



## Article

# Study on Performance Evaluation and Prediction of Francis Turbine Units Considering Low-Quality Data and Variable Operating Conditions

Ran Duan , Jie Liu , Jianzhong Zhou \*, Yi Liu, Pei Wang and Xinqiang Niu

School of Civil and Hydraulic Engineering, Huazhong University of Science and Technology, Wuhan 430074, China; duan\_ran@hust.edu.cn (R.D.); jie\_liu@hust.edu.cn (J.L.); liu\_yi@hust.edu.cn (Y.L.); w\_sara@hust.edu.cn (P.W.); xq\_niu\_hust@163.com (X.N.)

\* Correspondence: jz2258842051@163.com

**Abstract:** The stable operation of the Francis turbine unit (FTU) determines the safety of the hydropower plant and the energy grid. The traditional FTU performance evaluation methods with a fixed threshold cannot avoid the influence of variable operating conditions. Meanwhile, anomaly samples and missing values in the low-quality on-site data distort the monitoring signals, which greatly affects the evaluation and prediction accuracy of the FTU. Therefore, an approach to the performance evaluation and prediction of the FTU considering low-quality data and variable operating conditions is proposed in this study. First, taking the variable operating conditions into consideration, a FTU on-site data-cleaning method based on DBSCAN is constructed to adaptively identify the anomaly samples. Second, the gate recurrent unit with decay mechanism (GRUD) and the Wasserstein generative adversarial network (WGAN) are combined to propose the GRUD–WGAN model for missing data imputation. Third, to reduce the impact of data randomness, the healthy-state probability model of the FTU is established based on the GPR. Fourth, the prediction model based on the temporal pattern attention–long short-term memory (TPA–LSTM) is constructed for accurate degradation trend forecasting. Ultimately, validity experiments were conducted with the on-site data set of a large FTU in production. The comparison experiments indicate that the proposed GRUD–WGAN has the highest accuracy at each data-missing rate. In addition, since the cleaning and imputation improve the data quality, the TPA–LSTM-based performance indicator prediction model has great accuracy and generalization performance.

**Keywords:** Francis turbine unit; performance state evaluation; degradation trend prediction; data imputation; healthy-state model



**Citation:** Duan, R.; Liu, J.; Zhou, J.; Liu, Y.; Wang, P.; Niu, X. Study on Performance Evaluation and Prediction of Francis Turbine Units Considering Low-Quality Data and Variable Operating Conditions. *Appl. Sci.* **2022**, *12*, 4866. <https://doi.org/10.3390/app12104866>

Academic Editor: José A.F.O. Correia

Received: 11 March 2022

Accepted: 10 May 2022

Published: 11 May 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Hydropower is an important renewable and clean energy. With the increasingly severe energy and climate challenges, it is imperative to develop hydropower energy safely and efficiently. As a critical equipment for hydropower energy utilization, the Francis turbine unit (FTU) also undertakes essential tasks such as peak frequency modulation and emergency standby in the power grid system. Therefore, ensuring the safe and stable operation of the FTU is of great significance in promoting the development of the national economy and ensuring the stability of the energy system [1]. Currently, the maintenance strategy of the FTU is mainly routine maintenance and repair after failures, which has a high cost and makes it difficult to recognize the early signs of fault in time. Therefore, the performance evaluation and the degradation trend prediction of the FTU have attracted more and more attention [2–4]. Although the studies of general rotating machinery prognostics are relatively mature [5,6], there are still two practical difficulties in the field of engineering applications of the FTU: (1) The quality of the on-site measured data is usually too low, characterized by low sampling frequency, missing data and anomaly

data. (2) The drastic variation in operating conditions makes it difficult to evaluate and predict the performance of FTU accurately.

The working environment of the FTU is hostile, including external interferences such as moisture, dust, vibration and electromagnetic disturbance. The supporting monitoring and data acquisition system of the FTU often involves multiple distributed modules with complex data transmission link structures and long physical distances. Hence, there are often many anomaly samples and missing values in the on-site raw data of FTUs, caused by sensor failure, short-term router failure or electromagnetic interference [7]. Meanwhile, the storage space of the data acquisition system is limited, and the short-term variation trend of each monitored quantity is not evident during the long-term working period of the FTU, so the storage frequency of the on-site data is often relatively low. Some studies about the performance evaluation and the remaining useful life prediction of rotating machinery have achieved excellent results in laboratory environments. However, these approaches often rely on high-quality data and are difficult to directly apply to the engineering practice of the FTU [8–10]. Focusing on the anomaly data, some current research has adopted denoising methods based on frequency domain analysis or energy spectrum analysis, which are effective while the sampling frequency is high and consistent [11–13]. However, the missing values and the variable condition significantly affect their effectiveness. The clustering method has been proved to have significant performance in the recognition of outliers of high-dimensional data [14–16]. As a density-based clustering method, the density-based spatial clustering of applications with noise (DBSCAN) algorithm is widely used in anomaly data detection due to its simple structure and good adaptability to high-dimensional data [17–20]. Hence, the DBSCAN is adopted to clean the raw monitoring data in this paper.

Except for the missing samples in the raw data set, data cleaning also increases the amount of missing data. If these missing values were simply deleted, the potentially important information might be discarded. In order to evaluate and predict the performance of the FTU more effectively, it is necessary to fill in the missing data appropriately. Traditional filling methods based on statistics, such as mean filling and median filling, mostly ignore the time sequence information between data. With the rapid development of machine learning theory and technology, more and more studies are processing sequences with missing values based on improved recurrent neural networks (RNN) [21–23]. Che et al. introduced a decay mechanism to the typical gate recurrent unit model (GRU) to construct the GRUD model. It was proved that the decay mechanism enables GRUD to effectively learn potential patterns in sequences with missing values [24]. GRUD has achieved good results in the prediction of incomplete sequences. However, these supervised regression methods cannot be directly used to generate complete sequences because the training targets cannot be set for the missing values. As one of the most promising models in unsupervised learning on complex distribution, the generative adversarial network (GAN) model has made outstanding achievements in nonlinear model analysis and image generation [25–27]. Based on the competitive learning between the generator and the discriminator, this model can adaptively learn the expression paradigm in the input data [28–30]. The classic GAN model has problems such as training difficulties and mode collapse. Martin introduced the Wasserstein distance to guide the training process, which significantly improved the performance of the GAN model [31,32]. At present, the time series data-generating ability of the WGAN remains to be studied. Therefore, in this paper, the GRUD–WGAN model is proposed to realize the missing data imputation of the on-site data.

Because of the influence of natural inflow conditions and the adjustment requirements of the power grid, the operating parameters of the FTU vary across a wide range and are generally of a high frequency. The monitoring data are highly correlated with operating conditions. Traditional performance evaluation methods of the FTU are primarily based on the overage alarm strategy and the fixed threshold, ignoring the correlation between monitoring data and working condition parameters [33]. To solve this problem, Shan et al. adopted the backpropagation neural network to construct the nonlinear mapping

relationship between operating parameters and the vibration amplitude of the lower bracket [34]. This research realizes the evaluation of the FTU under variable operating conditions. However, the definite numerical mapping relation is susceptible to the random fluctuation of signals. Therefore, the Gaussian process regression (GPR) is introduced to establish a probability mapping model between operating parameters and the probability density distribution of monitored values, so as to improve the robustness of the performance evaluation model against random noise.

After quantifying the abstract FTU performance into performance degradation indicators (PDI), the degradation trend prediction problem is essentially transformed into a time series forecasting task. Because RNNs, such as the GRU and the long short-term memory (LSTM) network, can learn the potential timing information, they are widely used in sequence prediction [35–38]. Shih et al. added the temporal pattern attention (TPA) mechanism based on the LSTM structure to further improve the performance of the LSTM model in mining temporal dependencies [39]. In this paper, the TPA–LSTM is adopted to construct the prediction model for the performance degradation indicator of the FTU.

To sum up, in the field of the evaluation and performance prediction of the FTU, there are few studies on the cleaning of low-quality on-site data. The data imputation method of the incomplete data set of the FTU is not yet mature. Meanwhile, few studies have considered the random fluctuation of data in establishing a healthy model of the FTU. In addition, there is room for the further improvement of the accuracy of PDI prediction models.

In this paper, an approach for the performance evaluation and prediction of FTU considering low-quality data and variable operating conditions is proposed. The main contributions are highlighted as follows:

- (1) Considering the variable operating conditions and the characteristics of the anomaly samples, an on-site data-cleaning method based on DBSCAN is constructed to adaptively detect both singulars and outliers.
- (2) Combining the incomplete sequence information mining ability of the GRUD and the hidden pattern learning ability of the WGAN, the GRUD–WGAN-based missing value imputation model is proposed to improve the low-quality data utilization value.
- (3) Based on the GPR, the mapping relationship between the operating parameters and the distribution of monitored data is established as the healthy-state probability model of the FTU. The robustness of the healthy model to the random noise is improved because the distribution probability, instead of a single value, is taken into consideration.
- (4) The TPA–LSTM-based PDI prediction model is constructed to realize the accurate degradation trend prediction as the basis for predictive maintenance of the FTU.

The rest of this article is organized as follows. The framework and procedures of the proposed approach are explained in detail in Section 2. The proposed method is applied on a large practical FTU, and the results are presented in Section 3. The performance of the proposed data imputation model and the trend prediction model is emphatically compared and discussed in Section 4. Finally, the conclusion is given in Section 5.

## 2. Proposed Method

In this paper, an approach for the performance evaluation and prediction of FTUs considering low-quality data and variable operating conditions is proposed. The proposed framework is illustrated in Figure 1. First, the DBSCAN algorithm is introduced to clean the anomaly samples in the raw monitoring data set of the FTU, which includes the water head ( $H$ ), the active power ( $P$ ) and the vibration amplitude ( $V$ ) of the top cover. Second, the GRUD–WGAN model is proposed to fill in the missing values in the raw data set or those caused by data cleaning. Third, the healthy-state probabilistic model of the FTU under complex operating conditions is established based on the complete data set and the GPR algorithm. The negative log-likelihood probability (NLLP) between the data to be evaluated and the healthy-state model is defined as the PDI of the FTU. Finally, to forecast the degradation trend of the FTU, the PDI prediction model is constructed based on the TPA–LSTM algorithm.

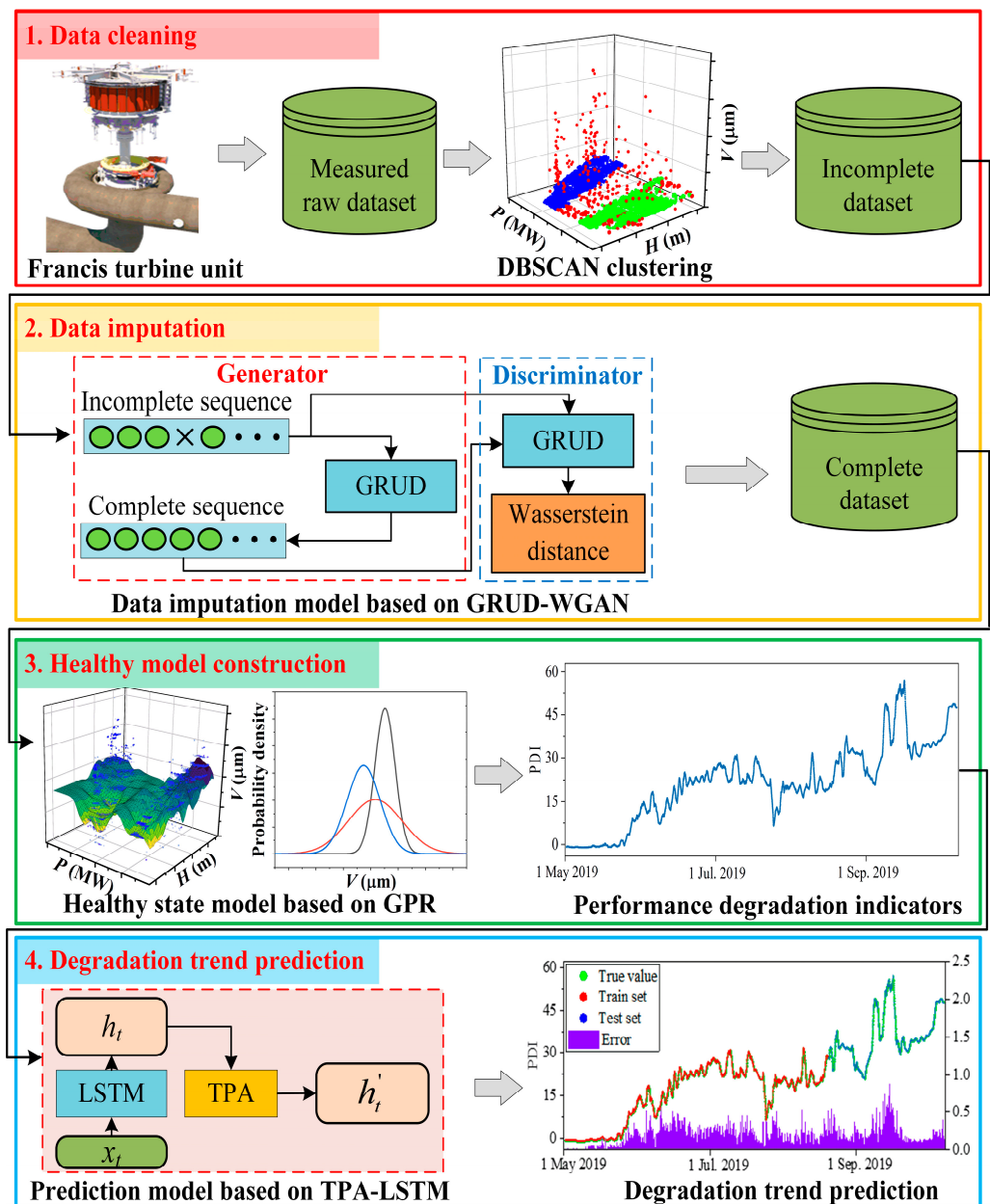


Figure 1. The framework of the proposed method.

### 2.1. Data Cleaning

The condition monitoring systems of large FTUs usually have a huge scale and complex structures. The state-monitoring data and operating condition parameters are usually distributed and monitored by several different monitoring modules, and collected into the computer monitoring system of the hydropower station through various data communication protocols and long-distance communication cables, which are prone to communication packet loss or short-term failure. Moreover, FTUs work in a humid, high-electromagnetic-interference and drastic-vibration environment. These factors lead to an apparent anomaly or missing values in the on-site raw data. Traditional signal denoising methods are mainly based on signal decomposition and reconstruction. Frequent changes in operating conditions will affect the effectiveness of these methods. Meanwhile, random missing values make these methods challenging to apply to engineering practice.

The DBSCAN algorithm is a kind of unsupervised clustering method. As an effective density-clustering method, DBSCAN can adaptively identify clusters with irregular shapes and automatically mark sample points with low density as noise. The DBSCAN is adopted

to adaptively recognize the anomaly values in the raw data set. The schematic diagram of the DBSCAN algorithm is illustrated in Figure 2.

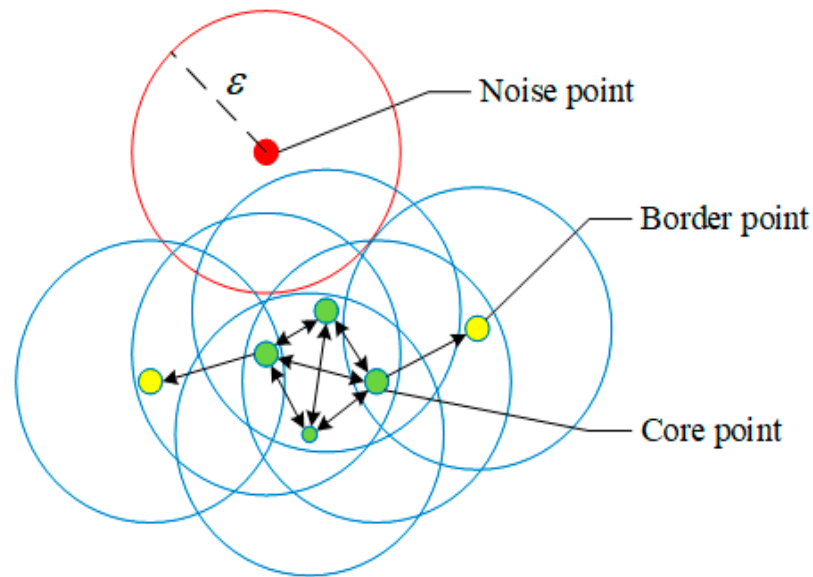


Figure 2. Schematic of the DBSCAN.

For each sample  $p$  in the data set  $\Psi$ , the region for which the Euclidean distance from the sample point  $p$  is less than  $\epsilon$  is defined as the  $\epsilon$ -neighborhood of  $p$ . Its element set is expressed as:

$$N_\epsilon(p) = \{q \in \Psi | \rho(p, q) \leq \epsilon\} \tag{1}$$

where  $\rho(p, q)$  represents the Euclidean distance between samples  $p$  and  $q$ .

If sample  $q$  is in the  $\epsilon$ -neighborhood of sample  $p$ ,  $p$  and  $q$  are called directly density-reachable to each other. If sample  $r$  is also directly density-reachable to  $q$ , but  $r$  is not directly density-reachable to  $p$ ,  $r$  and  $p$  are called density-reachable to each other. If the element number of  $N_\epsilon(p)$  is greater than the minimum density threshold  $Z$ , sample  $p$  is defined as a core point. If sample  $r$  is located in the  $\epsilon$ -neighborhood of a particular core point, but  $r$  is not a core point, then  $r$  is defined as a border point.

The samples which are directly density-reachable or density-reachable to the core point construct a cluster, as illustrated by blue circles in Figure 2. The samples that do not belong to any clusters are noise points, marked as red points in Figure 2. The recognized noise points in the raw data set are dropped out.

### 2.2. Missing Value Imputation

To provide a complete data set for subsequent health status evaluation, the GRUD-WGAN model is proposed to fill in the missing values. The main inspiration of the GRUD-WGAN is to use GRUD to receive incomplete sequences with missing values and convert them into complete hidden sequences. Then, through the antagonistic training of the generator and the discriminator under the WGAN framework, the distribution of valid values is learned adaptively, so as to generate a proper complete sequence.

#### 2.2.1. GRUD

GRU has been widely proved to have an excellent ability to capture dependencies between time series data [37]. However, the traditional GRU cannot handle sequences with missing values. Based on the GRU model, GRUD adds the decay mechanism to estimate the missing values according to the previous sequence [24]. The schematic diagram of the GRUD model is shown in Figure 3. The trainable decay coefficient  $\gamma$  is defined as:

$$\gamma_t = \exp(-\max(0, W_\gamma \delta_t + b_\gamma)) \tag{2}$$

where  $\delta_t$  is the time interval between the current moment  $t$  and the last non-missing value,  $W_\gamma$  and  $b_\gamma$  represent the weight and bias of a neural network so that that  $\gamma$  can be updated during the training process.

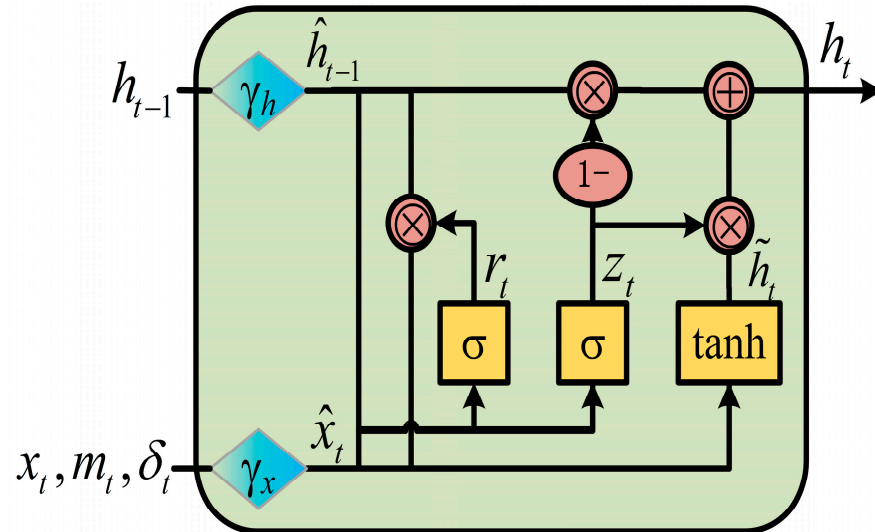


Figure 3. Schematic of the GRUD.

The missing values of the input data  $x$  are replaced by  $\hat{x}$ , expressed as:

$$\hat{x}_t = m_t x_t + (1 - m_t)(\gamma_x x_{t'} + (1 - \gamma_x)\bar{x}) \tag{3}$$

$$m_t = \begin{cases} 0, & \text{if } x_t \text{ is missing} \\ 1, & \text{otherwise} \end{cases} \tag{4}$$

where  $m_t$  is the mask code,  $x_{t'}$  is the last non-missing value, and  $\bar{x}$  indicates the mean value of  $x$ .

The decay mechanism is also applied to the hidden state  $h$  to enhance the learning of missing value patterns:

$$\hat{h}_t = \gamma_h \odot h_t \tag{5}$$

where  $\odot$  represents element multiplication.

Moreover, the mask code  $m$  is fed into the GRU cell directly. Finally, the update functions of GRUD are as follows:

$$r_t = \sigma(W_r [\hat{x}_t, \hat{h}_{t-1}, m_t] + b_r) \tag{6}$$

$$z_t = \sigma(W_z [\hat{x}_t, \hat{h}_{t-1}, m_t] + b_z) \tag{7}$$

$$\tilde{h}_t = \tanh(W [\hat{x}_t, r_t \odot \hat{h}_{t-1}, m_t] + b) \tag{8}$$

$$h_t = (1 - z_t) \odot \hat{h}_{t-1} + z_t \odot \tilde{h}_t \tag{9}$$

where  $r$  and  $z$  represent the reset gate and the update gate, and  $\sigma()$  and  $\tanh()$  indicate the sigmoid and tanh activation function.

### 2.2.2. WGAN

The GAN model is inspired by the two-player zero-sum game [26]. The typical structure of the GAN includes a discriminator and a generator. The goal of the discriminator is to correctly distinguish between actual data sampled from the input data set and fake data generated by the generator. The purpose of the generator is to produce fake data that can deceive the discriminator. GAN is trained by alternating adversarial learning between

discriminator and generator, and the optimum objective is to achieve Nash equilibrium. Finally, the generator can accurately estimate the distribution of data samples. However, the original GAN has some problems, such as training difficulty and mode collapse. Therefore, Arjovsky et al. proposed the WGAN model, which adopts the Wasserstein distance instead of the Jensen–Shannon divergence to indicate the difference between the actual samples and the generated samples [32]. Specifically, the Wasserstein distance of the WGAN can be simplistically expressed as:

$$W = \max(E[f(x_r)] - E[f(x_g)]) \tag{10}$$

where  $E[\ ]$  represents the mathematical expectation,  $x_r$  and  $x_g$  represent samples of the real data and the generated data, respectively, and  $f(\ )$  indicates a neural network model for which the last layer is not a nonlinear activation layer.

The introduction of the Wasserstein distance solves the problem of gradient extinction. To minimize the Wasserstein distance, the loss function of the generator  $G(\ )$  and the discriminator  $D(\ )$  of WGAN can be expressed as:

$$L_G = -D(G(x)) \tag{11}$$

$$L_D = D(G(x)) - D(x_r) \tag{12}$$

In addition, to satisfy the Lipschitz continuity condition, the parameters of the discriminator network need to be clipped to  $[-c, c]$ .  $c$  is a fixed constant, and its value does not affect the direction of the gradient.

### 2.2.3. GRUD–WGAN Model

The GRUD and the WGAN models are combined to establish the GRUD–WGAN model, as shown in Figure 4. The GRUD model is adopted to be the essential component of both the generator and the discriminator of the WGAN framework. Specifically, the generator is constructed with a GRUD layer and a linear layer. The incomplete input sequence  $X_t$  at time  $t$  with length  $l$  is fed into the GRUD model, and  $n$ -dimensional hidden state vectors are outputted. The linear layer maps these hidden state vectors to a reconstructed sequence  $\tilde{X}_t$  of the same shape as the input sequence. Then, the complete output sequence  $\hat{X}_t$  is calculated according to the mask code matrix  $M_t$ , expressed as:

$$\hat{X}_t = M_t \odot X_t + (1 - M_t) \odot \tilde{X}_t \tag{13}$$

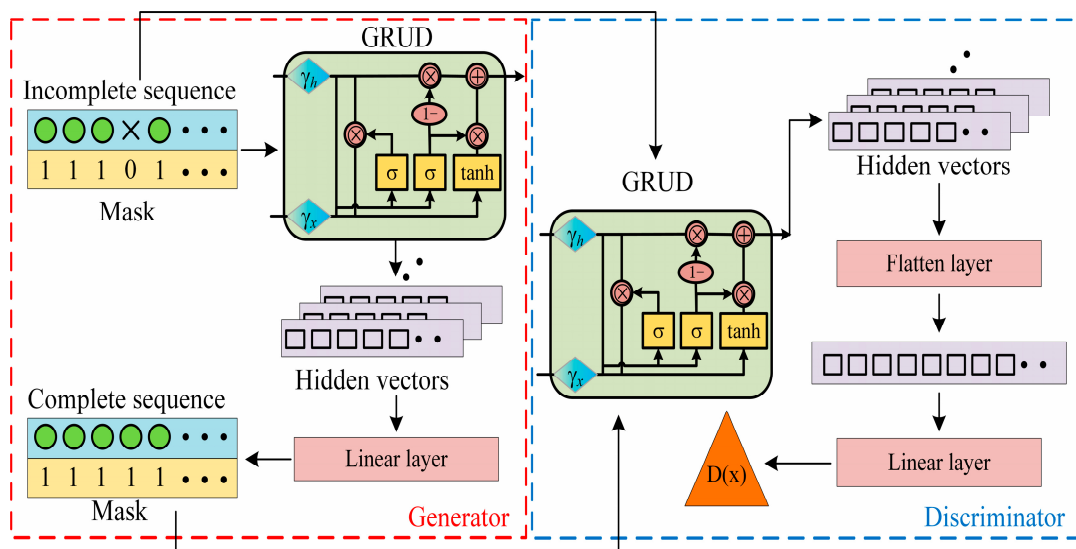


Figure 4. The framework of the proposed GRUD-WGAN.

The discriminator is also built by a GRUD layer and a linear layer. The GRUD is adopted to accept the incomplete sequences of the real data set or the complete imputed data set produced by the generator. The linear layer is used to map the hidden state vectors to a single value, which indicates the Wasserstein distance.

To make full use of the available data to accelerate convergence, the reconstruction error  $L_r$  is added to the loss function of the generator, given by:

$$L_r = \|(\tilde{X}_t - X_t) \odot M_t\|_2 \tag{14}$$

$$L_G = -D(G(X_t)) + L_r \tag{15}$$

when the training process converges, a generator that receives incomplete sequences and outputs complete sequences can be obtained.

### 2.3. Healthy-State Model Construction

To realize the accurate performance evaluation of FTUs under variable operating conditions, the mapping relationship between operating parameters and the probability density distribution of the vibration amplitude is constructed based on the GPR algorithm. GPR is a non-parametric regression method based on the probability statistical theory, which shows a strong generalization ability and adaptability in dealing with complex fitting and regression tasks. The GPR constructs the time series model through the Gaussian prior knowledge. The Gaussian prior is the distribution of  $f(X)$  values corresponding to each independent variable  $X$ . It can be described by the mean function  $\mu(\cdot)$  and the covariance function  $\kappa(\cdot)$ , expressed as:

$$Y = f(X) \sim N(\mu, \kappa) \tag{16}$$

According to the Bayesian inference, the joint distribution of actual observation samples  $Y^*$  and the dependent variable  $Y$  also obeys the Gaussian distribution, given by:

$$\begin{bmatrix} Y \\ Y^* \end{bmatrix} \sim N\left(\begin{bmatrix} \mu(Y) \\ \mu(Y^*) \end{bmatrix}, \begin{bmatrix} \kappa(X, X) & \kappa(X, X^*) \\ \kappa(X, X^*)^T & \kappa(X^*, X^*) \end{bmatrix}\right) \tag{17}$$

where the superscript \* indicates the actual observation value.

After expanding Equation (17), the mean and the variance of  $Y$  can be expressed as:

$$Y \sim N\left(\kappa_{fy}^T \kappa_{ff}^{-1} \mu(Y^*), \kappa_{yy} - \kappa_{fy}^T \kappa_{ff}^{-1} \kappa_{fy}\right) \tag{18}$$

$$\mu(Y) = \kappa_{fy}^T \kappa_{ff}^{-1} \mu(Y^*) \tag{19}$$

$$var(Y) = \kappa_{yy} - \kappa_{fy}^T \kappa_{ff}^{-1} \kappa_{fy} \tag{20}$$

According to the Gaussian distribution formula, the probability density of  $Y$  can be expressed as:

$$P(Y) = \frac{1}{\sqrt{2\pi var(Y)}} \exp\left(-\frac{(Y - \mu(Y))^2}{2var(Y)}\right) \tag{21}$$

The healthy standard distribution model of vibration amplitude is constructed by the monitoring data acquired during the normal working period of the FTU. The *NLLP* between the healthy standard distribution and the data to be evaluated  $Y'$  is defined as the PDI, given by:

$$NLLP = -\log(P(Y')) \tag{22}$$

The smaller the *NLLP* value is, the more similar the distribution of the data to be evaluated is to the healthy standard distribution, and the better the FTU status is, and vice versa.



2.4. Degradation Trend Prediction

After quantifying the differences between the data to be evaluated and the health model as the PDIs of the FTU, the degradation trend prediction task is converted to a time series forecasting problem. The TPA-LSTM introduced a temporal pattern attention mechanism based on the 1DCNN to the traditional LSTM model. The TPA-LSTM effectively improves the accuracy and stability of time series prediction by learning the attention weights of previous hidden states [39]. The basic framework of the TPA-LSTM model is shown in Figure 5.

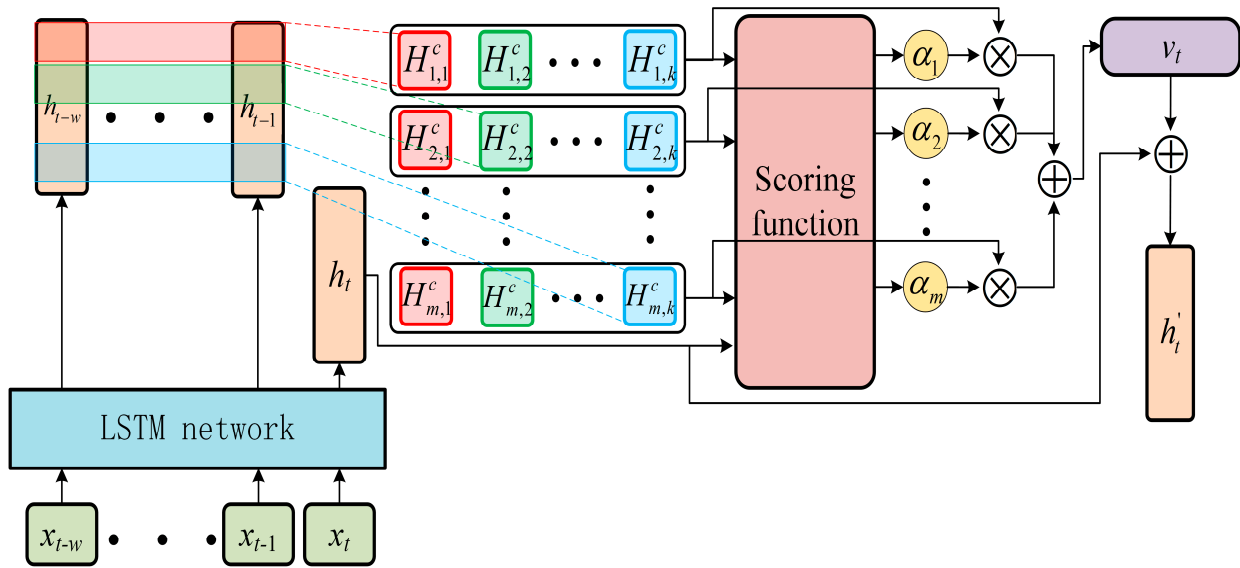


Figure 5. Schematic diagram of the TPA-LSTM.

The classic LSTM network is adopted to calculate the hidden state  $h$  with  $m$  dimensions according to the input data  $x$ . Then, a one-dimensional convolution operation with  $k$  kernels is performed on the matrix constructed with previous hidden state vectors to gain various temporal patterns, expressed as:

$$H_{i,j}^c = \sum_{n=1}^w h_{i,t-w-1+n} \times c_j \tag{23}$$

where  $w$  is the window length of the input data,  $t$  represents the time stamp, and  $c_j$  is the  $j^{th}$  1DCNN kernel.

The attention weights vector  $\alpha$  is calculated by the scoring function, which is essentially a fully connected layer.

$$\alpha_i = \text{sigmoid}(W_\alpha h_t H_i^c) \tag{24}$$

The context vector  $v_t$  and the final hidden state  $h'_t$  are defined as:

$$v_t = \sum_{n=1}^m \alpha_n H_n^c \tag{25}$$

$$h'_t = W_h h_t + W_v v_t \tag{26}$$

3. Engineering Application

To verify the effectiveness of the proposed method, a large-scale FTU in the actual engineering environment was selected as the research object. This section begins with a brief introduction to the basic information of the FTU. Then, the long-term monitoring records of the water head ( $H$ ), the active power ( $P$ ) and the vibration amplitude ( $V$ ) of

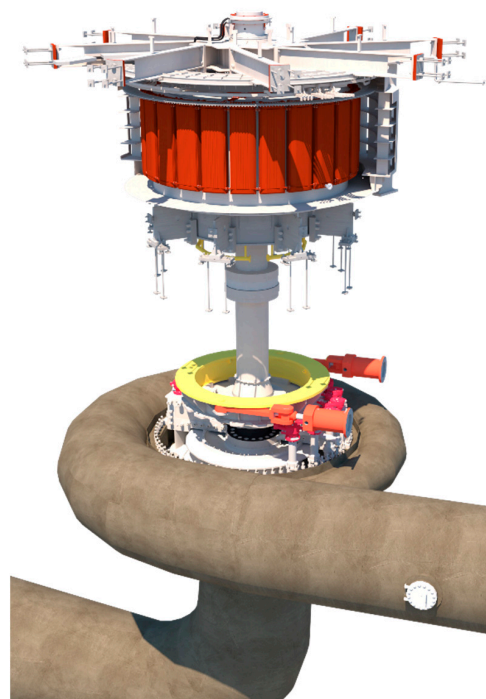
the top cover were obtained from the computer monitoring system of the hydropower station to form the raw data set. Next, anomaly samples in the raw data were removed based on DBSCAN. In addition, the GRUD–WGAN model was established to fill in the missing values. On this basis, the health state probability model of FTU was constructed based on the GPR algorithm, and the *NLLP* was defined as the PDI of the FTU. Finally, the TPA–LSTM model was built to realize the degradation trend prediction of the FTU.

### 3.1. Research Object

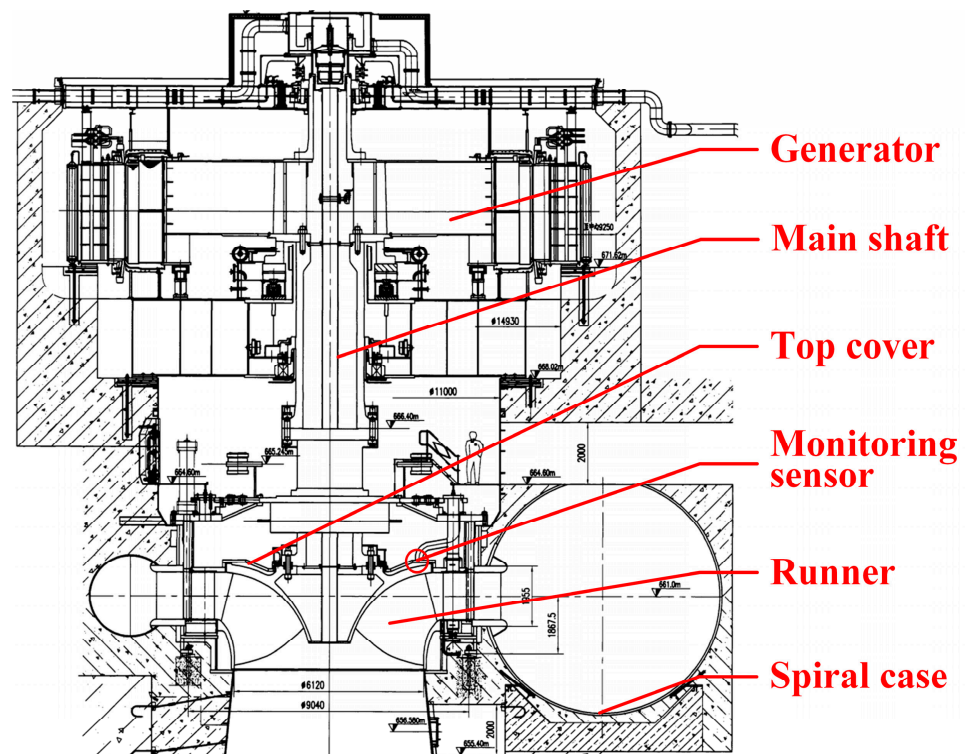
The researched FTU is located in the upper reaches of the Dadu River, Sichuan province, west of China. It is a large-capacity unit with a medium-high water head. Its essential performance parameters are listed in Table 1, and the basic structure is shown in Figure 6. The working state of the FTU is closely related to the operating parameters. Therefore, operating conditions must be considered. The operating condition of the FTU can be described by the water head ( $H$ ) and the active power ( $P$ ). The top cover is located between the turbine and the generator. It seals the runner chamber and connects the main shaft. As a critical component, its vibration amplitude ( $V$ ) can reflect the working state of the FTU. The position of the monitoring point is illustrated in Figure 7. Therefore, the raw sample set  $\Psi$  including both operating parameters and monitoring data is formed by  $(H, P, V)$ .

**Table 1.** Basic performance parameters of the FTU.

Parameters	Values	Units
Type	HLD416A-LJ-696	\
Rated power	600	MW
Rated water head	156.7	m
Rated flow rate	435	m <sup>3</sup> /s
Rated rotate speed	125	r/min
Rated efficiency	96.43	%
Inlet diameter of the runner	6.964	m
Number of runner blades	15	\
Guide vane distribution diameter	8.0	m
Number of guide vanes	24	\



**Figure 6.** The basic structure of the FTU.

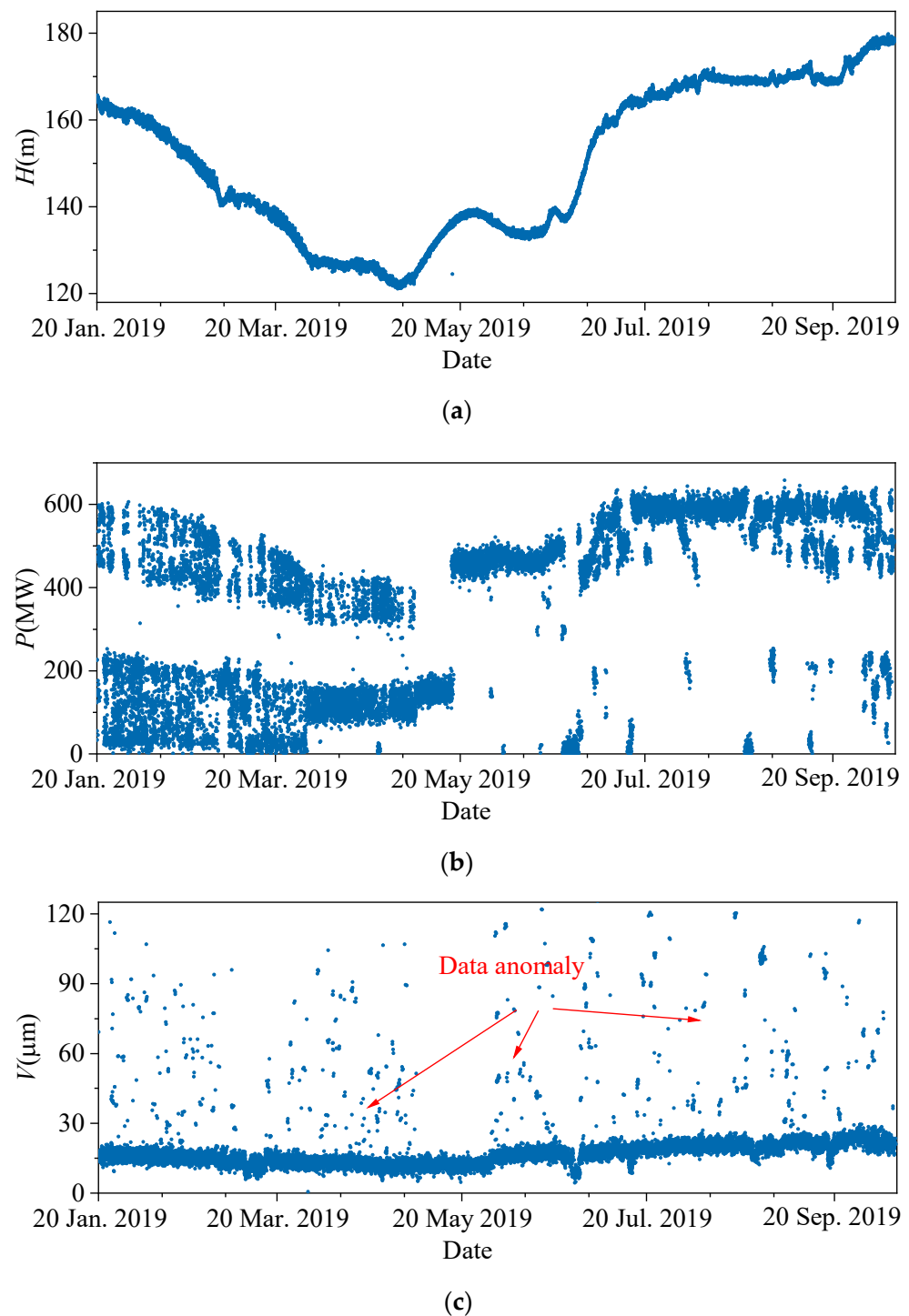


**Figure 7.** The position of the monitoring sensor.

The operating parameters of the FTU are monitored by the supervisory control system of the hydropower station. The vibration signals of critical components are acquired by a PSTA-2100 state monitoring system. They are transmitted to the computer monitoring system of the hydropower station through the TCP/IP protocol and Modbus 485 protocol, respectively. The physical distance of the transmission link is usually above several thousand meters, including multiple switches, routers, and different transceiver devices. In addition, these monitoring and communication systems work in the extreme environment of high humidity, strong vibration, and high electromagnetic interference, which may result in short-term failures. Consequently, the raw data directly exported from the computer monitoring system are often low quality, which manifests as data anomalies and data loss.

### 3.2. On-Site Data Cleaning

The acquired data include the  $(H, P, V)$  records from 20 January 2019, to 11 October 2019, and the sampling frequency is 30 min per sample, including 12,638 samples. The raw data are shown in Figure 8. There is a long-term fluctuation trend in the water head data because of the seasonal fluctuation of upstream and downstream water levels. The active power is specified by the dispatching center according to the real-time power network load demand. Hence, it has high-frequency short-term fluctuation characteristics. The vibration amplitude is affected by the variation in these operating parameters, so its variation is complicated. Therefore, the operating condition parameters should be considered in data cleaning and subsequent evaluation. In addition, the overall missing rate of the raw data is 0.243, and the data anomaly in the vibration data is obvious.



**Figure 8.** On-site raw data: (a) water head  $H$ ; (b) active power  $P$ ; and (c) vibration amplitude of the top cover  $V$ .

The raw data ( $H$ ,  $P$ ,  $V$ ) were combined into a three-dimensional point set  $\Psi$ , as shown in Figure 9. Due to the characteristics of the FTU, the oblique blank area is the restricted operating region. Usually, the FTU would avoid working in this restricted operating region because of the high vibration and low efficiency. The valid data are concentrated in the operating condition area on both sides. In addition, the anomaly data include singulars whose amplitude is different from the standard values and the outliers whose values are within the normal range, but with a distribution inconsistent with the standard values.

The dataset  $\Psi$  was inputted into the DBSCAN model, and the radius  $\epsilon$  and the minimum number of samples within cluster  $Z$  were determined by the silhouette score  $s$ , defined as:

$$s = \frac{1}{n} \sum_{i=1}^n \frac{b_i - a_i}{\max(a_i, b_i)} \tag{27}$$

where  $a_i$  is the average distance between the  $i$ th sample and other samples in the same cluster, and  $b_i$  is the average distance between the  $i$ th sample and all other samples in the nearest cluster.  $s \in [-1, 1]$  A larger  $s$  indicates that the samples within clusters are condensed, and the samples between clusters are dispersed.

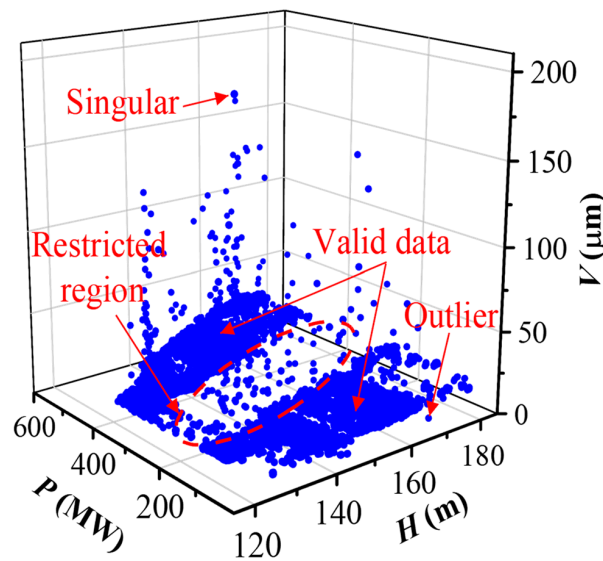


Figure 9. The three-dimensional sample set.

After several experiments,  $s$  reached its maximum of 0.63 when  $\epsilon = 7.5$ ,  $Z = 196$ . The clustering result is shown in Figure 10. The DBSCAN model effectively identified two valid data agglomerations and marked both singulars and outliers as noise points. The noise point was dropped out, and the valid data were retained, defined as the valid data set  $\Psi'$ , for subsequent analysis.

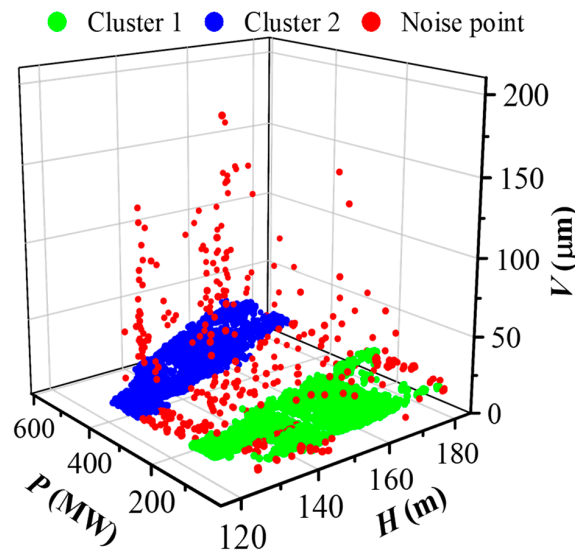
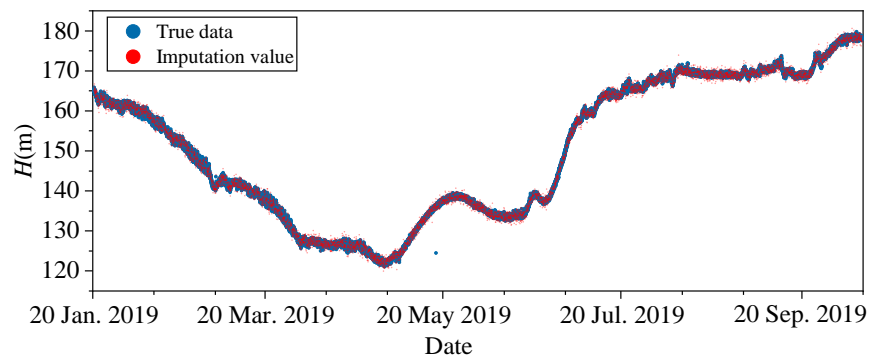


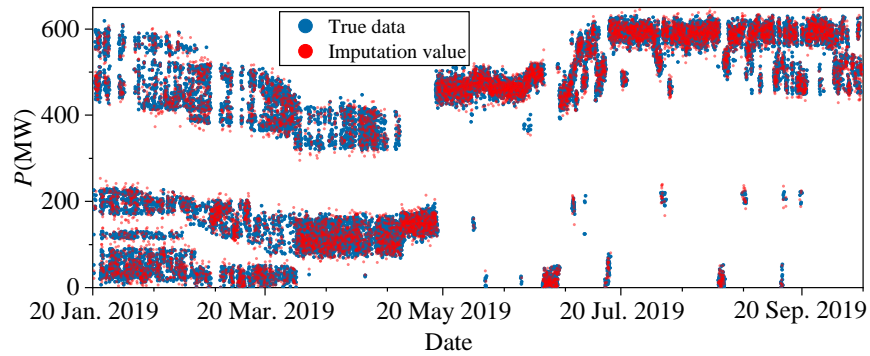
Figure 10. Clustering result of DBSCAN.

### 3.3. Missing Value Imputation

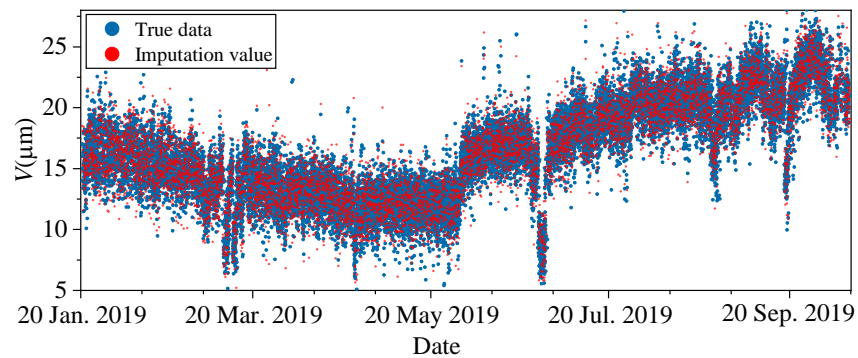
The missing data rates of the raw dataset  $\Psi$  and the cleaned valid data set  $\Psi'$  were 0.243 and 0.292, respectively. The existence of missing values greatly impacts the subsequent evaluation and prediction procedures. The  $H'$ ,  $P'$  and  $V'$  of the valid data set  $\Psi'$  were inputted into the proposed GRUD-WGAN model to fill the missing values. The main parameters are listed in Table 2. The result of data imputation is shown in Figure 11, where imputation values are marked as red points. It can be seen that the amplitudes of imputation values are similar to the actual values nearby. The distribution of the imputation values is also similar to the actual values, as shown in Figure 12. This indicates that the generator successfully learned the distribution of actual data. The complete data set after data imputation was defined as  $\Psi_C$ . The validity of the proposed GRUD-WGAN and other data imputation methods is further compared and discussed in Section 4.



(a)

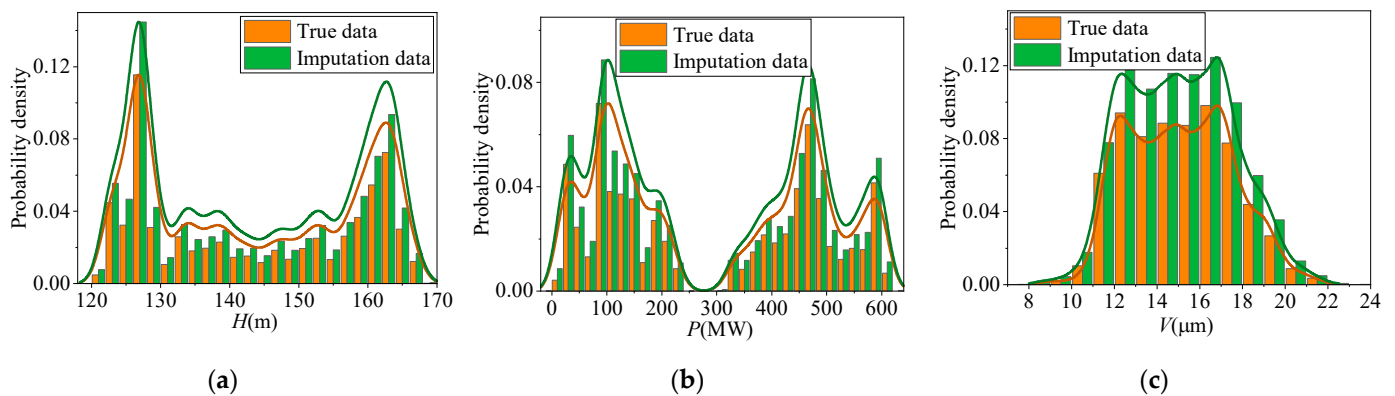


(b)



(c)

**Figure 11.** The results of data imputation: (a) water head  $H$ ; (b) active power  $P$ ; and (c) vibration amplitude of the top cover  $V$ .



**Figure 12.** The distribution of data imputation results: (a) water head  $H$ ; (b) active power  $P$ ; and (c) vibration amplitude of the top cover  $V$ .

**Table 2.** Main parameters of the GRUD-WGAN.

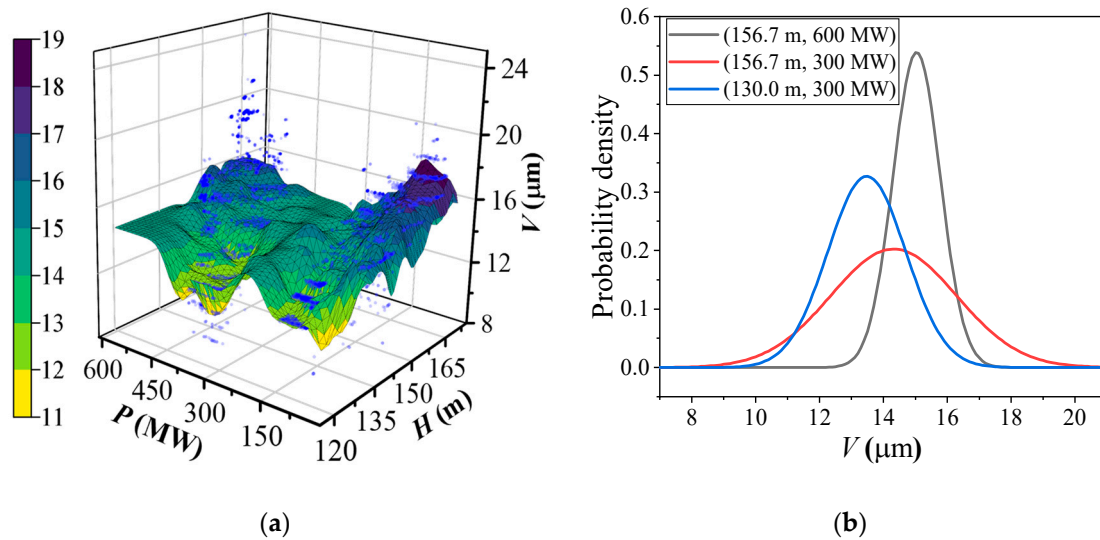
Parameter	Value
Length of the input sequence	20
Number of hidden units in the generator	32
Number of hidden units in the discriminator	32
Clipping coefficient	0.01
Epochs of training	250

### 3.4. Performance Evaluation of the FTU

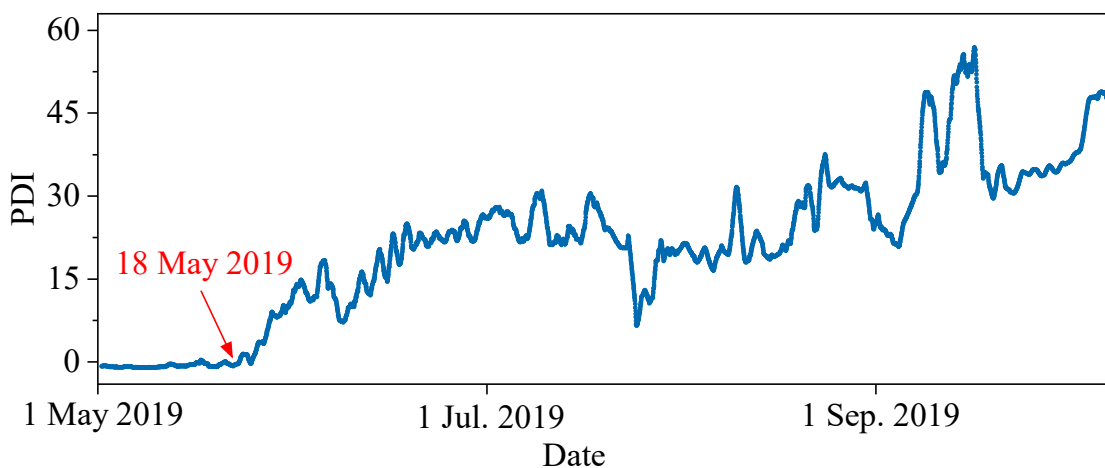
The data from 20 January 2019 to 1 May 2019 in the complete data set  $\Psi_C$  were defined as the healthy standard data set  $\Psi_H$ , including 4814 samples. The FTU was maintained before 20 January 2019, and it performed well in the restart test. Meanwhile, this period includes all possible operating conditions, especially the water head. The rest of the data were defined as the evaluated data set  $\Psi_E$ , including 7824 samples.

The operating parameters  $H_H$  and  $P_H$  in  $\Psi_H$  were selected as two independent variables, and  $V_H$  was selected as the dependent variable. The GPR algorithm was adopted to fit the mapping relationship between  $(H_H, P_H)$  and the probability density distribution of  $V_H$ , as the healthy-state model. Figure 13a shows the three-dimensional surface formed by the mean value of  $prob_{H_H, P_H}(V_H)$ . In addition, three operating conditions  $((H = 156.7 \text{ m}, P = 600 \text{ MW}), (H = 156.7 \text{ m}, P = 300 \text{ MW}), \text{ and } (H = 130.0 \text{ m}, P = 300 \text{ MW}))$  were selected as the example to draw the probability density distribution curve of  $prob_{H_H, P_H}(V_H)$ , as shown in Figure 13b. Obviously, the distribution of the vibration amplitude of the top cover is highly related to the operating condition parameters. At the rated operating condition  $(H = 156.7 \text{ m}, P = 600 \text{ MW})$ , the vibration amplitude distribution is more concentrated.

The operating parameters  $(H_E, P_E)$  of the evaluated data set  $\Psi_E$  were inputted into the constructed healthy-state model to calculate the healthy standard distribution function of the vibration amplitude  $prob_{H_E, P_E}(V)$ . Then,  $V_E$  was put into the function, and the  $NLLP = -\log(prob_{H_E, P_E}(V_E))$  was calculated as the PDI of the FTU. Moreover, considering that sufficient samples can reflect the characteristics of the probability density, one day (48 samples) was taken as the time window to generate a moving average for the calculated PDIs, and the finally obtained PDI curve is shown in Figure 14. The defined PDI based on the NLLP represents the difference in the probability density distribution between the current state and the healthy state. The PDI indicates the relative degradation trend of the FTU, so it is a dimensionless value. It can be seen that the PDI remains stable from 1 May 2019 to 18 May 2019, and the curve shows an apparent upward trend with oscillation after 18 May 2019.



**Figure 13.** The healthy-state model based on GPR: (a) the three-dimensional surface of the mean value; and (b) the standard distribution at specific conditions.



**Figure 14.** PDI curve of the FTU.

3.5. Degradation Trend Prediction of the FTU

To forecast the degradation trend of the FTU according to the historical PDIs, the TPA-LSTM model was established. The main parameters are listed in Table 3. The mean square error was selected as the loss function, and the Adam optimizer was adopted to obtain a dynamic update of the learning rate. The obtained PDI curve included 7776 points, which were divided into a training set and test set in a ratio of 7:3. The model was trained for 300 epochs and the final prediction result is shown in Figure 15.

**Table 3.** Main parameters of the TPA-LSTM.

Parameter	Value
Length of the input	12
Length of the output sequence	1
Number of hidden units in the LSTM layer	32
Number of kernels in the 1DCNN layer	10
Batch size	64



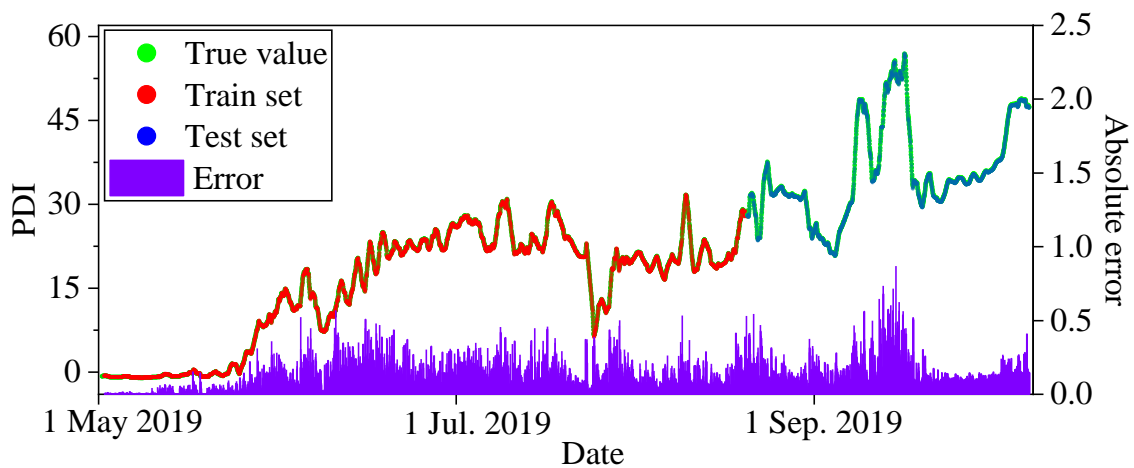


Figure 15. Prediction result of the TPA-LSTM.

The root mean square error (*RMSE*), mean absolute error (*MAE*) and  $R^2$  were selected as the metrics of the prediction result, defined as:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i^*)^2} \tag{28}$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i^*| \tag{29}$$

$$R^2 = \frac{\sum_{i=1}^N (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^N (y_i^* - \bar{y})^2} \tag{30}$$

where  $N$  is the length of the sequence,  $\hat{y}_i$  means the prediction value,  $y_i^*$  is the actual value and  $\bar{y}$  is the mean value of the actual sequence. Lower *RMSEs* and *MAEs* indicate the better accuracy of the prediction result. The  $R^2$  value is between 0 and 1. An  $R^2$  close to 1 means the correlation between the predicted sequence and the actual sequence is strong. The metrics of the prediction result are listed in Table 4. The accuracy and the correlation of the TPA-LSTM model are high in the degradation trend prediction task. The criteria of the training set and the test set are similar, indicating that the model has good generalization performance. The performances of the TPA-LSTM model and other prediction methods are further compared and discussed in Section 4.

Table 4. Criteria of the prediction result.

Data Set	RMSE	MAE	$R^2$
Training data set	0.094	0.064	0.999
Test data set	0.104	0.068	0.999

#### 4. Discussion

In order to prove the effectiveness of the proposed method in terms of missing data imputation and the degradation trend prediction of the FTU, the necessity of data imputation is analyzed first. Then, the effects of different data imputation methods on the complete measured data set are compared in this section. Next, the influence of different input sequence lengths on the GRUD-WGAN model is discussed. Then, the forecasting efficiency of different prediction models on the PDI curve is compared.

#### 4.1. Necessity Analysis of Data Imputation

To validate the necessity of the data imputation procedure, this step was left out of the approach framework, and the samples from 20 January 2019 to 1 May 2019 in the cleaned data set  $\Psi'$  were selected to construct the healthy-state model of the FTU. The valid samples of each day from 1 May 2019 to 11 October 2019 were used for evaluation. The number of samples per day was inconsistent due to the missing data, as shown in Figure 16. Missing data may make the PDI curve unstable.

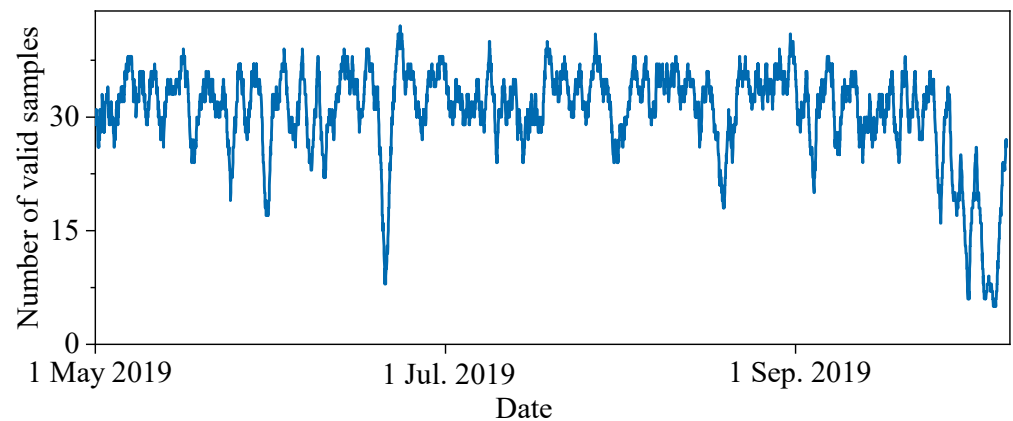


Figure 16. The number of valid samples per day.

The obtained PDI curves are shown in Figure 17. The data imputation step makes the trend of the PDI curve more obvious. In order to quantify the influence of data imputation on the PDI curve, the standard deviation (STD) and the mean absolute difference (MAD) were introduced to describe the stability and the smoothness of the PDI curve, defined as:

$$STD = \sqrt{\frac{1}{N} \sum_{i=1}^N (I_i - \bar{I})^2} \tag{31}$$

$$MAD = \frac{\sum_{i=1}^{N-1} |I_{i+1} - I_i|}{N - 1} \tag{32}$$

where  $I$  indicates the PDI value, and  $N$  represents the sample number of the PDI sequence. The smaller the  $STD$  and the  $MAD$  values are, the more significant the trend of the PDI curve is. The criteria of the results obtained by different procedures are listed in Table 5. The proposed data imputation method reduces  $STD$  and  $MAD$  on the PDI curve by 17.9% and 67.8%, respectively. Therefore, it is necessary to introduce the data imputation procedure.

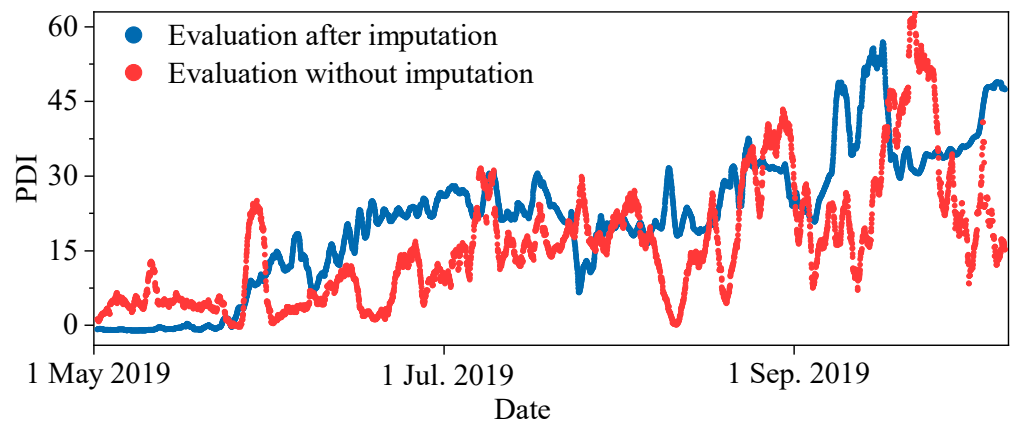


Figure 17. The influence of data imputation on PDI curve.

**Table 5.** Criteria of different procedures.

	STD	MAD
PDI curve obtained after data imputation	11.023	0.098
PDI curve obtained without data imputation	13.429	0.305

4.2. Comparison of Different Imputation Methods

To verify the efficiency of the proposed GRUD–WGAN missing data imputation model, various different methods were compared. In order to allow the imputation results to be evaluated, the complete measured water head data of a large hydropower station in the middle reaches of the Yangtze River were adopted. The data set included the daily mean value of the water head from 1 January 2014 to 1 January 2020. Various missing rates,  $r$ , of samples, including 0.1, 0.3, 0.5 and 0.7, were randomly selected from the complete dataset to construct the incomplete data set. The baseline methods are as follows:

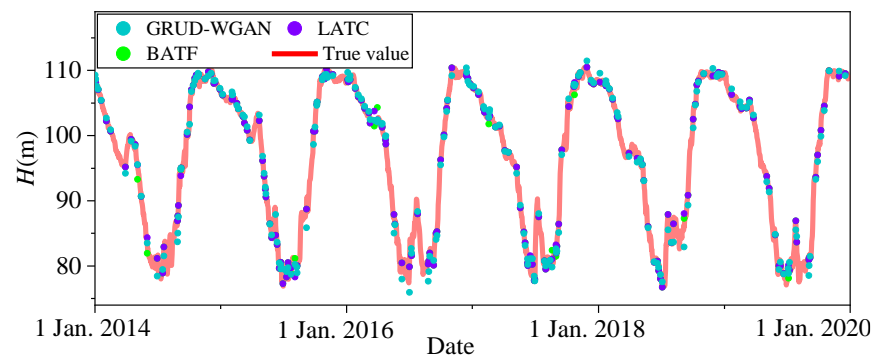
1. **Low-rank autoregressive tensor completion (LATC):** This model selects the temporal variation as a new regularization term, which makes it better able to capture the global consistency of data [40].
2. **Bayesian-augmented tensor factorization model (BATF):** This model establishes a full Bayesian framework, and the variational Bayesian algorithm is used to adaptively optimize the parameters [41]. BATF combines explicit patterns and latent factors together, and it has good generalization performance.
3. **Mean filling:** As the most straightforward statistical approach, the missing values are filled by the mean value of the sequence data.

The parameters of the GRUD–WGAN model were consistent with those in Table 3. The main configurations of the other models are listed in Table 6.

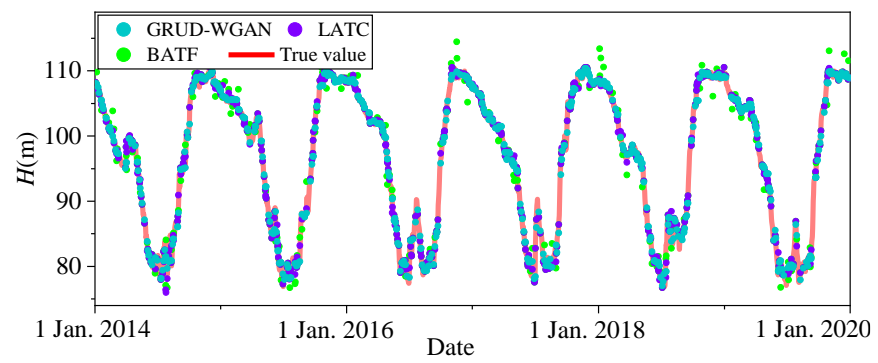
**Table 6.** Main parameters of the compared imputation methods.

Model	Parameter	Value
LATC	Weight for tensors’ nuclear norm	5
	Truncation coefficient for nuclear norm	30
	Stop tolerance	0.001
BATF	Rank of the factorization matrix	80
	Number of iterations	1000

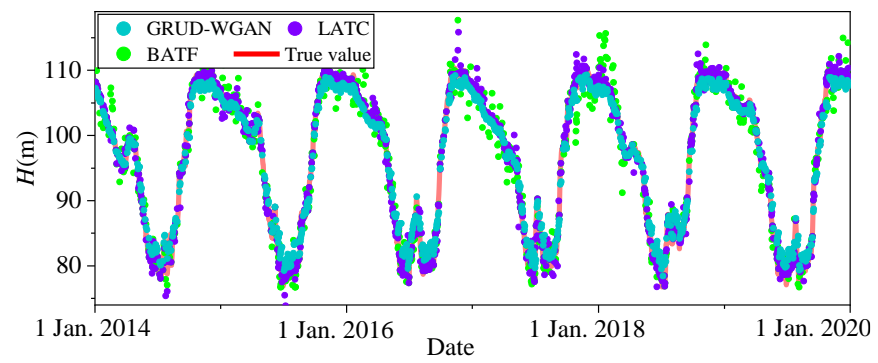
The specific imputation results of each model are shown in Figure 18. For each missing rate, the imputation values of GRUD–WGAN are closest to the actual values. The RMSEs between the imputation results and the actual values under different rates of missing data are compared in Figure 19. Under each missing rate, the RMSE of the GRUD–WGAN model achieves the lowest values, while the mean filling achieves the highest. The imputation RMSEs of all the compared methods increase along with the rise in the missing rate  $r$ , and the GRUD–WGAN model has the most minor error increase amplitude. This suggests that the proposed GRUD–WGAN has the highest accuracy and robustness among all the compared models.



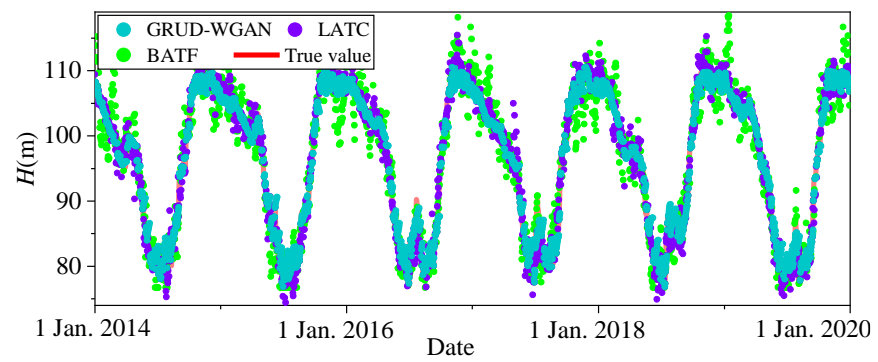
(a)



(b)



(c)



(d)

**Figure 18.** The result of data imputation with various missing rates: (a)  $r = 0.1$ ; (b)  $r = 0.3$ ; (c)  $r = 0.5$ ; and (d)  $r = 0.7$ .

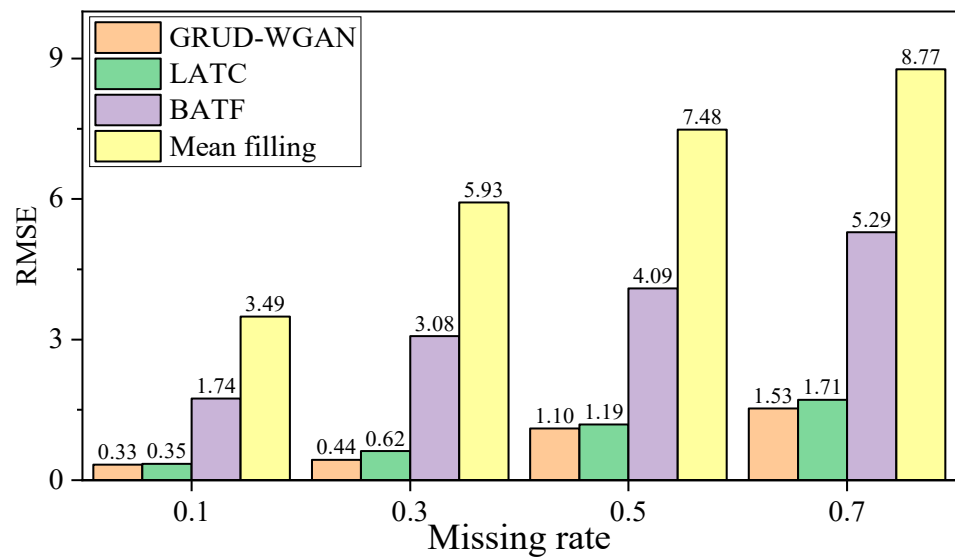


Figure 19. RMSEs of different data imputation models.

4.3. Influence of the Input Length

In the GRUD-WGAN model, the GRUD, which is the core component of both the generator and discriminator, receives sequences of length  $l$  and mines the temporal dependencies in the inputs. Hence, the parameter  $l$  determines the width of the receptive field of the GRUD model. Figure 20 shows the influence of different input length  $l$  on the GRUD-WGAN model at various missing rates. Obviously, when the missing rate is low ( $r = 0.1$ ), the filling error decreases with the increase in  $l$ , because longer input sequences contain more valid information. However, when the missing rate increases ( $r = 0.3, 0.5, 0.7$ ), the imputation accuracy of the GRUD-WGAN increases first and then decreases with the increase in  $l$ , reaching the minimum value at  $l = 20$ . This is because the input of too many missing values affects the learning of potential patterns. It is also worth noting that the training time cost rises with the increase in  $l$ . Hence, in order to improve the accuracy and efficiency of the model, it is critical to select the appropriate input length  $l$ .

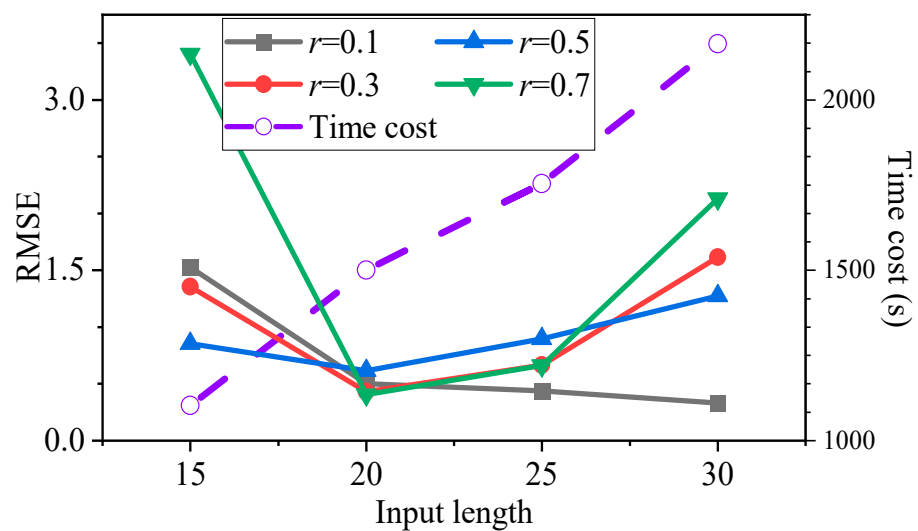


Figure 20. RMSEs of different input lengths.

4.4. Comparison of Different Prediction Methods

To validate the effectiveness of the PDI forecasting model based on the TPA-LSTM, various prediction methods were compared:

1. **RNN:** The classic RNN is the basic framework of other recurrent models. For each timestamp, the structure and the parameters of the RNN cell are shared [42].
    - **LSTM:** This model is an improved version based on the RNN. The LSTM calculates the hidden state for the long-term memory and the cell state for the short-term memory. The input gate, the forget gate and the output gate are introduced to control the transmission of information [43].
  2. **GRU:** This model is a variant of LSTM and has a simpler structure. The forget gate and the input gate are merged to form the update gate. Hence, the calculation efficiency of the GRU is relatively higher [44].
  3. **Support vector regression (SVR):** This model reflects the low-dimensional time series data to the high-dimensional feature space through the kernel function. Then, the regression result with the minimum error is searched through iterations [45].
- The main parameters of the above models are listed in Table 7.

Table 7. Main parameters of the compared prediction methods.

Model	Parameter	Value
RNN	Number of hidden units	32
	Number of hidden layers	2
LSTM	Number of hidden units	32
	Number of hidden layers	2
GRU	Number of hidden units	32
	Number of hidden layers	2
SVR	Regularization coefficient	10
	Epsilon	0.01

The next sample was predicted with every 12 previous points. The prediction results of the above methods are shown in Figure 21. Additionally, the criteria are listed in Table 8. For this single-step time series prediction task, all the compared models achieved good results on the training set. Most of the methods also performed well in the test set, except for the SVR. Due to the introduction of the temporal pattern attention mechanism, the TPA-LSTM achieved the best results on both the training set and test set, which indicates that the TPA-LSTM has better accuracy, as well as a better generalization ability. However, due to the more complex structure, TPA-LSTM is more time-consuming.

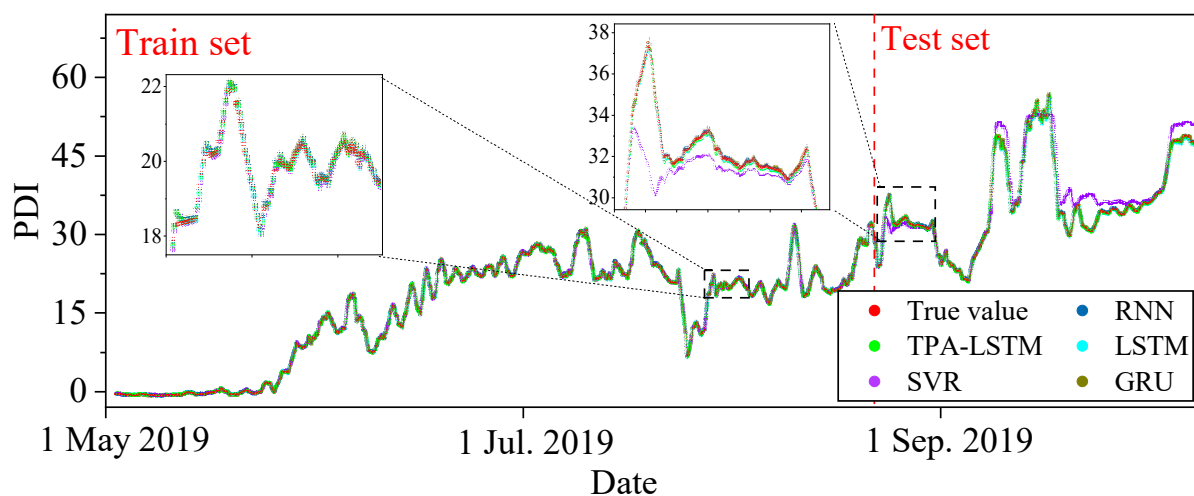


Figure 21. Prediction results of the compared models.

**Table 8.** Criteria of different prediction models.

Model	Train Set			Test Set			Time Cost (s)
	RMSE	MAE	$R^2$	RMSE	MAE	$R^2$	
TPA-LSTM	0.094	0.064	0.999	0.104	0.068	0.999	108.36
RNN	0.111	0.080	0.999	0.133	0.097	0.997	48.36
LSTM	0.099	0.067	0.998	0.284	0.214	0.998	49.82
GRU	0.096	0.066	0.999	0.136	0.082	0.997	48.07
SVR	0.098	0.065	0.999	2.589	1.703	0.908	2.3

## 5. Conclusions

Focusing on the practical problems of low-quality data and the frequently changing operating conditions of the fields of engineering applications of the FTU, an approach to the performance evaluation and prediction of the FTU considering low-quality data and variable operating conditions is proposed in this study. First, the on-site data set is constructed by the operating parameters and the vibration amplitude, and the DBSCAN algorithm is adopted to clean the anomaly data under variable operating conditions. Second, combining the incomplete sequence information mining ability of the GRUD and the hidden pattern learning ability of the WGAN, the GRUD-WGAN based missing value imputation model is proposed to improve the low-quality data utilization value. Third, the probability healthy-state model of the FTU is constructed based on the GPR to reduce the impact of data randomness. Additionally, the NLLP is calculated as the PDI of the FTU. Fourth, the degradation trend prediction model of the FTU is established based on the TPA-LSTM. Finally, a set of comparison experiments were carried out. The verification results demonstrate that the proposed data imputation method enhances the stability and the smoothness of the obtained PDI curve. Among the compared methods, the proposed GRUD-WGAN for data imputation has the highest accuracy at each experimental rate of missing data,  $r$ . In addition, when  $r = 0.1$ , the accuracy of GRUD-WGAN rises with the increase in the input length  $l$ . When  $r \geq 0.3$ , the imputation accuracy reaches the maximum while  $l = 20$ . In addition, the constructed prediction model based on TPA-LSTM achieves the lowest RMSE and MAE, and the highest  $R^2$  on both the training set and test set, indicating that the model has good accuracy and generalization performance.

The relative trend of the current state of the FTU against the healthy standard state is identified in this study. In the next phase of our research, if the long-term maintenance records can be obtained, the PDI curve can be correlated with the actual state of the FTU. Furthermore, a multistage degradation alarm model based on the PDI values can be constructed, so as to lay the foundation for state-based maintenance.

**Author Contributions:** Conceptualization, R.D.; Data curation, Y.L.; Formal analysis, J.L.; Funding acquisition, J.Z.; Investigation, P.W.; Methodology, R.D.; Project administration, J.Z.; Resources, J.Z.; Validation, J.L., P.W. and X.N.; Visualization, Y.L.; Writing—original draft, R.D.; Writing—review and editing, J.L., J.Z. and X.N. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the key project of the Natural Science Foundation of China (U1865202) and the Fundamental Research Funds for the Central Universities (HUST: 2020JYCXJJ046).

**Institutional Review Board Statement:** The study does not require ethical approval.

**Informed Consent Statement:** The study does not involve humans.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

FTU	Francis turbine unit
DBSCAN	Density-based spatial clustering of applications with noise
RNN	Recurrent neural networks
GRU	Gate recurrent unit model
GRUD	Gate recurrent unit model with decay mechanism
WGAN	Wasserstein generative adversarial network
GPR	Gaussian process regression
PDI	Performance degradation indicator
LSTM	Long short-term memory
TPA-LSTM	Temporal pattern attention-long short-term memory
NLLP	Negative log-likelihood probability
RMSE	Root mean square error
MAE	Mean absolute error
STD	Standard deviation
MAD	Mean absolute difference
LATC	Low-rank autoregressive tensor completion
BATF	Bayesian augmented tensor factorization model

## References

- Huang, H.; Qin, A.; Mao, H.; Fu, J.; Huang, Z.; Yang, Y.; Li, X.; Huang, H. The Prediction Method on the Early Failure of Hydropower Units Based on Gaussian Process Regression Driven by Monitoring Data. *Appl. Sci.* **2021**, *11*, 153. [[CrossRef](#)]
- Cordova, M.M.; Finardi, E.C.; Ribas, F.A.C.; de Matos, V.L.; Scuzziato, M.R. Performance evaluation and energy production optimization in the real-time operation of hydropower plants. *Electr. Power Syst. Res.* **2014**, *116*, 201–207. [[CrossRef](#)]
- Li, H.; Xu, B.; Riasi, A.; Szulc, P.; Chen, D.; M'Zoughi, F.; Skjelbred, H.I.; Kong, J.; Tazraei, P. Performance evaluation in enabling safety for a hydropower generation system. *Renew. Energy* **2019**, *143*, 1628–1642. [[CrossRef](#)]
- Xu, B.; Chen, D.; Li, H.; Zhuang, K.; Hu, X.; Li, J.; Skjelbred, H.I.; Kong, J.; Patelli, E. Priority analysis for risk factors of equipment in a hydraulic turbine generator unit. *J. Loss Prev. Process Ind.* **2019**, *58*, 1–7. [[CrossRef](#)]
- He, Z.Y.; Shao, H.D.; Ding, Z.Y.; Jiang, H.K.; Cheng, J.S. Modified Deep Autoencoder Driven by Multisource Parameters for Fault Transfer Prognosis of Aeroengine. *IEEE Trans. Ind. Electron.* **2022**, *69*, 845–855. [[CrossRef](#)]
- He, Z.Y.; Shao, H.D.; Zhong, X.; Zhao, X.Z. Ensemble transfer CNNs driven by multi-channel signals for fault diagnosis of rotating machinery cross working conditions. *Knowl.-Based Syst.* **2020**, *207*, 106396. [[CrossRef](#)]
- Duan, R.; Liu, J.; Zhou, J.Z.; Wang, P.; Liu, W. An Ensemble Prognostic Method of Francis Turbine Units Using Low-Quality Data under Variable Operating Conditions. *Sensors* **2022**, *22*, 525. [[CrossRef](#)]
- Guo, L.; Yu, Y.; Duan, A.; Gao, H.; Zhang, J. An unsupervised feature learning based health indicator construction method for performance assessment of machines. *Mech. Syst. Signal Process.* **2022**, *167*, 108573. [[CrossRef](#)]
- Qin, Y.; Wu, X.; Luo, J. Data-Model Combined Driven Digital Twin of Life-Cycle Rolling Bearing. *IEEE Trans. Ind. Inform.* **2022**, *18*, 1530–1540. [[CrossRef](#)]
- Ding, P.; Jia, M. Mechatronics Equipment Performance Degradation Assessment Using Limited and Unlabeled Data. *IEEE Trans. Ind. Inform.* **2022**, *18*, 2374–2385. [[CrossRef](#)]
- Li, H.; Liu, T.; Wu, X.; Chen, Q. A Bearing Fault Diagnosis Method Based on Enhanced Singular Value Decomposition. *IEEE Trans. Ind. Inform.* **2021**, *17*, 3220–3230. [[CrossRef](#)]
- Li, Y.; Cheng, G.; Liu, C. Research on bearing fault diagnosis based on spectrum characteristics under strong noise interference. *Measurement* **2021**, *169*, 108509. [[CrossRef](#)]
- Wang, Q.; Wang, L.; Yu, H.; Wang, D.; Nandi, A.K. Utilizing SVD and VMD for Denoising Non-Stationary Signals of Roller Bearings. *Sensors* **2022**, *22*, 195. [[CrossRef](#)]
- Wang, W.; Hu, X.; Du, Y. Algorithm optimization and anomaly detection simulation based on extended Jarvis-Patrick clustering and outlier detection. *Alex. Eng. J.* **2022**, *61*, 2106–2115. [[CrossRef](#)]
- Chen, H.; Ma, H.; Chu, X.; Xue, D. Anomaly detection and critical attributes identification for products with multiple operating conditions based on isolation forest. *Adv. Eng. Inform.* **2020**, *46*, 101139. [[CrossRef](#)]
- Liu, X.; Lu, S.; Ren, Y.; Wu, Z. Wind Turbine Anomaly Detection Based on SCADA Data Mining. *Electronics* **2020**, *9*, 751. [[CrossRef](#)]
- Chen, Y.; Tang, S.; Bouguila, N.; Wang, C.; Du, J.; Li, H. A fast clustering algorithm based on pruning unnecessary distance computations in DBSCAN for high-dimensional data. *Pattern Recognit.* **2018**, *83*, 375–387. [[CrossRef](#)]
- Li, S. An Improved DBSCAN Algorithm Based on the Neighbor Similarity and Fast Nearest Neighbor Query. *IEEE Access* **2020**, *8*, 47468–47476. [[CrossRef](#)]
- Nasibov, E.N.; Ulutagay, G. Robustness of density-based clustering methods with various neighborhood relations. *Fuzzy Sets Syst.* **2009**, *160*, 3601–3615. [[CrossRef](#)]



20. Luchi, D.; Rodrigues, A.L.; Varejao, F.M. Sampling approaches for applying DBSCAN to large datasets. *Pattern Recognit. Lett.* **2019**, *117*, 90–96. [[CrossRef](#)]
21. Kim, H.; Jang, G.; Choi, H.; Lim, M.; Choi, J. Medical examination data prediction with missing information imputation based on recurrent neural networks. *Int. J. Data Min. Bioinform.* **2018**, *19*, 202–220. [[CrossRef](#)]
22. Cui, Z.; Ke, R.; Pu, Z.; Wang, Y. Stacked bidirectional and unidirectional LSTM recurrent neural network for forecasting network-wide traffic state with missing values. *Transp. Res. Part C Emerg. Technol.* **2020**, *118*, 102674. [[CrossRef](#)]
23. Zhang, J.; Mu, X.; Fang, J.; Yang, Y. Time Series Imputation via Integration of Revealed Information Based on the Residual Shortcut Connection. *IEEE Access* **2019**, *7*, 102397–102405. [[CrossRef](#)]
24. Che, Z.; Purushotham, S.; Cho, K.; Sontag, D.; Liu, Y. Recurrent Neural Networks for Multivariate Time Series with Missing Values. *Sci. Rep.* **2018**, *8*, 6085. [[CrossRef](#)]
25. Pham, Q.T.M.; Ahn, S.; Shin, J.; Song, S.J. Generating future fundus images for early age-related macular degeneration based on generative adversarial networks. *Comput. Methods Programs Biomed.* **2022**, *216*, 106648. [[CrossRef](#)]
26. Saxena, D.; Cao, J. Generative Adversarial Networks (GANs): Challenges, Solutions, and Future Directions. *ACM Comput. Surv.* **2022**, *54*, 1–42. [[CrossRef](#)]
27. Baskerville, N.P.; Keating, J.P.; Mezzadri, F.; Najnudel, J. A Spin Glass Model for the Loss Surfaces of Generative Adversarial Networks. *J. Stat. Phys.* **2022**, *186*, 29. [[CrossRef](#)]
28. Tsialiamanis, G.; Champneys, M.D.; Dervilis, N.; Wagg, D.J.; Worden, K. On the application of generative adversarial networks for nonlinear modal analysis. *Mech. Syst. Signal Process.* **2022**, *166*, 108473. [[CrossRef](#)]
29. Ma, Y.; Zhong, P.; Xu, B.; Zhu, F.; Yang, L.; Wang, H.; Lu, Q. Stochastic generation of runoff series for multiple reservoirs based on generative adversarial networks. *J. Hydrol.* **2022**, *605*, 127326. [[CrossRef](#)]
30. Fan, X.; Zhang, W.; Sun, B.; Zhang, J.; He, X. Battery pack consistency modeling based on generative adversarial networks. *Energy* **2022**, *239*, 122419. [[CrossRef](#)]
31. Ishaan, G.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; Courville, A. Improved Training of Wasserstein GANs. In *Advances in Neural Information Processing Systems 30 (NIPS 2017), Proceedings of the 31st Annual Conference on Neural Information Processing Systems (NIPS), Long Beach, CA, USA, 4–9 December 2017*; Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R., Eds.; MIT Press: Long Beach CA, USA, 2017; Volume 30.
32. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein Generative Adversarial Networks. In *International Conference on Machine Learning, Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017*; Precup, D., Teh, Y.W., Eds.; PMLR: Sydney, Australia, 2017; Volume 70.
33. Mao, G.; Wang, S.; Teng, Q.; Zuo, J.; Tan, X.; Wang, H.; Liu, Z. The sustainable future of hydropower: A critical analysis of cooling units via the Theory of Inventive Problem Solving and Life Cycle Assessment methods. *J. Clean. Prod.* **2017**, *142*, 2446–2453. [[CrossRef](#)]
34. Shan, Y.; Liu, J.; Xu, Y.; Zhou, J. A combined multi-objective optimization model for degradation trend prediction of pumped storage unit. *Measurement* **2021**, *169*, 108373. [[CrossRef](#)]
35. Li, G.; Zhao, X.; Fan, C.; Fang, X.; Li, F.; Wu, Y. Assessment of long short-term memory and its modifications for enhanced short-term building energy predictions. *J. Build. Eng.* **2021**, *43*, 103182. [[CrossRef](#)]
36. Sayah, M.; Guebli, D.; Al Masry, Z.; Zerhouni, N. Robustness testing framework for RUL prediction Deep LSTM networks. *ISA Trans.* **2021**, *113*, 28–38. [[CrossRef](#)]
37. Gao, S.; Huang, Y.; Zhang, S.; Han, J.; Wang, G.; Zhang, M.; Lin, Q. Short-term runoff prediction with GRU and LSTM networks without requiring time step optimization during sample generation. *J. Hydrol.* **2020**, *589*, 125188. [[CrossRef](#)]
38. Wu, Q.; Jiang, Z.; Hong, K.; Liu, H.; Yang, L.T.; Ding, J. Tensor-Based Recurrent Neural Network and Multi-Modal Prediction With Its Applications in Traffic Network Management. *IEEE Trans. Netw. Serv. Manag.* **2021**, *18*, 780–792. [[CrossRef](#)]
39. Shih, S.; Sun, F.; Lee, H. Temporal pattern attention for multivariate time series forecasting. *Mach. Learn.* **2019**, *108*, 1421–1441. [[CrossRef](#)]
40. Chen, X.; Lei, M.; Saunier, N.; Sun, L. Low-Rank Autoregressive Tensor Completion for Spatiotemporal Traffic Data Imputation. *IEEE Trans. Intell. Transp. Syst.* **2021**. early access. [[CrossRef](#)]
41. Chen, X.; He, Z.; Chen, Y.; Lu, Y.; Wang, J. Missing traffic data imputation and pattern discovery with a Bayesian augmented tensor factorization model. *Transp. Res. Part C Emerg. Technol.* **2019**, *104*, 66–77. [[CrossRef](#)]
42. Zhao, R.; Yan, R.; Chen, Z.; Mao, K.; Wang, P.; Gao, R.X. Deep learning and its applications to machine health monitoring. *Mech. Syst. Signal Process.* **2019**, *115*, 213–237. [[CrossRef](#)]
43. Yan, H.; Qin, Y.; Xiang, S.; Wang, Y.; Chen, H. Long-term gear life prediction based on ordered neurons LSTM neural networks. *Measurement* **2020**, *165*, 108205. [[CrossRef](#)]
44. Dai, G.; Ma, C.; Xu, X. Short-Term Traffic Flow Prediction Method for Urban Road Sections Based on SpaceTime Analysis and GRU. *IEEE Access* **2019**, *7*, 143025–143035. [[CrossRef](#)]
45. Nguyen, H.; Choi, Y.; Bui, X.N.; Trung, N.T. Predicting Blast-Induced Ground Vibration in Open-Pit Mines Using Vibration Sensors and Support Vector Regression-Based Optimization Algorithms. *Sensors* **2020**, *20*, 132. [[CrossRef](#)]