag{7b}$$

where $\boldsymbol{X} = \{\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_N\}$ and $\boldsymbol{y} = \{y_1, y_2, \cdots, y_N\}$ are N couples of training samples, and denote the predictors and response, respectively; $\boldsymbol{\varepsilon}$ denotes the error term which is assumed to follow a normal distribution denoted as $N(0, \sigma_n^2 \boldsymbol{I}_N)$ with zero mean and covariance $\sigma_n^2 \boldsymbol{I}_N$; $\{\boldsymbol{\alpha}, \sigma_n^2\}$ are the parameters to be estimated or updated, and the form of posterior depends on the assumption of joint prior $p(\boldsymbol{\alpha}, \sigma_n^2)$.

Without loss of generality, a Jeffreys' non-informative joint prior (Jeffreys 1946) can be used as below.

$$p\left(\boldsymbol{\alpha}, \sigma_n^2\right) \propto \frac{1}{\sigma_n^2} \tag{8}$$

Then, the marginal posterior can be written as

$$p\left(\boldsymbol{\alpha} \mid \boldsymbol{y}, \boldsymbol{X}\right) = t_{loc-scale}\left(\hat{\boldsymbol{\alpha}}, \frac{\left(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\alpha}}\right)^{\mathrm{T}}\left(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\alpha}}\right)\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right)^{-1}}{N-q-1}, N-q-1\right) \tag{9a}$$

$$p\left(\sigma_n^2 \mid \boldsymbol{y}, \boldsymbol{X}\right) = \Gamma^{-1}\left(\frac{N-q-1}{2}, \left[\frac{1}{2}\left(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\alpha}}\right)^{\mathrm{T}}\left(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\alpha}}\right)\right]^{-1}\right) \tag{9b}$$

where q represents the number of predictors and $q+1$ is equal to the number of regression coefficients $\boldsymbol{\alpha}$ including an intercept. $\hat{\boldsymbol{\alpha}} = (\boldsymbol{X}^{\mathrm{T}}, \boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y}$ denotes the least squares estimate of $\boldsymbol{\alpha}$.

Gaussian process regression

When the linear correlation is not strong because of the disturbance of latent variables, using a non-parametric model is more flexible. Gaussian process regression (GPR), which is a nonparametric and fully Bayesian approach (Gelman et al. 2014), offers one way to address the issue of an unclear model form in a function space.

The basic assumption of the model is presented in Eq. (10).

$$\boldsymbol{y} = f(\boldsymbol{X}) + \boldsymbol{\varepsilon},\ f(\boldsymbol{X}) \sim GP\left(0, k\left(\boldsymbol{X}_i, \boldsymbol{X}_j\right)\right),\ \boldsymbol{\varepsilon} \sim N\left(0, \sigma_n^2 \boldsymbol{I}_N\right) \tag{10}$$

where $\boldsymbol{X} = \{\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_N\}$ and $\boldsymbol{y} = \{y_1, y_2, \cdots, y_N\}$ are N couples of training samples; $f(\boldsymbol{X})$ is a latent function which follows a Gaussian process; $k(\boldsymbol{X}_i, \boldsymbol{X}_j)$ is a covariance function (a kernel function) specifying the covariance between $f(\boldsymbol{X}_i)$ and $f(\boldsymbol{X}_j)$ at any two points $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$ (i and j can be the same) in the process; $\boldsymbol{\varepsilon}$ denotes the residual term which follows normal distribution and i.i.d. The form of covariance function and hyper-parameters $\boldsymbol{\eta}$ including kernel parameters and σ_n^2 defines the mapping from $\boldsymbol{X}$ to $\boldsymbol{y}$ in the function space.

The selection of kernel function depends on the feature of data itself. As a generally used kernel function, radial basis function kernel (RBF kernel), which has a ready interpretation as a similarity measure related with the squared Euclidean distance of two feature vectors in input space, is adopted. Eq. (11) shows an RBF kernel.

$$k\left(\boldsymbol{X}_i, \boldsymbol{X}_j\right) = \sigma_f^2 \exp\left(-\frac{\left|\boldsymbol{X}_i - \boldsymbol{X}_j\right|^2}{2l^2}\right) \tag{11}$$

where σ_f^2 and l are kernel parameters which denote width and characteristic length-scale, respectively.

A key step in the GPR is to learn the hyper-parameters with the training dataset $\boldsymbol{X}$ and $\boldsymbol{y}$. Generally, hyper-parameter estimation can be done by methods of two kinds (Bachoc 2013): Maximum Likelihood Estimation (MLE) (as shown in Eq. (12a)), and Maximum Pseudo-likelihood Estimation (MPLE) (as shown in Eq. (12b)) which combines a cross-validation in a leave-one-out way and is regarded to perform better when there is weak confidence for the model assumptions.

$$\hat{\boldsymbol{\eta}} = \arg\max_{\eta} \log p\left(\boldsymbol{y} \middle| \boldsymbol{X}, \boldsymbol{\eta}\right) \tag{12a}$$

$$\hat{\boldsymbol{\eta}} = \arg\max_{\eta} \sum_{i=1}^{N} \log p\left(y_i \middle| \boldsymbol{X}, \boldsymbol{y}_{-i}, \boldsymbol{\eta}\right) \tag{12b}$$

Therein, $\boldsymbol{\eta}$ includes the hyper-parameters $\{\sigma_f^2, l, \sigma_n^2\}$, $\boldsymbol{y}_{-i}$ denotes the dataset without y_i, and N represents the number of observations.

Based on the property by which the conditional distribution of a multi-dimensional joint Gaussian distribution is still Gaussian, a posterior predictive distribution on new predictors $\boldsymbol{X}_*$ can be given as follows.

$$p\left(\boldsymbol{f}_* \middle| \boldsymbol{X}, \boldsymbol{y}, \boldsymbol{X}_*, \sigma_n^2\right) = N\left(\bar{\boldsymbol{f}}_*, \mathrm{cov}(\boldsymbol{f}_*)\right) \tag{13a}$$

$$\bar{\boldsymbol{f}}_* = \boldsymbol{K}(\boldsymbol{X}_*, \boldsymbol{X})\left[\boldsymbol{K}(\boldsymbol{X}, \boldsymbol{X}) + \sigma_n^2 \boldsymbol{I}\right]^{-1} \boldsymbol{y} \tag{13b}$$

$$\mathrm{cov}(\boldsymbol{f}_*) = \boldsymbol{K}(\boldsymbol{X}_*, \boldsymbol{X}_*) - \boldsymbol{K}(\boldsymbol{X}_*, \boldsymbol{X})\left[\boldsymbol{K}(\boldsymbol{X}, \boldsymbol{X}) + \sigma_n^2 \boldsymbol{I}\right]^{-1} \boldsymbol{K}(\boldsymbol{X}, \boldsymbol{X}_*) \tag{13c}$$

where $\boldsymbol{f}_* = f(\boldsymbol{X}_*)$, $\boldsymbol{K}$ denotes a kernel matrix with entries $k_{ij} = k(\boldsymbol{X}_i, \boldsymbol{X}_j)$ and $\boldsymbol{I}$ is an identity matrix.

Generalized autoregressive model

As previously mentioned, deficient measurement of the predictor variables might sometimes lead to a non-stationary residual process in EOVs-driven models, which severely undermines the reliability of residual-based anomaly detection. This section introduces two popular generalized autoregressive models for variability research in a non-EOVs way: SARIMA and LSTM.

Seasonal Autoregressive Integrated Moving Average (SARIMA) model

The SARIMA model is a classical linear time-series model which is able to track the variability of data with trend, seasonality and random components (Box et al. 2015). A general multiplicative representation of a $\mathrm{SARIMA}(p,d,q)\times(P,D,Q)_s$ model can be written as Eq. (14).

$$\phi_p(B)\Phi_P\left(B^s\right)\nabla^d\nabla_s^D y_t=\theta_q(B)\Theta_Q\left(B^s\right)\varepsilon_t \tag{14}$$

Therein, B denotes the backward shift operator with $By_t=y_{t-1}$. Also, s stands for the period of seasonality, $\nabla=1-\mathrm{B}$ is the difference operator, and $\nabla_s=1-B^s$ is the seasonal difference operator. The polynomials $\phi_p(\cdot)$, $\Phi_P(\cdot)$, $\theta_q(\cdot)$ and $\Theta_Q(\cdot)$ are ordered respectively as p, P, q and Q with unknown coefficients. Also, d and D respectively represent the order for differencing and seasonal differencing. Innovation ε_t is generally assumed to follow a Gaussian or Student's t-distribution.

A plot involving autocorrelation function (ACF) and partial autocorrelation function (PACF) can be referred to specify a proper set initially for the order parameters of a SARIMA model. Furthermore, some indices such as the Akaike information criterion (AIC) and Bayesian information criterion (BIC) are useful to assess the optimal structure of the model. Generally speaking, model order parameters with small AIC and BIC can be adopted. Given k as the number of coefficients to be estimated, n as the sample size and $\hat{L}$ as the maximum likelihood, AIC and BIC can be estimated using Eq. (15).

$$\mathrm{AIC}=2k-2\ln\left(\hat{L}\right) \tag{15a}$$

$$\mathrm{BIC}=k\ln(n)-2\ln\left(\hat{L}\right) \tag{15b}$$

Long Short-Term Memory (LSTM)

Classical time series models including SARIMA have been used widely. However, there are still some defects in the classical time series models. A linear form might restrict the distribution of the predicted response and engender heteroscedasticity in the residuals. Consideration of finite order in the time series model as an approximation might affect the performance of residuals. In that case, a nonlinear neural network with an LSTM layer (Horchreiter and Schmidhuber 1997; Gers et al. 2000) might be a good substitute for SARIMA.

An LSTM neural network is a kind of Recurrent Neural Network (RNN) that is able to learn the long-term dependencies in time series data. In traditional RNN, learning long-term dependencies practically is difficult because of the computation issue known as gradient vanishing and exploding in the backpropagation through time (BPTT) algorithm. The LSTM network solves this issue well by introducing a cell state and some special gates to forget and update information from each time step.

A common LSTM network for time series regression issue has a layer structure as presented in Fig. 2(a). Compared with traditional feedforward neural networks like multilayer perceptron (MLP) or convolutional neural network (CNN), the major difference is a recurrent structure in LSTM layer. A schematic diagram for the local details of a vanilla LSTM layer (Gers et al. 2000; Wu et al. 2018; Guo et al. 2020) is presented in Fig. 2(b). The vertical direction shows the data flow in time, whereas the horizontal direction shows the data flow from input to output. The architecture is composed mainly of a cell and three gates, respectively called a forget gate, input gate, and output gate. The cell is used to track and record the dependencies in the time series, whereas the gates regulate the information flow into or out of the cell. The detailed forward pass operation in an LSTM layer can be described as follows referring to reports by Horchreiter and Schmidhuber (1997), Gers et al. (2000), and Graves (2012).

$$\boldsymbol{f}_t = \sigma_g\left(\boldsymbol{W}_f \boldsymbol{X}_t + \boldsymbol{U}_f \boldsymbol{h}_{t-1} + \boldsymbol{b}_f\right) \tag{16a}$$

$$\boldsymbol{i}_t = \sigma_g\left(\boldsymbol{W}_i \boldsymbol{X}_t + \boldsymbol{U}_i \boldsymbol{h}_{t-1} + \boldsymbol{b}_i\right) \tag{16b}$$

$$\boldsymbol{o}_t = \sigma_g\left(\boldsymbol{W}_o \boldsymbol{X}_t + \boldsymbol{U}_o \boldsymbol{h}_{t-1} + \boldsymbol{b}_o\right) \tag{16c}$$

$$\tilde{\boldsymbol{c}}_t = \sigma_c\left(\boldsymbol{W}_c \boldsymbol{X}_t + \boldsymbol{U}_c \boldsymbol{h}_{t-1} + \boldsymbol{b}_c\right) \tag{16d}$$

$$\boldsymbol{c}_t = \boldsymbol{f}_t \circ \boldsymbol{c}_{t-1} + \boldsymbol{i}_t \circ \tilde{\boldsymbol{c}}_t \tag{16e}$$

$$\boldsymbol{h}_t = \boldsymbol{o}_t \circ \sigma_c\left(\boldsymbol{c}_t\right) \tag{16f}$$

where the operator $\circ$ denotes the Hadamard product; $\boldsymbol{X}_t \in R^{d\times 1}$ is the input vector with d features; $\boldsymbol{f}_t \in R^{h\times 1}$, $\boldsymbol{i}_t \in R^{h\times 1}$, and $\boldsymbol{o}_t \in R^{h\times 1}$ are the output of forget gate, input gate, and output gate, respectively with h dimensions which equals to the number of hidden units in LSTM layer; $\boldsymbol{h}_t \in R^{h\times 1}$ is the hidden state vector which is delivered to both output and next step; $\tilde{\boldsymbol{c}}_t \in R^{h\times 1}$ is the candidate cell state input; $\boldsymbol{c}_t \in R^{h\times 1}$ is the cell state vector which records long-term dependency; $\boldsymbol{W} \in R^{h\times d}$, $\boldsymbol{U} \in R^{h\times h}$ and $\boldsymbol{b} \in R^{h\times 1}$ are input weight matrix, recurrent weight matrix and bias vector to be learned, respectively; σ_g is the gate activation function which is a sigmoid function, while σ_c denotes the state activation function generally using a hyperbolic tangent function.

For a one-step ahead prediction case, the learned correlation between sequential input $\{\boldsymbol{X}_i\}_{i=2}^{t} = \{\boldsymbol{y}_i\}_{i=1}^{t-1}$ and one-step ahead output $\boldsymbol{y}_t$ can be finally written as Eq. (17).

$$\boldsymbol{y}_t = \hat{\boldsymbol{y}}_t + \boldsymbol{\varepsilon}_t = \varphi\left(\left\{\boldsymbol{y}_i\right\}_{i=1}^{t-1}, \boldsymbol{h}_0, \boldsymbol{c}_0\right) + \boldsymbol{\varepsilon}_t \tag{17}$$

which can be regarded as a nonlinear and no-cut-off autoregressive model, with initial states $\boldsymbol{h}_0$ and $\boldsymbol{c}_0$ generally set to be 0 as two hyper-parameters. $\varphi(\cdot)$ denotes a nonlinear function displaying the long-term correlation between one-step ahead prediction $\hat{\boldsymbol{y}}_t$ and all the past observations $\{\boldsymbol{y}_i\}_{i=1}^{t-1}$. $\boldsymbol{\varepsilon}_t$ is the residual term between the observation $\boldsymbol{y}_t$ and the prediction $\hat{\boldsymbol{y}}_t$.

Residual-based anomaly detection

From the viewpoint of classical statistics and in long-term SHM, the MPV of the modal frequency obtained using the Bayesian FFT at a certain time t is a maximum likelihood estimator correlated with the estimators of FFTs, $\{\hat{F}_k\}_t$, which comply with independent complex Gaussian distributions. This arrangement leads to treatment of MPV (denoted as $\hat{f}_t$) as a random variable with a naive form such as Eq. (18).

$$\hat{f}_t = G\left(\left\{\hat{F}_k\right\}_t\right) = \underbrace{f_t}_{EOVs-related} + \underbrace{\varepsilon_t}_{noise\&approx.} \tag{18}$$

Therein, $G(\cdot)$ is a connection function. f_t denotes the real value (fixed but unknown) of modal frequency at time t and is EOVs-related; ε_t is the random error term which mainly contains the influence from noise of vibration data and approximation in identification algorithm, and is generally independent.

As shown in Eq. (18), the total variability of identified MPV $\hat{f}_t$ in long-term SHM comprises the

EOVs-related variability in f_t, and identification uncertainty in ε_t which is mainly caused by noise in vibration data as well as approximations in identification algorithms and is generally independent. Consequently, theoretically speaking, if a model that ideally captures the EOVs-related variability of modal frequency in long-term SHM can be formulated, the residuals of the model can be regarded as a component after removing the EOVs-related variability, and might include most effects of noise and approximations in system identification. In that case, it can be expected that the residuals get close to a white noise process or at least a stationary process regardless of heteroscedasticity in some cases, and a change in the features of residual distribution might indicate an anomalous state of structure.

Above discussion can be regarded as an interpretation of data normalization in SHM from a probabilistic viewpoint with various sources of uncertainties. Therefore, by comparing the introduced four models on modeling the EOVs-related variability and data normalization, the efficacy of LSTM for data normalization and residual-based anomaly detection can be demonstrated.

Case study

Background

A case study is conducted using monitoring data collected from a simply supported steel plate-girder bridge in Japan [see Fig. 3]. The target bridge was constructed in 1957 with span length of 40.5 m and width of 4.5 m. Vehicle was the major excitation to the bridge from the view of strength, while other EOVs such as wind occupied the major duration of ambient excitation. Long-term health monitoring was implemented for this bridge with duration of about half a year before it is removed. As portrayed in Fig. 4, sensors are deployed mainly to the two steel girders to measure bridge responses including acceleration, strain and displacement along with temperature. The vertical acceleration data collected by accelerometers (A1–A10) are taken to reflect the vibration response and used for operational modal analysis. Temperature records are denoted as T_1, T_2, T_3, T_4, and T_5, where T_5 represents the air temperature.

Operational modal analysis

Operational modal analysis is conducted with acceleration records from 10 channels based on the Fast Bayesian FFT algorithm. The signal length for the analysis was set as 0.5 h, whereas the identification of modal properties was conducted per 0.5 h in the long-term SHM. To apply the Bayesian FFT algorithm, frequency bands within which the modal frequencies probably exist, the number of modes, and the corresponding initialization set should be previously specified. The frequency band generally can be determined with a singular value (SV) spectrum inferred from the cross power spectral density (CPSD) matrix (Au 2017). The SV spectrum is a plot of the eigenvalues of the real part of CPSD matrix versus frequency. Each spectral line in an SV spectrum denotes a certain-order eigenvalue versus frequency. Because correlation among signal channels is not involved in a PSD spectrum, it is difficult to ascertain the number of modes around a peak displaying dynamic amplification, especially when some close modes exist. In that case, the amount of dominating eigenvalues (lines) around a peak in the SV spectrum might indicate the number of close modes around the peak.

As presented in Fig. 5, both the PSD and SV spectrum provide the potential resonance bands of the bridge, while in the SV spectrum, the number of lines (eigenvalues) significantly above the remaining ones around a peak represents the number of modes around the corresponding frequency. The captions denoted in the SV spectrum show the identified modes. The identified 'partial' mode shapes (confined to the measured DOFs) are depicted in Fig. 6. The 2nd* bending mode is named for its high similarity to the 2nd bending mode with measured DOFs.

As previously described, Bayesian operational modal analysis including the Bayesian FFT provides information to quantify the uncertainty in identification such as posterior variance and the signal-to-noise ratio (SNR). The posterior variance and SNR at each modal frequency are portrayed in Fig. 7. The results suggest a phenomenon of ascending identification uncertainty and rapidly descending SNR with respect to the mode order. It is noteworthy that bending modes tend to possess higher SNR than torsional modes. These results provide valuable information for selecting indicators as response variables in later studies of long-term variability. In terms of data quality, modal frequencies with less identification uncertainty and higher SNR are preferred. A mode with lower SNR does not guarantee stable identification in every dataset. Consequently, modal frequencies of the first and the second bending modes (denoted respectively as f_1 and f_2) are considered. The third bending modal frequency (mode 7 in Fig. 6) denoted as f_3 with the most significant uncertainty and lowest SNR as depicted in Fig. 7 is also selected for comparison. Long-term MPVs of those three bending frequencies are presented in Fig. 8. Each point corresponds to an identification conducted per 0.5 h. The blank segments are attributable to intermittent monitoring. To check the effects of EOVs-driven models in modeling the variability of modal frequencies in long-term SHM, the dataset is divided into two subsets: a training set (left part of the partition line in Fig. 8) and a validation set (right part of the partition line in Fig. 8). To verify if there is interference of multimodality in the estimates of each bending mode in the long-term SHM, Modal Assurance Criterion (MAC) is calculated along with each identification and plotted in Fig. 9. It is noted that most of the estimates lead to a MAC value more than 0.99, which indicated that most of the estimates follow the same mode in long-term SHM and thus are reliable.

Correlation analysis and temperature-driven model

In the monitoring of the target bridge, temperature is the only measured EOVs with records of different locations denoted as T_1, T_2, T_3, T_4, and T_5. Fig. 10 shows that the temperature records from two girders (T_1 to T_4) show more variance than air temperature (T_5). Even between two girders, the temperature variance mutually differs. It is noted that obvious difference of temperature records between two girders occurs around the lower points at night. A possible reason for this phenomenon is inferred as a result of the topography around the target bridge. The bridge lies along north-south direction, with nearby a dam located in east [see Fig. 3]. The girder with records of T_1 and T_2 is also in the east of bridge and just faced with the drain opening of dam. Therefore, it can be inferred a comprehensive effect of wind, water spray and sunshine may make a difference between the two girders on temperature records, in particular for the case at night. Fig. 11 shows a correlation plot including temperature records and the first bending modal frequency, from which higher similarity was observed between T_1 and T_2, as it was between T_3 and T_4. Therefore, to avoid the influence of multicollinearity in regression, T_2, T_4, and T_5 are chosen as predictors in the temperature-driven models. A negative linear relation was observed for the relation between the modal frequency and temperature. However, a large variance was also noted, which is probably attributable to interference from other EOVs such as traffic and humidity.

To get a clearer perception on the correlation, standardized temperature records and modal frequencies about five days were investigated. In Fig. 12, one might observe that the valleys of identified modal frequencies generally appear in daytime with good correspondence to the peaks of temperature records, while the peaks of identified modal frequencies are usually observed around midnight and hours ahead of the valleys of temperature records, which further complicated the research of EOVs-related variability. Generally, thermal inertia in large structure may lead to an inconsistency between temperature records and modal frequencies, but for that case the temperature records from outside usually do not lag behind but go ahead of the modal frequencies. In the target bridge, it is observed that the valleys of temperature records severely lag behind the peaks of modal frequencies, and no obvious lag effect among three temperature records either. Therefore, it is inferred as a result of interference from some latent EOVs such as traffic and humidity. In addition, it is noteworthy that the EOVs-related variability with daily periodicity is heavily scattered in the third bending frequency compared to the first and second bending frequencies because of its high identification uncertainty and low SNR. It indicated the influence of high identification uncertainty in OMA blurred the feature of EOVs-related variability in long-term SHM.

Using data of the training set shown in Fig. 8, temperature-driven models were established. As previously mention, temperature records T_2, T_4, and T_5 were selected as an input vector with a consideration of the distributed thermal effect. The identified MPVs of modal frequencies f_1, f_2 and f_3 were taken as a response vector. In the training of BLR model, a non-informative joint prior of coefficients and error variance is used for posterior estimation which is analytically tractable. Then, the posterior is further used to calculate a predictive distribution for prediction with belief. In the training of GPR model, MPLE is implemented to estimate the hyper-parameters including kernel parameters of an RBF kernel and error variance. The quasi-Newton method is used in the optimization of objective function. Then, with the learned hyper-parameters, a conditional normal distribution is calculated as the posterior predictive distribution for prediction with belief. The prediction results of the validation set presented in Fig. 13 demonstrated that the GPR model performed better than the BLR model, with a capacity of predicting more local variations of the modal frequencies in long-term SHM. When considering the third bending frequency polluted by higher identification uncertainty, the prediction almost failed to reflect the EOVs-related variability which generally has daily periodicity.

In long-term SHM, a well-performed model is needed for tracking EOVs-related variability to reproduce residuals with stationary features, which is the prerequisite in residual-based anomaly detection. Therefore, the residuals of prediction models for the first and the second bending frequencies are investigated comprehensively. As presented in Fig. 14, obvious differences such as biased mean values were observed between the residual distributions on training dataset and validation dataset for each prediction model, which suggests the existence of a non-stationary feature in the residuals and an unsatisfactory generalization capacity of these models. Quantile–quantile plots (Q–Q plots) of the residual distribution are presented in Fig. 15. The plots demonstrated that the residual distribution of each model performed as right-skewed and failed to meet a Gaussian distribution. It might be inferred that interference of some latent EOVs remained in the residuals of temperature-driven models and led to non-stationary and non-Gaussian behavior. Therefore, for the SHM of target bridge with some crucial latent EOVs unmeasured, these temperature-driven models tend to be insufficient for data normalization and residual-based anomaly detection in the target bridge.

Generalized autoregressive model

As the use of temperature-driven models might lead to an issue of latent EOVs effect in residuals, autocorrelation of modal frequencies corresponding to autocorrelation in general EOVs offered another way for data normalization and reducing latent EOVs effects in residuals. It can be expected that the residuals of generalized autoregressive models including SARIMA and LSTM might perform more stationarily in a one-step ahead predictive way. These generalized autoregressive models generally demand a time series without missing data. Therefore, a data segment with few missing data from original intermittent sequence of modal frequencies is used for model establishment after data interpolation using the daily periodicity.

As presented in Fig. 16, the data segment is divided into three subsets: a training dataset, validation dataset, and testing dataset. To conduct residual-based anomaly detection, five simulated anomaly scenarios were introduced into the testing dataset as marked as D1–D5. They denote the following: D1 adds a step signal with amplitude of 0.3 σ to the original testing data; D2 adds a linear slope signal from 0 to 0.6 σ; D3 amplifies the original testing data by a factor of 1.05; D4 is a test for the robustness of the proposed anomaly detection method to white noise with SNR equal to 5; D5 simulates changes in the autocorrelation of the original testing data as a kind of anomaly by adding an AR(1) component to the original testing data. Also, INT denotes an intact scenario, which is the original testing data. The severity of these artificially introduced data anomalies was limited to represent weak anomalies which are difficult to detect intuitively.

A SARIMA model is first constructed as a benchmark model. According to the ACF and PACF of f_1 presented in Fig. 17(a), the seasonal frequency of 48 time steps corresponding to 24 h (= 48 × 0.5 h) in the sequence is observed, which indicates a seasonal component with daily periodicity. Then, a seasonal differencing operation is conducted on the original sequence of f_1, and the obtained seasonal differencing sequence is further investigated. Fig. 17(b) shows the ACF and PACF of the seasonal differencing sequence of f_1, which proposes a seasonal moving average component SMA(1) and an autoregressive component AR(p) while p denotes the order of AR component. The optimal AR model order was decided as p = 3 using both AIC and BIC values of the model as shown in Table 1. Finally, a $\mathrm{SARIMA}(3,0,0) \times (0,1,1)_{48}$ model was used to capture the EOVs-related variability.

As for the vanilla LSTM neural network, which is a popular deep learning technique, the structure of the network and corresponding hyper-parameters are determined after sufficient trials of hyper-parameter adjustment. The network is defined as a univariate one-step ahead predictive structure. Optimal network parameters are learned by stochastic optimization on the objective function, specifically the mean squared error (MSE) with the Adam algorithm (Kingma and Ba 2015). The number of hidden units in the LSTM layer was set as 30. The maximum number of epochs was defined as 250. The initial learning rate was set as 0.001, but it was dropped after half of the epochs with a factor of 0.1. To avoid gradient explosion, the gradient threshold was set as 1. To avoid overfitting, a trick called early stopping was used. The training process is presented in Fig. 18.

To show a more intuitive comparison, BLR and GPR models are also established on the same datasets shown in Fig. 16. The performance of EOVs-related variability compensation and data normalization using the trained SARIMA model and LSTM network is investigated compared with temperature-driven models. Observed from Fig. 19, both SARIMA and LSTM models traced the EOV-related variability well compared with temperature-driven models, while the prediction by the SARIMA model displayed more local variations than the LSTM network. Fig. 20 shows the residual sequences of these four models on both training set and validation set. It can be noted that residual sequences of SARIMA and LSTM tend to be a stationary process, while residual sequences of BLR and GPR still performed differently between training set and validation set with an obvious remaining trend term. To quantitatively reflect the distinction of the four models in explaining the variance of original sequence, coefficient of determination denoted as R^2 is computed in Fig. 21. It is known that a high value of R^2 close to 1 gives a proof of good explanation of data variance, and close R^2 values between training set and validation set implies a good generalization ability of model. Hence, the results shown in Fig. 21 indicated the trained LSTM model made both good interpretation and generalization ability, while SARIMA model resulted in a poor interpretation and temperature-driven models showed a bad generalization. From the residual distributions on training dataset and validation dataset presented in Fig. 22, it is noteworthy that the stationarity of residuals in both models performed much better than those in temperature-driven models, especially the mean values of residual distributions. The trained LSTM network led to better performance than the SARIMA model with less variance in residuals. The ACF and PACF of the residuals presented in Fig. 23 further demonstrated better performance of the LSTM network than that of SARIMA. Fig. 23(a) shows that the residuals of the trained SARIMA model belong to a SMA(1) component which is a stationary but autocorrelated process, while Fig. 23(b) shows that there was almost no autocorrelation in the residuals of the trained LSTM network. An interesting observation from the cumulative distributions of residuals presented in Fig. 24 is that the residual distributions of both models tend to be symmetric but heavy-tailed rather than a Gaussian distribution, while a *t*-location-scale distribution fits well. However, this phenomenon is explainable by an effect of heteroscedasticity which can be found between the daytime and nighttime (e.g. valleys and peaks of modal frequencies in Fig. 12 and Fig. 16) with daily periodicity.

Since both SARIMA and LSTM models produced residuals with un-biased mean values in residual distributions after data normalization, it is expected the mean of residual distribution can be taken as an indicator for anomaly detection with statistical test. Because the residual distributions tend to be symmetric but non-Gaussian as shown in Fig. 24, *t*-test which requests a normal population may lose the efficacy. Therefore, in this study a nonparametric test method called Mann–Whitney *U* test (Mann and Whitney 1947; Siegel 1956) was introduced to ascertain whether significant changes had occurred, or not, in the medians (equal to means in symmetric distributions) of residual distributions from different datasets.

Mann–Whitney *U* test offers a flexible way without any assumption on the distribution type of populations. It tests a null hypothesis H_0 (the samples are from two populations with equal medians), against an alternative H_1 (the samples are from two populations with different medians). For populations that are approximately symmetric, this test is equivalent to a test of the equality of means. The test statistic *U* is defined as

$$U=\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} S\left(X_i,Y_j\right), \text{ with } S\left(X,Y\right)=\begin{cases}1, & \text{if } Y<X,\\ 1/2, & \text{if } Y=X,\\ 0, & \text{if } Y>X.\end{cases} \tag{19}$$

where $\{X_i\}_{i=1}^{n_1}$ and $\{Y_j\}_{j=1}^{n_2}$ respectively represent two sets of samples from two populations. n_1 and n_2 are corresponding sample size.

The distribution of *U* is known under the null hypothesis, and generally tabulated in the case of a small sample size. For a large sample size, it has been demonstrated that *U* is approximately normally distributed, in which case a *z*-test can be used. The standardized test statistic *z* is given as

$$z=\frac{U-\mu_U}{\sigma_U} \tag{20}$$

where μ_U and σ_U are the mean and standard deviation of *U*, respectively, given by

$$\mu_U=\frac{n_1 n_2}{2} \tag{21a}$$

$$\sigma_U=\sqrt{\frac{n_1 n_2\left(n_1+n_2+1\right)}{12}} \tag{21b}$$

Fig. 25 shows the fitted *t*-location-scale distributions of the residuals of two models on different data subsets. It is noteworthy that the mean values of the residual distributions of LSTM network show more distinctions among different scenarios than those in SARIMA model. Based on the statistical test, the residuals of two models are tested between training dataset and testing dataset with six scenarios. Those tested results are summarized in Table 2 for the Mann–Whitney *U* test. Value 1 denotes rejection of null hypothesis, which indicates that significant difference exists between the means of two populations and can be treated as a proof of anomaly occurrence. As shown in the table, the results indicated that LSTM network which is a nonlinear and no-cut-off model tends to be more sensitive to anomalies than a linear and finite-order SARIMA model in the procedure of data normalization. By means of the Mann–Whitney *U* test on the residuals of trained LSTM network, all simulated anomaly scenarios were detected except for the D4 scenario which is defined by introducing additional white noise, whereas this situation can be regarded as evidence for the robustness of the residual-based anomaly detection method associating Mann–Whitney *U* test with the LSTM network.

Conclusions

This study investigated the efficacy of LSTM for data normalization and anomaly detection in a steel plate-girder bridge comparing with some traditional methods, and proposed a novel anomaly detection approach. Modal frequency identified by a fast Bayesian FFT method under ambient vibration is taken as the basic damage-sensitive feature, and the properties of EOVs-related variability in the identified modal frequencies in long-term SHM is elaborated with data normalization methods. Major conclusions can be summarized as follows.

(1) The quantified identification uncertainty and estimated signal-to-noise ratio from the Bayesian FFT method provide useful information when selecting parameters for subsequent long-term SHM.

(2) Among EOVs-driven models, the GPR model performed better than the BLR model in capturing the EOVs-related variability.

(3) Lack of information of dominant EOVs caused a latent variable issue that severely affects the EOVs-driven model performance. The residuals involving the effects of latent EOVs might not comply with the Gaussian assumption of random error and further affect the accuracy of MLE, or become nonstationary and further affect the generalization ability of models.

(4) Under the condition of existence of latent dominant EOVs, both classical SARIMA model and vanilla LSTM network well captured the seasonality and random variability of the modal frequency in long-term SHM.

(5) Results of residual-based anomaly detection demonstrated that the LSTM model considering nonlinearity and long-term correlation is more sensitive to the anomaly which occurs in the pattern of EOVs-related variability of modal frequency compared to the classical SARIMA model.

(6) An anomaly detection method combining the residuals of one-step ahead prediction by LSTM and Mann–Whitney *U* test showed a good performance for detecting anomalies in the long-term SHM of the steel plate-girder bridge.

Data Availability Statement

All data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments

This study was partly sponsored by a Japanese Society for the Promotion of Science (JSPS) Grant-in-Aid for Scientific Research (B) under project No. 19H02225 and National Natural Science Foundation of China (51878484). That financial support is gratefully acknowledged.

References

Au, S. K., F. L. Zhang, and Y. C. Ni. 2013. “Bayesian operational modal analysis: theory, computation, practice.” *Computers and Structures*, 126, 3–14. https://doi.org/10.1016/j.compstruc.2012.12.015.

Au, S. K. 2017. *Operational Modal Analysis: Modeling*, *Bayesian Inference*, *Uncertainty Laws*. Singapore: Springer. https://doi.org/10.1007/978-981-10-4118-1.

Avendaño-Valencia, L. D. and E. N. Chatzi. 2020. “Multivariate GP-VAR models for robust structural identification under operational variability.” *Probab*. *Eng*. *Mech*., 60, 103035. https://doi.org/10.1016/j.probengmech.2020.103035.

Bachoc, F., 2013. “Cross validation and maximum likelihood estimations of hyper-parameters of Gaussian processes with model misspecification.” *Computational Statistics & Data Analysis*, 66, 55–69.

Box, G. E. P., G. M. Jenkins, G. C. Reinsel, and G. M. Ljung. 2015. *Time Series Analysis: Forecasting and Control (Fifth edition)*. Wiley.

Cornwell, P., C. R. Farrar, S. W. Doebling, and H. Sohn. 1999. “Environmental variability of modal properties.” *Experimental Techniques*, 23 (6), 45–48. https://doi.org/10.1111/j.1747-1567.1999.tb01320.x.

Deraemaeker, A. and K. Worden. 2018. “A comparison of linear approaches to filter out environmental effects in structural health monitoring.” *Mech*. *Syst*. *Signal Process*., 105, 1–15.

Fan, W. and P. Qiao. 2011. “Vibration-based damage identification methods: a review and comparative study.” *Struct*. *Health Monit*., 10 (1), 83–111. https://doi.org/10.1016/j.ymssp.2017.11.045.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2014. *Bayesian Data Analysis (Third edition)*. Boca Raton, FL: Chapman and Hall/CRC.

Gers, F. A., J. Schmidhuber, and F. Cummins. 2000. “Learning to Forget: Continual Prediction with LSTM.” *Neural Computation*, 12 (10), 2451–2471. https://doi.org/10.1162/089976600300015015.

Graves, A., M. Liwicki, S. Fernandez, R. Bertolami, H. Bunke and J. Schmidhuber. 2009. “A Novel Connectionist System for Improved Unconstrained Handwriting Recognition.” *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 31 (5): 855–868.

Graves, A. 2012. *Supervised Sequence Labelling with Recurrent Neural Networks*. Berlin Heidelberg: Springer.

Guo, A., A. Jiang, J. Lin, and X. Li. 2020. “Data mining algorithms for bridge health monitoring: Kohonen clustering and LSTM prediction approaches.” *The Journal of Supercomputing*, 76, 932–947. https://doi.org/10.1007/s11227-019-03045-8.

Hochreiter, S. and J. Schmidhuber. 1997. “Long short-term memory.” *Neural Computation*, 9 (8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

Jeffreys, H. 1946. “An Invariant Form for the Prior Probability in Estimation Problems.” *Proceedings of the Royal Society of London*. *Series A*, *Mathematical and Physical Sciences*, 186 (1007), 453–461. https://doi.org/10.1098/rspa.1946.0056. https://doi.org/10.1098/rspa.1946.0056.

Kim, C. W., Y. Zhang, Z. Wang, Y. Oshima, and T. Morita. 2018. “Long-term bridge health monitoring and performance assessment based on a Bayesian approach.” *Struct*. *Infrastruct*. *Eng*., 14 (7), 883–894. https://doi.org/10.1080/15732479.2018.1436572.

Kingma, D. P. and J. L. Ba. 2015. “ADAM: A method for stochastic optimization.” Proc. of the Third International Conference for Learning Representations. San Diego, USA.

Liang, Y., D. Li, G. Song, and Q. Feng. 2018. “Frequency Co-integration-based damage detection for bridges under the influence of environmental temperature variation.” *Measurement*, 125, 163–175. https://doi.org/10.1016/j.measurement.2018.04.034.

Mann, H. B. and D. R. Whitney. 1947. "On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other." *Annals of Mathematical Statistics*, 18 (1), 50–60. https://doi.org/10.1214/aoms/1177730491.

Marin, J. M. and C. P. Robert. Bayesian Core: A Practical Approach to Computational Bayesian Statistics. New York: Springer Science+Business Media, LLC, 2007.

Nandan, H. and M. P. Singh. 2014. "Effects of thermal environment on structural frequencies: Part II – A system identification model." *Eng. Struct.*, 81, 491–498. https://doi.org/10.1016/j.engstruct.2014.07.042.

Ralbovsky, M., S. Deix, and R. Flesch. 2010. "Frequency changes in frequency-based damage identification." *Struct. Infrastruct. Eng.*, 6 (5), 611–619. https://doi.org/10.1080/15732470903068854.

Sen, D., K. Erazo, W. Zhang, S. Nagarajaiah, and L. Sun. 2019. "On the effectiveness of principal component analysis for decoupling structural damage and environmental effects in bridge structures." *J. Sound Vibra.*, 457, 280–298. https://doi.org/10.1016/j.jsv.2019.06.003.

Siegel, S. 1956. *Nonparametric statistics for the behavioral sciences.* McGraw-Hill.

Sohn, H. 2007. "Effects of environmental and operational variability on structural health monitoring." *Philos. Trans. R. Soc. A: Math. Phys. Eng. Sci.*, 365 (1851), 539–560. https://doi.org/10.1098/rsta.2006.1935.

Wierstra, D., J. Schmidhuber and F. J. Gomez. 2005. "Evolino: Hybrid Neuroevolution/Optimal Linear Search for Sequence Learning." Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI), Edinburgh: 853–858.

Wu, Y., M. Yuan, S. Dong, L. Lin, and Y. Liu. 2018. "Remaining useful life estimation of engineered systems using vanilla LSTM neural networks." *Neurocomputing*, 275, 167–179. https://doi.org/10.1016/j.neucom.2017.05.063.

Xin, J., J. Zhou, S. X. Yang, X. Li, and Y. Wang. 2018. "Bridge Structure Deformation Prediction Based on GNSS Data Using Kalman-ARIMA-GARCH Model." *Sensors*, 18 (1), 298. https://doi.org/10.3390/s18010298.

Yuen, K. V. and L. S. Katafygiotis. 2001a. "Bayesian spectral density approach for modal updating using ambient data." *Earthq. Eng. Struct. Dyn.*, 30 (8), 1103–1123. https://doi.org/10.1002/eqe.53.

Yuen, K. V. and L. S. Katafygiotis. 2001b. "Bayesian time-domain approach for modal updating using ambient data." *Probab. Eng. Mech.*, 16 (3), 219–231. https://doi.org/10.1016/S0266-8920(01)00004-2.

Yuen, K. V. and L. S. Katafygiotis. 2003. "Bayesian Fast Fourier Transform approach for modal updating using ambient data." *Adv. Struct. Eng.*, 6 (2), 81–95. https://doi.org/10.1260/136943303769013183.

Zhou, Y. and L. Sun. 2019. "Effects of environmental and operational actions on the modal frequency variations of a sea-crossing bridge: A periodicity perspective." *Mech. Syst. Signal Process.*, 131, 505–523. https://doi.org/10.1016/j.ymssp.2019.05.063.

Zia, T. and U. Zahid. 2019. "Long short-term memory recurrent neural network architectures for Urdu acoustic modeling." *International Journal of Speech Technology*, 22, 21–30. https://doi.org/10.1007/s10772-018-09573-7.

Table 1. AIC and BIC for the SARIMA models with a component AR(p)

Criterion	Order of AR component					
	$p = 1$	$p = 2$	$p = 3$	$p = 4$	$p = 5$	$p = 6$
AIC	-1678.5	-1734.3	-1774.8	-1773.6	-1777.2	-1778.6
BIC	-1658.6	-1709.5	-1745.0	-1738.8	-1737.5	-1734.0

Table 2. Test results based on Mann–Whitney U test

Model	Scenarios of testing data					
	INT	D1	D2	D3	D4	D5
LSTM	0	1	1	1	0	1
SARIMA	0	0	0	0	0	0

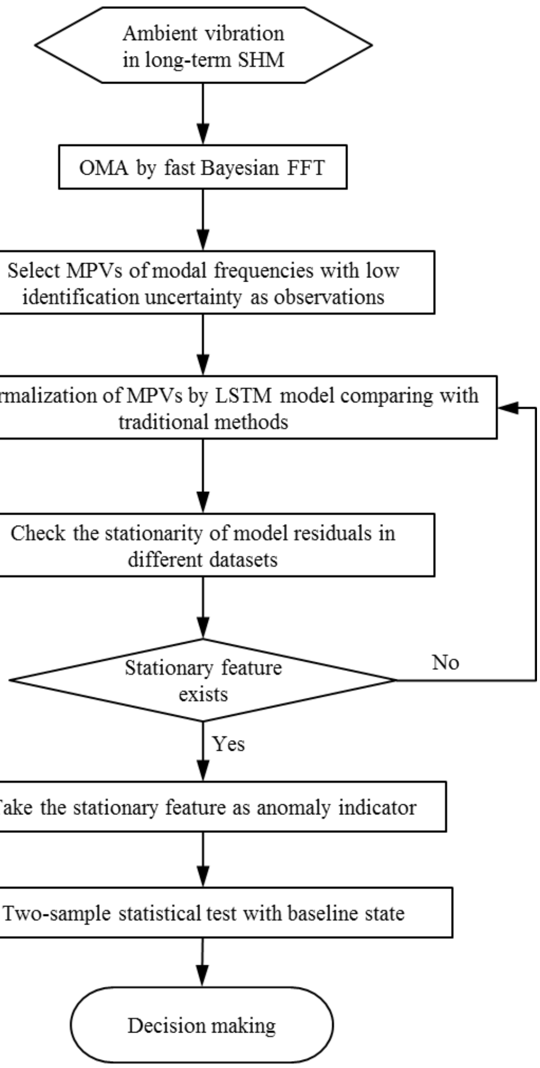


Fig. 1. Flowchart for the proposed residual-based anomaly detection procedure.

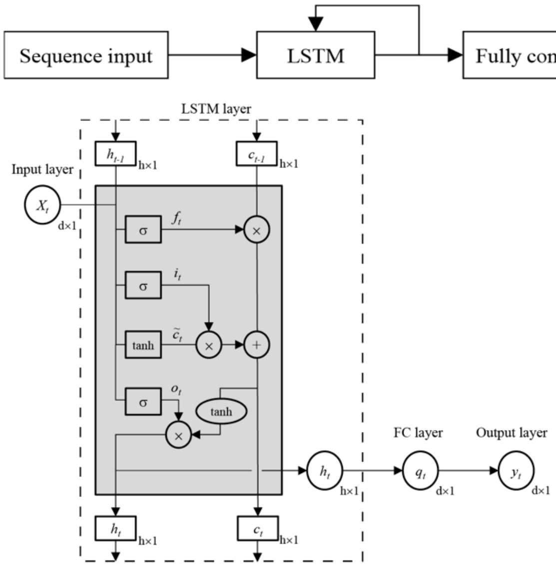


Fig. 2. (a) Common LSTM network for regression and (b) schematic diagram of a vanilla LSTM layer.

Fig. 3. Elevation view of the target bridge.

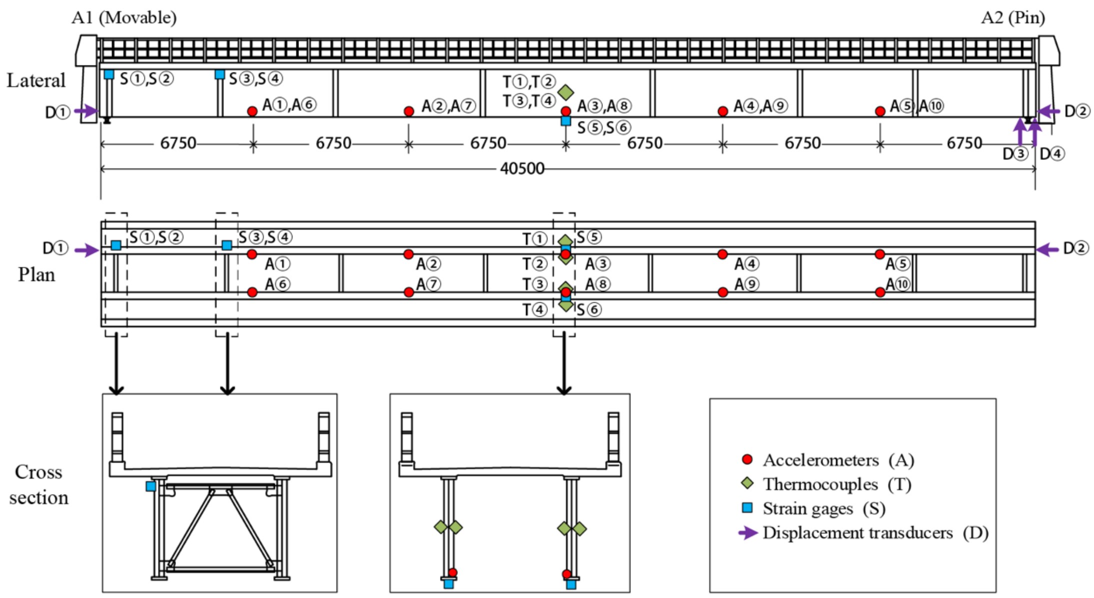


Fig. 4. Sensor deploying map.

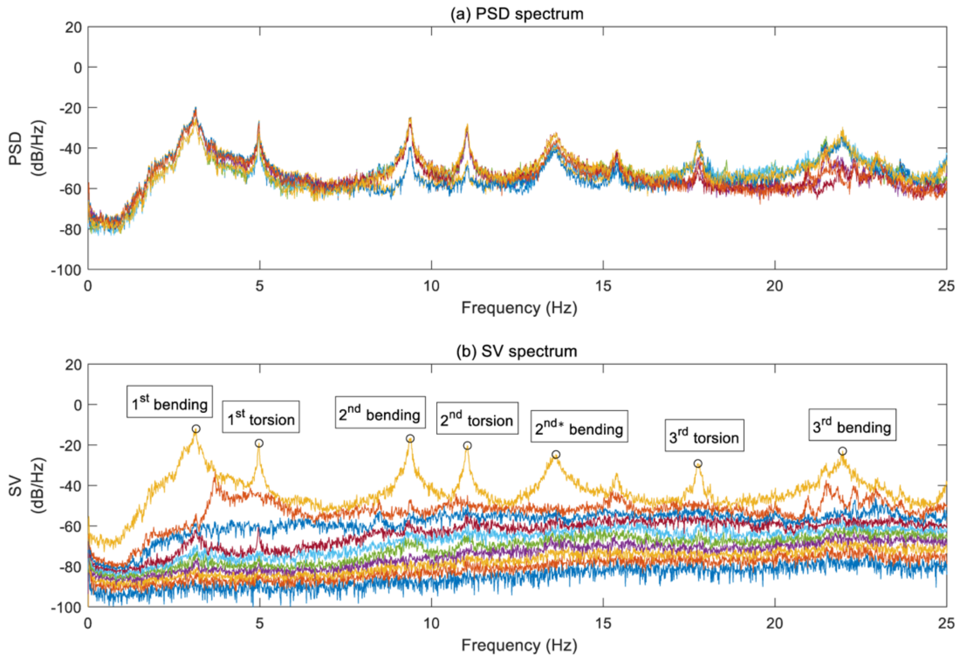


Fig. 5. PSD and SV spectrum.

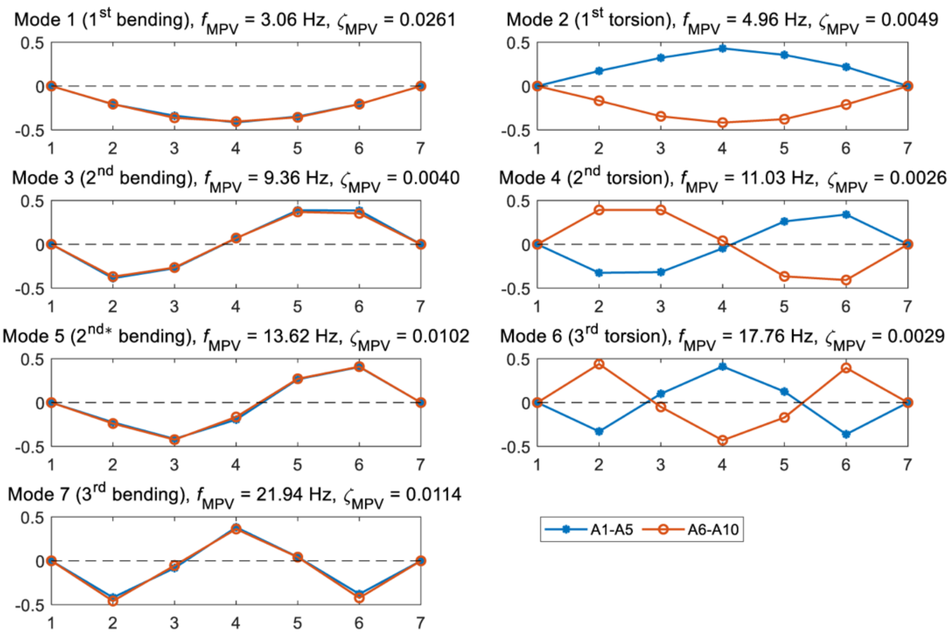


Fig. 6. Identified mode shapes.

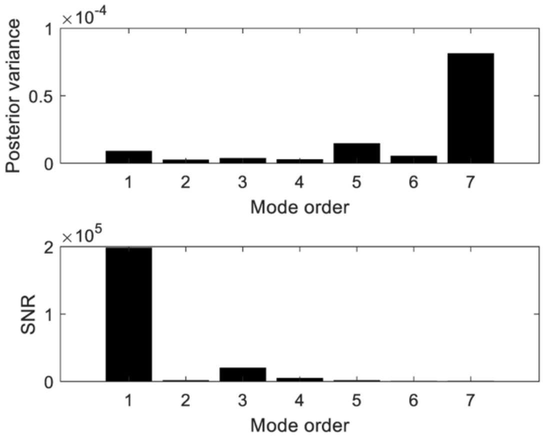


Fig. 7. Identified posterior variance and SNR.

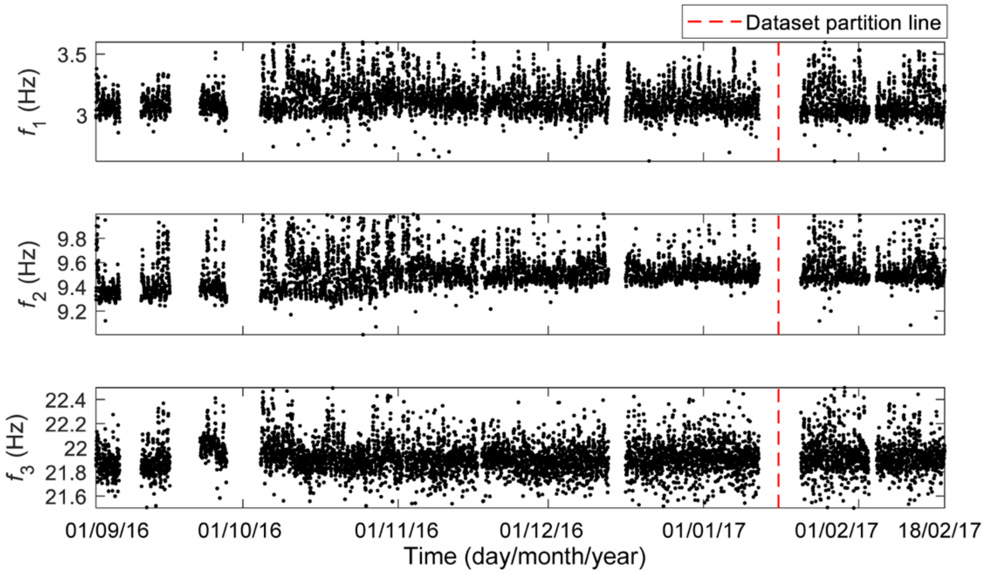


Fig. 8. MPVs of identified modal frequencies in long-term SHM.

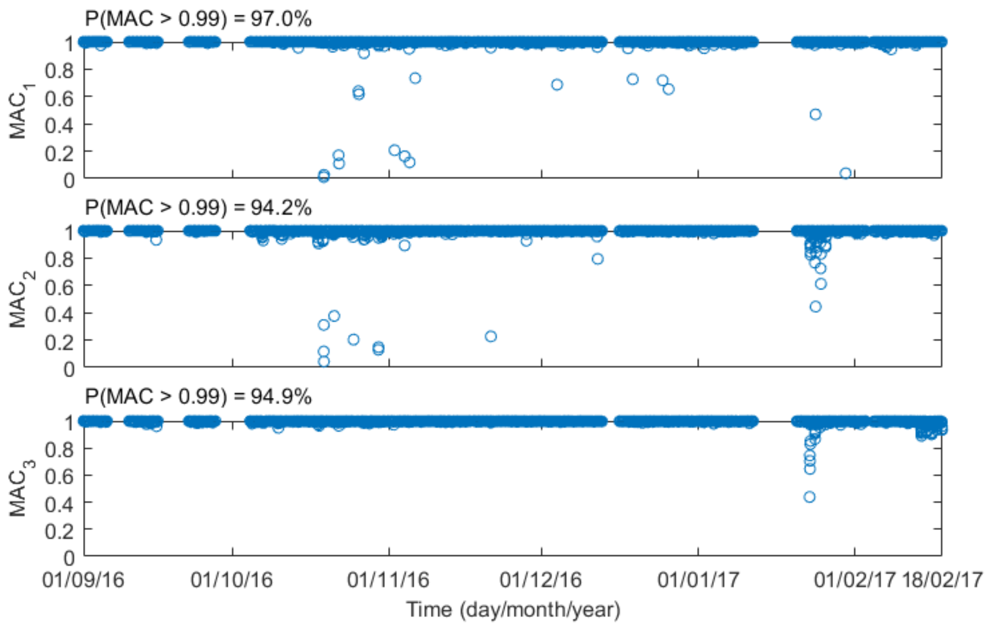


Fig. 9. Temperature records in long-term SHM.

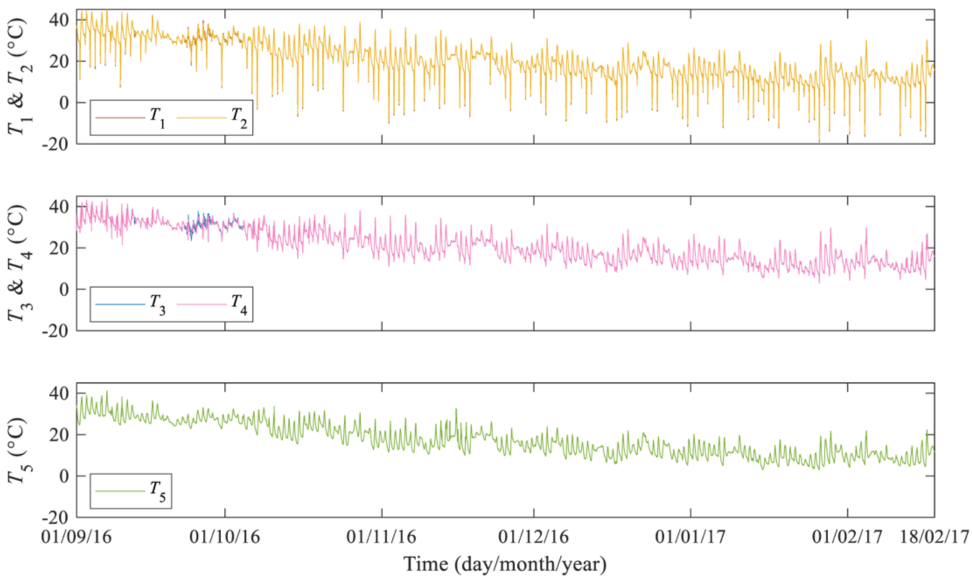


Fig. 10. Correlation plot among temperature records and f_1.

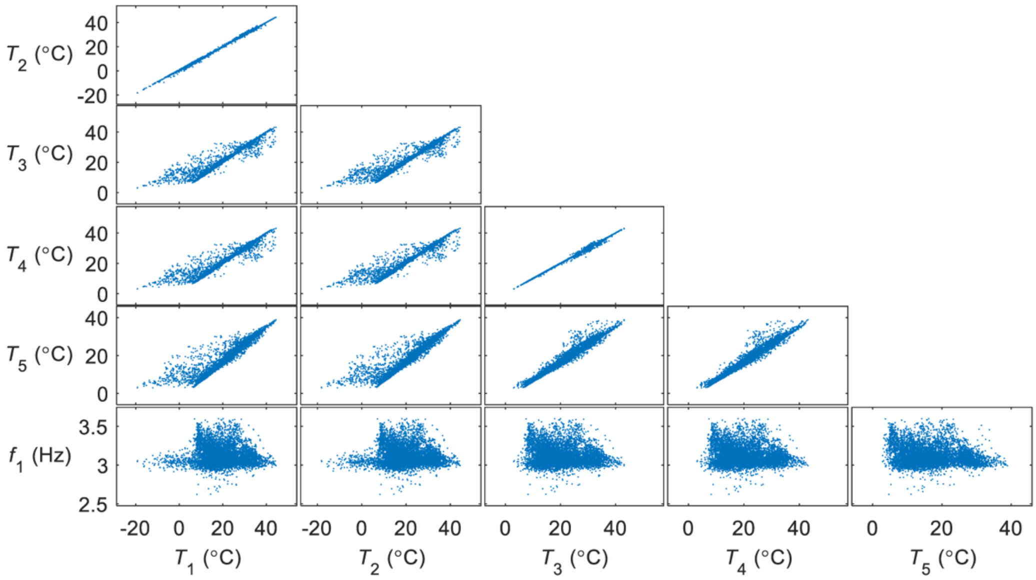


Fig. 11. Correspondence of standardized modal frequencies and temperatures.

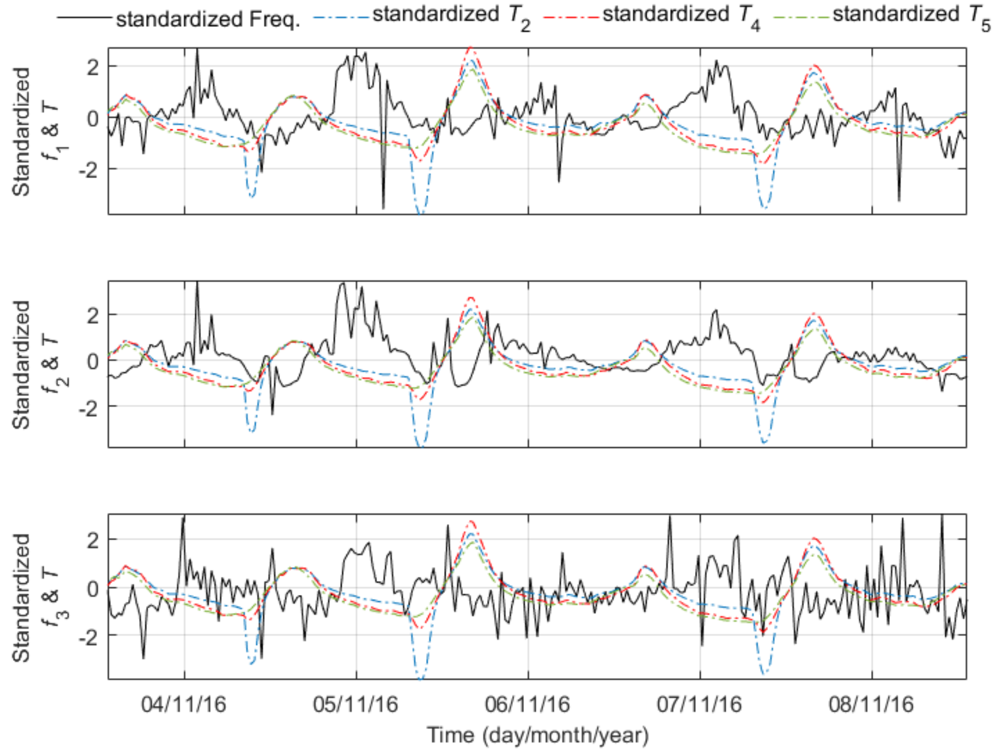
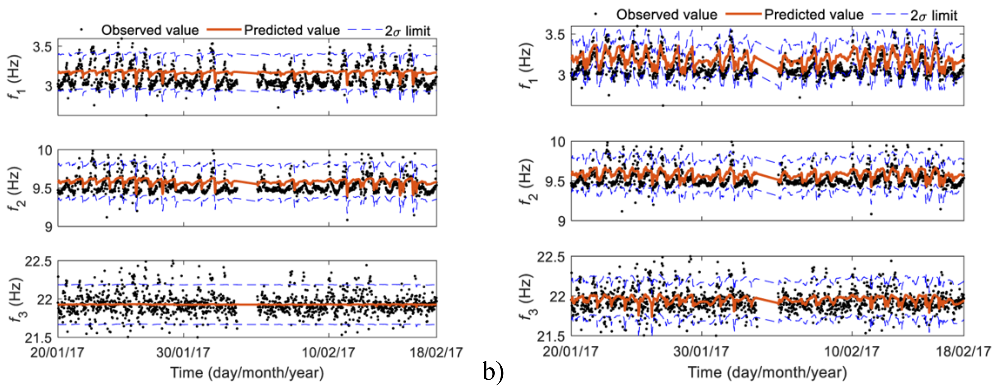
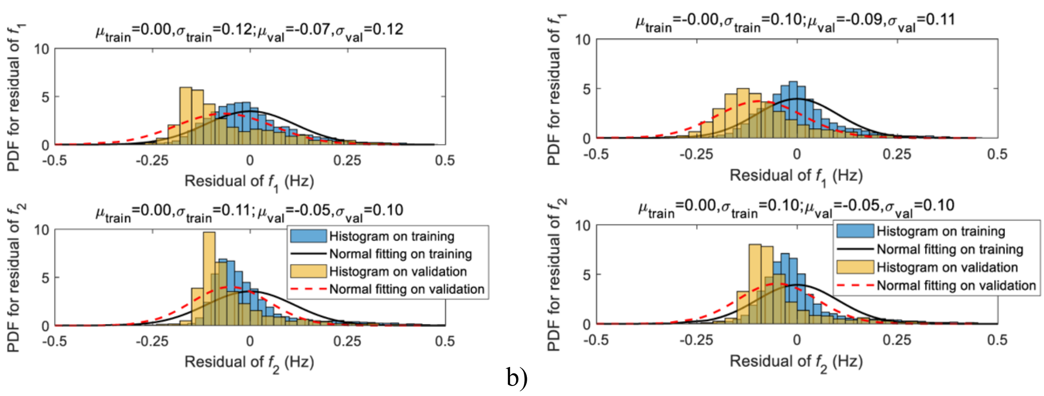


Fig. 12. Correspondence of standardized modal frequencies and temperatures.



a) b)

Fig. 13. Prediction on validation set: (a) BLR with non-informative prior and (b) GPR.



a) b)

Fig. 14. Residual distributions on training set and validation set: (a) BLR and (b) GPR.

a) b)

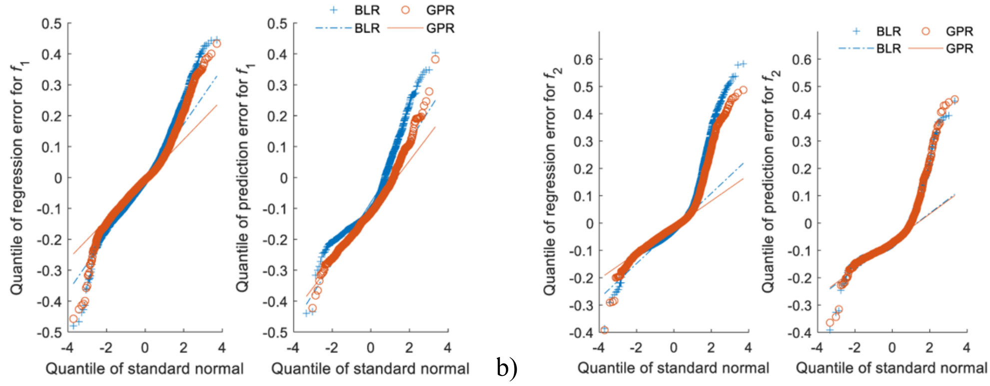


Fig. 15. Q-Q plots of residuals on training set and validation set: (a) for f1; and (b) for f2.

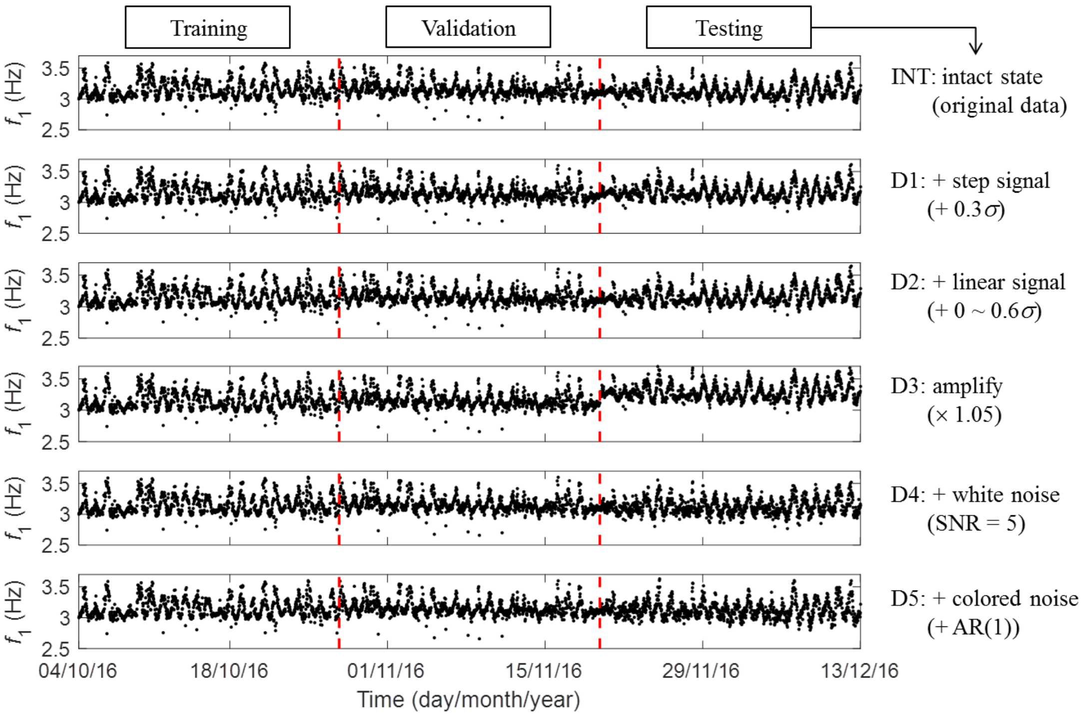


Fig. 16. Continuous segment of f1 with five simulated anomaly scenarios. σ = standard deviation of original data.

a) b)

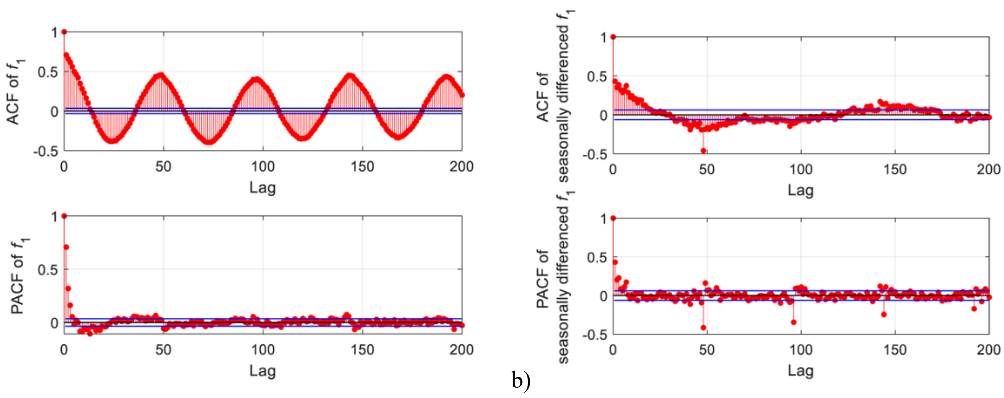


Fig. 17. ACF and PACF: (a) for f1; and (b) for seasonal differencing sequence of f1.

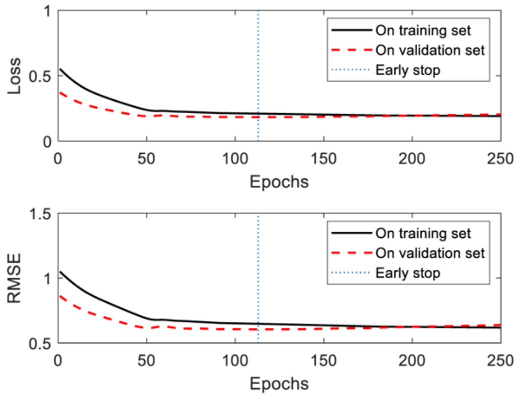


Fig. 18. Training process of LSTM network.

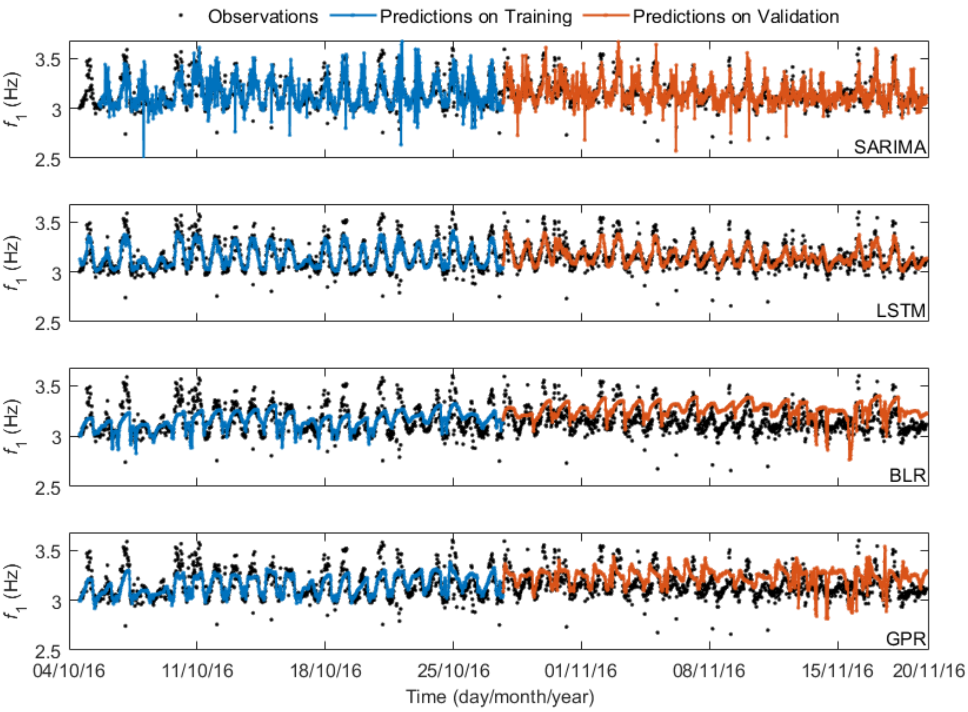


Fig. 19. Predictions of four models on training set and validation set.

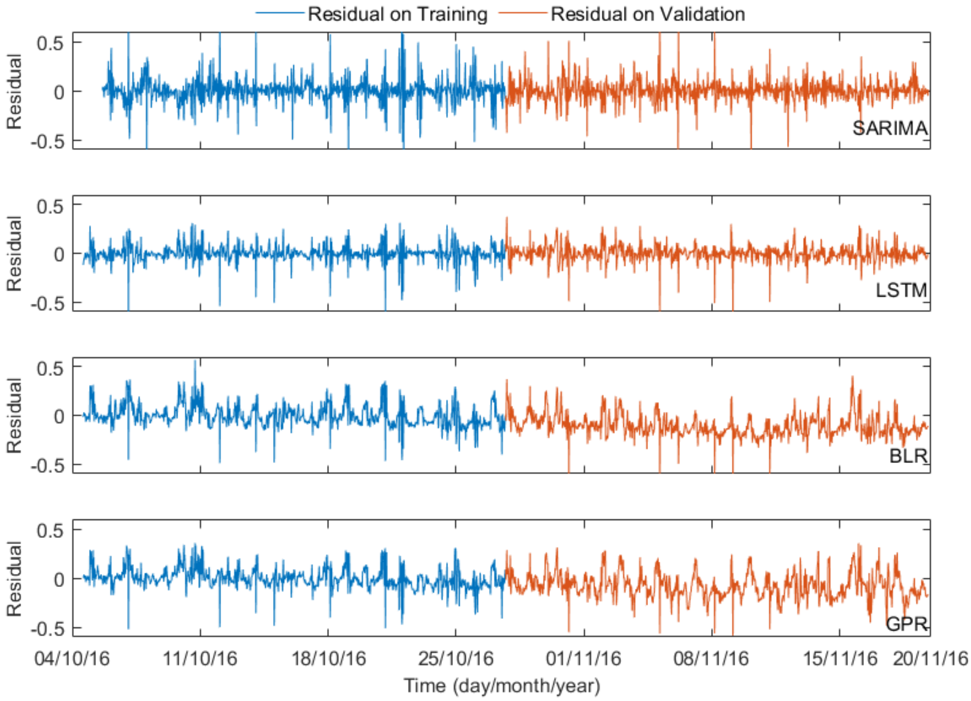


Fig. 20. Residual sequences of four models on training set and validation set.

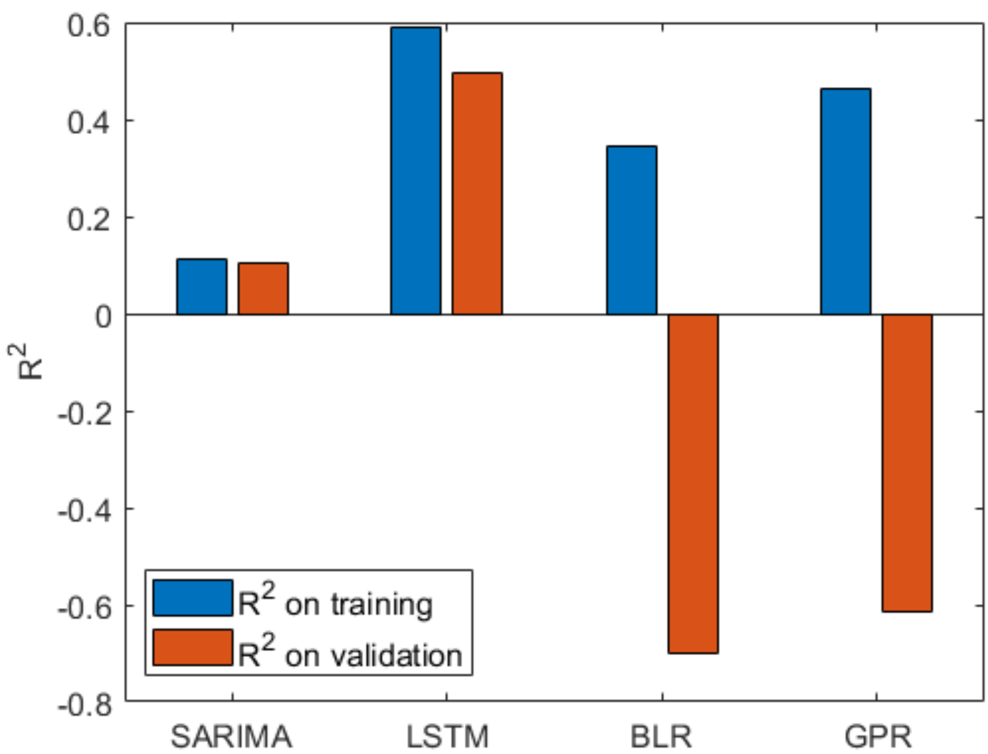


Fig. 21. Coefficient of determination R2 of four models on training set and validation set.

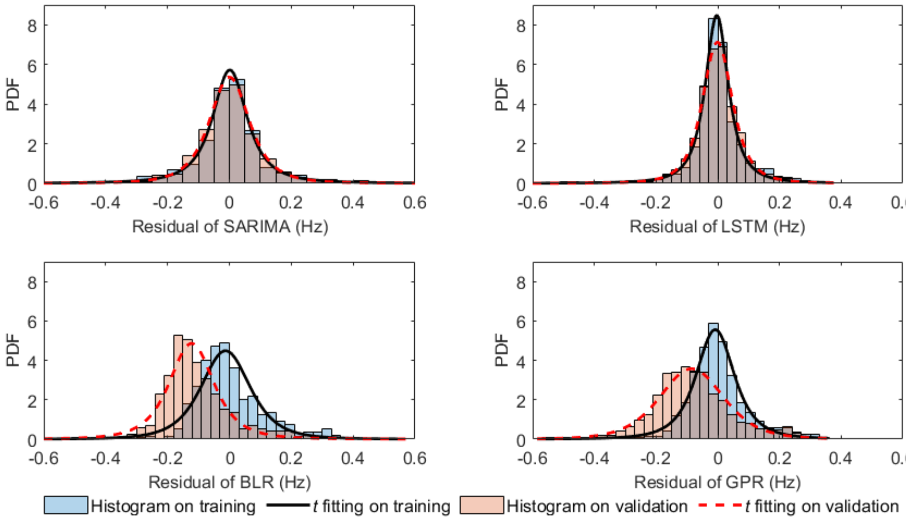
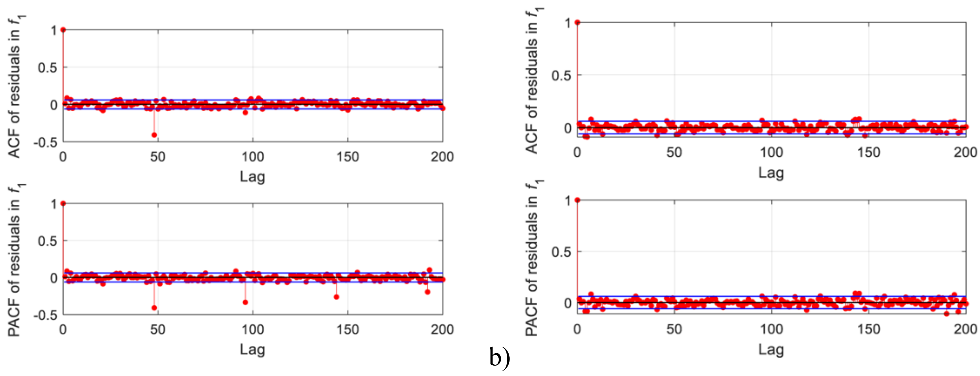


Fig. 22. Residual distributions on training set and validation set.



a) b)

Fig. 23. ACF and PACF of residuals: (a) for SARIMA; and (b) for LSTM

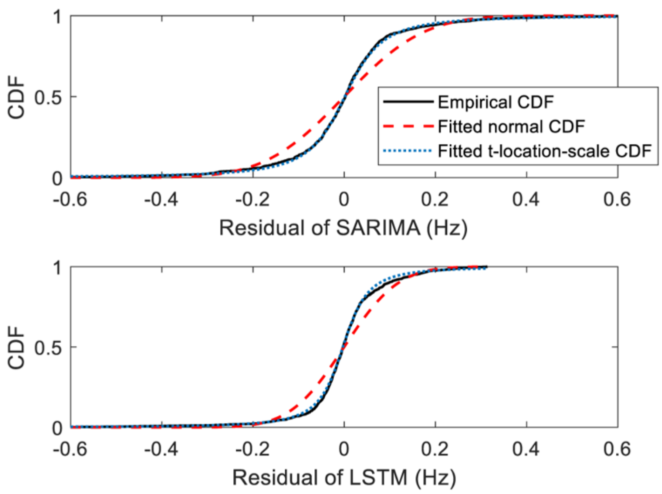


Fig. 24. Cumulative distribution functions (CDFs) of residuals.

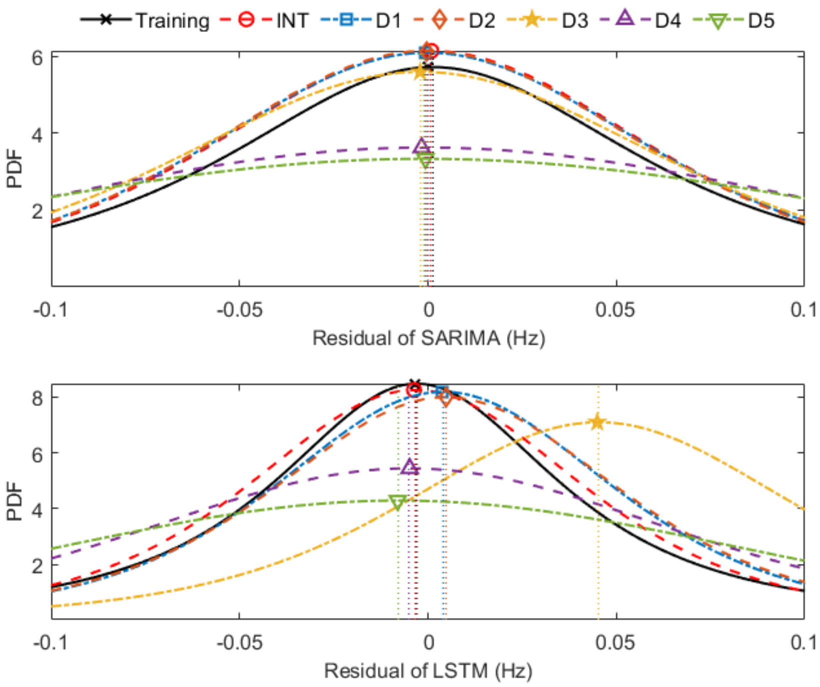


Fig. 25. Statistical feature of residuals: fitted t-location-scale distributions. Vertical dotted lines denote the means of each scenario.