Reliability Engineering and System Safety 234 (2023) 109162

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

Adaptive transfer learning for multimode process monitoring and unsupervised anomaly detection in steam turbines

Zhen Chen [a,d], Di Zhou [b], Enrico Zio [c,d], Tangbin Xia [a], Ershun Pan [a,*]

[a] State Key Laboratory of Mechanical System and Vibration, Department of Industrial Engineering & Management, Shanghai Jiao Tong University, Shanghai 200240, China

[b] College of Mechanical Engineering, Donghua University, Shanghai 200051, China

[c] Centre for Research on Risk and Crises, Ecole des Mines, ParisTech, PSL University, 75006 Paris, France

[d] Department of Energy, Politecnico di Milano, 20133 Milan, Italy

ARTICLE INFO

Keywords:
Condition monitoring
Unsupervised anomaly detection
Multimode
Transfer learning
Steam turbine

ABSTRACT

Through condition-based maintenance strategy, engineers can monitor the health states of equipment and take actions based on the sensor data. Limited by the low failure frequency and high monitoring costs, it is difficult to obtain sufficient historical data of all fault types for condition monitoring (CM). In the steam turbine operation, environmental factors, varying power consumption and manual adjustments can lead to a multimode process, which consists of multiple normal and abnormal conditions. This paper proposes a framework for online unsupervised CM and anomaly detection, not relying on expert knowledge or labeled historical data. Since there are often few monitoring data at the beginning of a new incoming operating mode, an adaptive self-transfer learning algorithm based on Gaussian processes is developed to model the monitoring data with uncertainty information, and to capture the cross-correlations between the different normal modes. A two-hierarchical identification criterion based on the predicted posterior intervals is introduced to first identify the change-points in the observations, and second to decide whether it is an anomaly or a transition between normal modes. The proposed framework is tested on a real steam turbine. The results illustrate its high effectiveness.

1. Introduction

Thermal power plants offer the primary source of electricity supply in the world. The occurrence of unexpected faults may cause the shutdown of power plants, resulting in economic loss and safety issues. As crucial equipment in power plants, steam turbine operation and maintenance have an important impact on power generation reliability, efficiency, and stability. Recently, condition monitoring (CM) has become essential to identify potential faults and reduce maintenance costs of much equipment [1–5].

A steam turbine is a system that converts the heat energy of hot steam into rotational mechanical energy [6,7]. The monitoring system integrates sensing devices in what is usually called the turbine supervisory instrumentation (TSI) system, as shown in Fig. 1. As the steam turbine operation is a multivariate dynamic process, various components are monitored and multiple parameters or variables are collected by the TSI, such as environmental parameters, vibration signals, temperatures, and control variables. The TSI data provide a cost-effective approach for the CM of steam turbines, without the need to introduce additional sensors. A successful CM should timely detect the turbine faults, based on the data collected by the TSI system.

However, a low sampling rate of the monitoring data impedes the use of many common signal processing techniques, such as spectral analysis [8]. Igor et al. [9] developed a machine-learning-based framework for detecting and classifying several fault types in a power-generation system. Since steam turbine faults occur randomly and the type of faults is diverse, it is difficult and expensive to collect sufficient historical fault data, and representative for all types of faults [10,11]. Therefore, it is necessary to develop anomaly detection techniques that do not rely on historical labeled data [12,13].

Generally, equipment anomalies produce patterns in data that deviate from well-defined normal conditions [14]. Anomaly detection is then the task of distinguishing abnormal and normal condition at the early stage [15–17]. It is also less reliant on labeled data and expert knowledge. A supervised model which needed sufficient normal or abnormal condition data tends to be impractical, as some abnormal data

* Corresponding author.
E-mail address: pes@sjtu.edu.cn (E. Pan).

https://doi.org/10.1016/j.ress.2023.109162
Received 24 May 2022; Received in revised form 10 February 2023; Accepted 11 February 2023
Available online 12 February 2023

are very rare. Generally, it is difficult to collect enough amount of abnormal data with limited cost and technical resources. Additionally, some abnormal conditions are unavailable before they occur [18]. A better anomaly detection method should develop a model that represents the normal operating conditions and detect the anomalies online once the system's condition significantly deviates from the normal prediction. Obviously, such unsupervised methods can perform anomaly detection in unlabeled data [19]. Therefore, the unsupervised anomaly detection method which is more suitable for online applications with only normal data will be studied in the following. Recently, many papers developed unsupervised anomaly detection methods based on residual analysis [20–23]. Due to the stochastic environmental disturbances, inevitably there are uncertainties in the practical monitoring process, monitoring data and the estimated residuals [24,25]. If these uncertainties are not quickly investigated, the accuracy of anomaly detection will be reduced significantly. Compared with many residual-analysis-based methods that provide deterministic estimation in the control variables without considering these uncertainties, a probabilistic method quantifies these uncertainties and develops the anomaly detection criterion based on the estimated intervals serves as a more accurate monitoring approach, which tends to identify the faults timely [26].

On the other hand, there are autocorrelations or temporal dependency in the monitoring data [27]. A large number of machine learning algorithms and signal processing techniques [28–30] have been proposed to account for this data autocorrelation for industrial equipment condition monitoring. Gaussian process (GP) has been used to develop anomaly detection methods for wind farm [31] and power plant [10] condition monitoring. Deep learning networks, such as deep auto-encoder networks, have been developed in [32–34] for the anomaly detection of wind turbines and steam turbines. Zhang et al. [35] proposed an anomaly detection and diagnosis method for wind turbines using long short-term memory-based stacked denoising autoencoders and extreme gradient boosting. Zhang et al. [36] developed an unsupervised, end-to-end approach to fault detection based on a flow-based model without the labels of faulty samples. A successful anomaly detection method depends on the accurate modeling of the normal operating conditions. For this, it is necessary to consider the associated uncertainties and autocorrelations in the monitoring data.

In the operating process of a steam turbine, there are usually multiple normal operating modes, due to electric grid fluctuations and manual adjustments [37]. The data characteristics of different modes are different. If one specific model is applied to different modes, it can result in unexpected false alarms and missing alarms. As most existing methods assume the monitoring data are generated from a single normal operating mode and the anomaly deviates only from this mode, they have been found to be not well-suited in practice [38]. To monitor the operating process and detect anomalies accurately, multimode process monitoring models have been constructed [38,39]. Peng et al. [40] proposed a multiple partial least square-based method to address the muti-mode problem in hot strip mill process with the Gaussian mixture model and achieve the quality prediction and monitoring goals. Quatrini et al. [41] developed a two-step method for anomaly detection in industrial processes with multiple phases based on machine learning classification algorithms.

Most of these methods are supervised learning methods, i.e., based on labeled data, where the number of normal modes is known and the training data are labeled for each mode. Since the operating modes are not fixed and the labeled data for each mode are difficult to collect, an unsupervised anomaly detection framework for the multimode process monitoring is more realistic. Moreover, for a running steam turbine, the historical normal data of each mode from identical system may be insufficient. Especially when a new normal mode comes, there is no any information. This makes it difficult to accurately implement mode identification and anomaly detection. Considering that cross-correlations and common attributes (invariant system structure and robust material performance) can exist in different modes due to the inherent attributes of steam turbines, the transfer learning which focuses on transferring the knowledge across domains [42,43] is employed. Then, the data from the past normal modes can be transferred to constructed the monitoring models for new modes. Therefore, the problem of data shortage can be solved effectively.

This study proposes an adaptive unsupervised anomaly detection framework for steam turbine condition monitoring. The main contributions are:

1) For the multimode operation processes of steam turbines, a transfer learning framework based on multiple GPs is developed to automatically transfer knowledge across modes, which solve the problem of data shortage at the beginning stage of each new mode. Different from the transfer learning with shared parameters, we learn a transfer kernel that can be a similarity measure to describe the cross-correlation of different modes more flexibly.
2) To achieve the goal of adaptive transfer, an adaptive self-transfer learning algorithm is developed to inferring multiple GPs, which not only yields deterministic predictions but also capture uncertainties and the nonlinear data autocorrelations. As the past mode is used as priori, the computational complexity for the GP of a new mode becomes lower.
3) To avoid false alarm and simultaneously detect anomaly and normal mode transition, a two-hierarchical identification criterion based on the predicted posterior intervals is proposed. It makes the proposed framework capable of an automatic online detection activity to timely identify different normal modes and anomalies.
4) An online unsupervised anomaly detection framework for the multimode process monitoring of steam turbine is built without labeled data and abnormal data, where a joint algorithm of data standardization and early stopping is developed to improve the efficiency. More reliable monitoring results are obtained and the potential faults are detected early.

The paper is arranged as follows. Section 2 provides a description of anomaly detection and stream turbine operating data. Section 3 introduces the adaptive transfer learning method. Section 4 proposes an online CM framework. Section 5 presents a real case study of steam turbine. Conclusions are drawn in Section 6.

2. Problem description

The TSI data includes various kinds of monitoring variables or

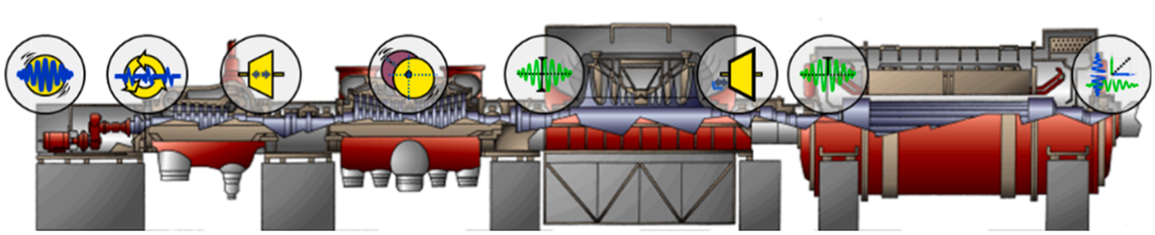

Fig. 1. The turbine supervisory instrumentation (TSI) system for steam turbine condition monitoring.

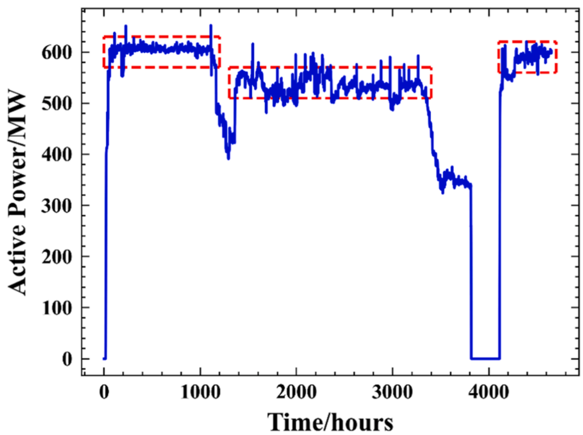

Fig. 2. Active power of a steam turbine generator during 7 months.

parameters. Considering the practical fault types and engineering experience, we only select the temperature signals here to conduct the anomaly detection in this study. Temperatures are significant and easily measured by TSI. They are also indicators of the health states of many steam turbine components, such as bearings, bushes and rotors. An unexpected rise in component temperatures can indicate poor lubrication, abrasion, overload, or cracking [44]. Bae et al. [45] proposed a CM scheme applying energy spectrums such as a wavelet analysis to only temperature signals for early abnormality detection of a steam turbine generator. In the power plants, when the incoming monitored temperatures are outside the confidence interval predicted as normal by the monitoring model, it would be probably construed as an early sign of a potential fault [10,46]. Then an anomaly is flagged.

In different seasons and weather, or even during the day and night, the power consumption can vary over wide ranges. Including the manual adjustments for mitigating the deterioration of equipment, the operating modes of steam turbines change to adapt to the different situations. Fig. 2 shows the active power profile of a steam turbine generator in a thermal power plant for 7 months. The rated power is 610 MW. Fig. 3 displays the metal temperatures of the #1 bearing bushes for 7 months. Obviously, there exists multiple different modes in the steam turbine operation. At the beginning, the steam turbine got its annual overhaul. Hence, both the active power and the temperature signal were close to their normal values in the complete health state. Then, due to the manual adjustment and system degradation, the active power declined and the temperature signal increased, which indicated that the steam turbine turned into degradation states. In the 5th month, the oil pressure gauge of TSI raised an alarm as shown in Fig. 4 and the steam

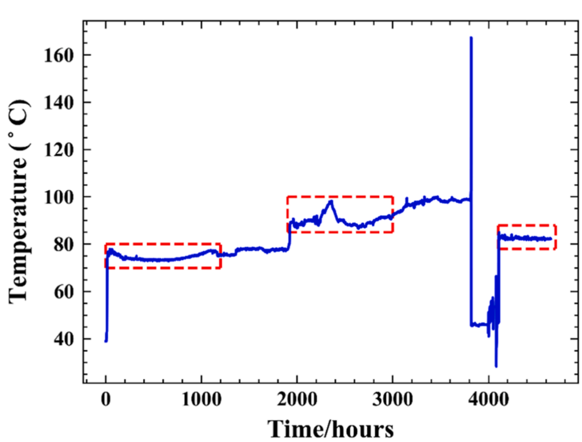

Fig. 3. Metal temperature of the #1 bearing bushes of a steam turbine generator during 7 months.

Fig. 4. The alarm raised by the oil pressure gauge of TSI.

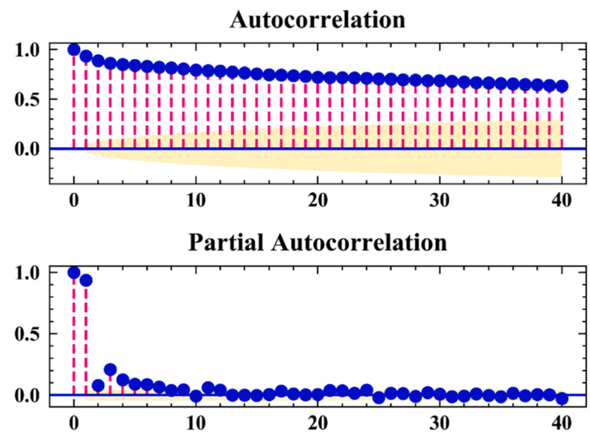

Fig. 5. Autocorrelation and partial autocorrelation in partial temperature data.

turbine was going to fail. Meanwhile, the active power dropped sharply and the temperature signal rose rapidly. Then, the steam turbine was shut down and repaired, when the active power dipped to zero and the temperature signal reduced to the ambient temperature. Finally, the steam turbine was back to the normal state after repairs and the monitoring signals were also in the normal levels.

Due to various influencing factors, the practical monitoring process inevitably exhibits multiple operating modes and stochastic properties. Fig. 5 shows the correlation analysis results of the temperature signals. In order to make the figures more clearly, the observations from the 1000th hours to the 3000 h are selected, when different operating modes are included. It can be seen that not only the temporal correlation among observations, but also the cross-correlation between different operating modes significantly exists. Therefore, it is necessary to consider these relationships in the modeling of CM.

GP is then used to construct the monitoring models for characterizing uncertainty and complicated correlations. To accurately describe the data of different modes, multiple GPs are built for each mode. Since there are few data monitored at the beginning stage of a new operating mode, an adaptive self-transfer learning framework for unsupervised anomaly detection based on multiple GPs is proposed to transferring the common attributes across modes, without relying on abnormal data or labeled data. Considering that there are no labeled data for each operating mode, a two-hierarchical identification criterion is developed to first identify the change-points and secondly detect the anomaly automatically. Based on the proposed methods, the abnormal conditions of steam turbines are expected to be detected timely so that the maintenance actions can be taken more quickly and efficiently.

3. Adaptive self-transfer learning

3.1. Gaussian processes

A GP describes a continuous sequence of random variables, any finite number of which follow a multivariate Gaussian distribution. Generally, a GP $f(\boldsymbol{x})$ can be characterized by a mean function $m(\boldsymbol{x})$ and a positive semi-definite covariance matrix $\boldsymbol{K}(\boldsymbol{x},\boldsymbol{x})$, i.e.,

$$f(\boldsymbol{x}) \sim \mathscr{GP}(m(\boldsymbol{x}), \boldsymbol{K}(\boldsymbol{x},\boldsymbol{x})), \tag{1}$$

where $\boldsymbol{x} = [x_1, x_2, \cdots, x_n]^{\mathrm{T}}$ is the input observation, $\boldsymbol{y} = f(\boldsymbol{x})$ is the target output observation of GP, $m(\boldsymbol{x}) = \mathbb{E}[\boldsymbol{y}]$ and the (*i, j*)-th element of $\boldsymbol{K}(\boldsymbol{x},\boldsymbol{x})$ is $k(x_i, x_j)$. $k(\cdot)$ is the covariance function usually defined in terms of a kernel function, and can be regarded as a similarity measure between the inputs.

Let $\boldsymbol{\theta}$ denote the vector of hyper-parameters of GP. Then, the log-likelihood function can be expressed as

$$\ell(\boldsymbol{\theta}) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\boldsymbol{K}| - \frac{1}{2}(\boldsymbol{y} - m(\boldsymbol{x}))^{\mathrm{T}}\boldsymbol{K}^{-1}(\boldsymbol{y} - m(\boldsymbol{x})). \tag{2}$$

The parameter $\boldsymbol{\theta}$ can be estimated by maximizing (Eq. (2)).

The GP regression model is a Bayesian model and widely applied in regression analysis, with prior $p(\boldsymbol{y}|\boldsymbol{x}) = \mathscr{N}(m(\boldsymbol{x}), \boldsymbol{K}(\boldsymbol{x},\boldsymbol{x}))$. According to the definition of a GP, when a new input $\boldsymbol{x}^*$ is available, $\boldsymbol{y} = f(\boldsymbol{x})$ and $y^* = f(\boldsymbol{x}^*)$are jointly Gaussian. Therefore, the conditional distribution $p(y^*|\boldsymbol{y})$ is calculated in closed form with the Gaussian identities, i.e.,

$$p(y^*|\boldsymbol{y}) = \mathscr{N}(\mu, \Sigma), \tag{3}$$

where $\mu = m(\boldsymbol{x}^*) + \boldsymbol{K}(x^*,\boldsymbol{x})\boldsymbol{K}(\boldsymbol{x},\boldsymbol{x})^{-1}(\boldsymbol{y} - m(\boldsymbol{x}))$,

$$\Sigma = k(x^*, x^*) - \boldsymbol{K}(x^*,\boldsymbol{x})\boldsymbol{K}(\boldsymbol{x},\boldsymbol{x})^{-1}\boldsymbol{K}(\boldsymbol{x}, x^*),$$

$$\boldsymbol{K}(x^*,\boldsymbol{x}) = \boldsymbol{K}(\boldsymbol{x}, x^*)^{\mathrm{T}} = [k(x^*, x_1), \cdots, k(x^*, x_n)].$$

As the focus is on the predictive result for the unknown output y^*, the posterior distribution $p(y^*|\boldsymbol{y})$ conditional on $\boldsymbol{y}$ can be calculated using Eq. (3). Note that Eq. (3) provides a flexible prediction for y^*, which can not only determine the point estimate with the distribution's mean, but also quantify the prediction uncertainty with its variance.

3.2. Dynamic time warping

Dynamic time warping (DTW) is a distance measure of similarity obtained by searching the optimal alignment between two time-dependent sequences. Using DTW, we can map one data point in a sequence to more than one data point in the other sequence, and thus, the distance between two time-sequences with different lengths can be calculated. If there is an anomaly in the operation process, the dissimilarity analysis by DTW between the online monitoring data sequence (varying length) and the historical sequences (fixed length) can be used for the detection of operating mode change.

Given a historical time series $\boldsymbol{v} = [v_1, \cdots, v_s]^{\mathrm{T}}$ and an online monitoring data $\boldsymbol{w} = [w_1, \cdots, w_n]^{\mathrm{T}}$, the similarity between $\boldsymbol{v}$ and $\boldsymbol{w}$ is measured by the DTW distance. First, a distance matrix is constructed, in which the (*i, j*)-th element denotes an alignment between two points v_i and w_j. Generally, the Euclidean distance is used as the alignment. Then, a warping path is defined as a cumulative distance of a sequence of elements from (1, 1) to (*l, n*), which should meet three constraints including boundary, monotonicity, and continuity. To optimize the best warping path and the minimum distance, a cost matrix is built based on dynamic programming, i.e.,

$$D(i,j) = d_{i,j} + \min\begin{cases} D(i,j-1) \\ D(i-1,j) \\ D(i-1,j-1) \end{cases}, \tag{4}$$

where $i = 1, \cdots, s, j = 1, \cdots, n$. Finally, the minimum distance $D(l, n)$ is the DTW distance, i.e., $DTW(\boldsymbol{v}, \boldsymbol{w}) = D(s, n)$.

To account for the varying warping paths, the DTW distance is normalized by l^*_{DTW}, the number of aligned pairs on the optimal warping path, to represent the dissimilarity of two data sequences as follows

$$dsim(\boldsymbol{v}, \boldsymbol{w}) = DTW(\boldsymbol{v}, \boldsymbol{w}) \,/\, l^*_{DTW}. \tag{5}$$

A lower $dsim(\cdot)$ indicates a higher similarity of v and w.

3.3. The proposed adaptive self-transfer learning method

At the initial stage of CM, a GP model (GP-1) is constructed to capture the current normal operating mode (called Mode-1), and the model parameters are updated as more monitoring data are collected. When some external factors or artificial adjustments occur, the steam turbine switches to a new normal operating mode (called Mode-2). At this time, a new GP model (GP-2) should be established for Mode-2. However, at the beginning of Mode-2, there are only a few monitoring data for parameter estimation, and it is difficult for the determined GP model to accurately capture the operating characteristics of Mode-2. Then, the monitoring efficiency and detection accuracy would decline significantly. To solve this problem, transfer learning is introduced.

Although the transfer learning has been combined with some neural networks and used in the anomaly detection of industrial equipment [47], the limited monitoring data with fault information were usual necessary and the uncertainties of operating processes were not modelled. Considering that there is often no abnormal data and the uncertainty quantification can improve the CM effectiveness, we propose a transfer learning based on GPs to solve these problems. Some previous works have used GP or sparse GP for transfer learning [48–51], all of which focused on transferring knowledge from source domains to target tasks for offline applications. Therefore, we develop an adaptive self-transfer learning algorithm based on Gaussian process models (ASTL-GP) for multimode process condition monitoring, which can adapt the online monitoring data in mode transfer automatically.

Using transfer learning, we can transfer knowledge from Mode-1 to Mode-2, especially at the beginning of Mode-2. Then, a more accurate GP-2 can be obtained with limited data by leveraging cross-correlations between the observations of different modes. There are two requirements that the expected transfer learning algorithm should meet. One is the shared knowledge between modes should be transferred as much as possible when these two modes are similar. The other is that negative transfer should be avoided as much as possible, if the two modes are unrelated. Obviously, a mechanism that can automatically adjust the transfer schemes would be preferred.

Denote $v_{1,1}, \cdots, v_{1,s}$ and $v_{2,1}, \cdots, v_{2,n}$ as the observations of Mode-1 collected at time $t_{1,1}, \cdots, t_{1,s}$ and the observations of Mode-2 collected at time $t_{2,1}, \cdots, t_{2,n}$, respectively. Suppose the input of GP-1 at $t_{1,s}$ is a vector of lagged observations $x_{1,s} = [v_{1,s-1}, \cdots, v_{1,s-L}]^{\mathrm{T}}$, where L is the number of lags to incorporate. The target output corresponding to $x_{1,s}$ is $y_{1,s} = v_{1,s}$. The input and the output for GP-2 at $t_{2,n}$ are $x_{2,n} = [v_{2,n-1}, \cdots, v_{2,n-L}]^{\mathrm{T}}$ and $y_{2,n} = v_{2,n}$, respectively. Note that the temporal dependency is considered and the past L observations are used to predict the next observation.

Let $\boldsymbol{y} = (\boldsymbol{y}_1^{\mathrm{T}}, \boldsymbol{y}_2^{\mathrm{T}})^{\mathrm{T}}$and define a GP over $\boldsymbol{y}$, i.e.,

$$\boldsymbol{y} \sim \mathscr{GP}(m(\boldsymbol{x}), \widetilde{\boldsymbol{K}}), \tag{6}$$

where $\boldsymbol{x} = [\boldsymbol{x}_1^{\mathrm{T}}, \boldsymbol{x}_2^{\mathrm{T}}]^{\mathrm{T}}, \boldsymbol{x}_1 = [x_{1,1}, \cdots, x_{1,s}]^{\mathrm{T}}, \boldsymbol{y}_1 = [y_{1,1}, \cdots, y_{1,s}]^{\mathrm{T}}, \boldsymbol{x}_2 = [x_{2,1}, \cdots, x_{2,n}]^{\mathrm{T}}, \boldsymbol{y}_2 = [y_{2,1}, \cdots, y_{2,n}]^{\mathrm{T}}$, $\widetilde{\boldsymbol{K}}$ is the kernel covariance matrix for transfer learning, and the (*i, j*)-th element of $\widetilde{\boldsymbol{K}}$ is defined as

$$\widetilde{k}(x_{\rho,i}, x_{\tau,j}) = \lambda_{ij} k(x_{\rho,i}, x_{\tau,j}), \tag{7}$$

where $\rho, \tau \in \{1,2\}$and the additional factor λ_{ij} is assumed to be

$$\lambda_{ij} = \frac{1}{1+\varsigma(x_{\rho,i}, x_{\tau,j}) \times dsim(y_1, y_2)},$$

with $\varsigma(x_{\rho,i}, x_{\tau,j}) = 0$ if $\rho = \tau$, otherwise $\varsigma(x_{\rho,i}, x_{\tau,j}) = 1$.

Note that Eq. (7) indicates that λ is expected to be 1 for highly correlated Mode-1 and Mode-2, and the cross-correlation between observations from the different modes equals to the correlation between the ones in the same mode. For nearly uncorrelated Mode-1 and Mode-2, λ is expected to be 0, and $\widetilde{k}(x_{\rho,i}, x_{\tau,j}) \to 0$. These behaviors are consistent with the two requirements proposed above for transfer learning and, thus, make the assumptions in Eq. (7) more flexible and convincing.

Using CM, we often want to predict the next target output $y_{2,n+1}$ at time $t_{2,n}$ by calculating $p(y_{2,n+1}|\boldsymbol{y})$. The corresponding inference is the same as that in Eq. (3). Let $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ be the hyperparameter vectors of GP-1 and GP-2, respectively. The mean and covariance of $(y_{2,n+1}|\boldsymbol{y})$ is given by

$$\mu_{2,n+1} = m(x_{2,n+1}) + \widetilde{\boldsymbol{K}}(x_{2,n+1}, \boldsymbol{x})\widetilde{\boldsymbol{K}}(\boldsymbol{x}, \boldsymbol{x})^{-1}(\boldsymbol{y} - m(\boldsymbol{x})), \tag{8}$$

$$\Sigma_{2,n+1} = k(x_{2,n+1}, x_{2,n+1}) - \widetilde{\boldsymbol{K}}(x_{2,n+1}, \boldsymbol{x})\widetilde{\boldsymbol{K}}(\boldsymbol{x}, \boldsymbol{x})^{-1}\widetilde{\boldsymbol{K}}(\boldsymbol{x}, x_{2,n+1}), \tag{9}$$

where $\widetilde{\boldsymbol{K}}(\boldsymbol{x}, \boldsymbol{x}) = \begin{bmatrix} \widetilde{\boldsymbol{K}}_{11} & \widetilde{\boldsymbol{K}}_{12} \\ \widetilde{\boldsymbol{K}}_{21} & \widetilde{\boldsymbol{K}}_{22} \end{bmatrix}$, $\widetilde{\boldsymbol{K}}_{11} = \boldsymbol{K}(\boldsymbol{x}_1, \boldsymbol{x}_1)$, $\widetilde{\boldsymbol{K}}_{22} = \boldsymbol{K}(\boldsymbol{x}_2, \boldsymbol{x}_2)$, $\widetilde{\boldsymbol{K}}(x_{2,n+1}, \boldsymbol{x}) = \widetilde{\boldsymbol{K}}(\boldsymbol{x}, x_{2,n+1})^{\mathrm{T}}$,

$\widetilde{\boldsymbol{K}}(x_{2,n+1}, \boldsymbol{x}) = [\widetilde{k}(x_{2,n+1}, x_{1,1}), \cdots, \widetilde{k}(x_{2,n+1}, x_{1,s}), k(x_{2,n+1}, x_{2,1}), \cdots, k(x_{2,n+1}, x_{2,n})]$, $\widetilde{\boldsymbol{K}}_{12} = \widetilde{\boldsymbol{K}}_{21}$ and

$$\widetilde{\boldsymbol{K}}_{12}(\boldsymbol{x}_1, \boldsymbol{x}_2) = \begin{bmatrix} \widetilde{k}(x_{1,1}, x_{2,1}) & \cdots & \widetilde{k}(x_{1,1}, x_{2,n}) \\ \vdots & \ddots & \vdots \\ \widetilde{k}(x_{1,l}, x_{2,1}) & \cdots & \widetilde{k}(x_{1,l}, x_{2,n}) \end{bmatrix}_{l \times n}.$$

Furthermore, $\mu_{2,n+1}$ can be decomposed as follows:

$$\mu_{2,n+1} = m(x_{2,n+1}) + \sum_{i=1}^{s} \lambda \alpha_i k(x_{2,n+1}, x_{1,i}) + \sum_{j=1}^{n} \alpha_{j+l} k(x_{2,n+1}, x_{2,j}), \tag{10}$$

where α_i is the ith element of $\widetilde{\boldsymbol{K}}(\boldsymbol{x}, \boldsymbol{x})^{-1}(\boldsymbol{y} - m(\boldsymbol{x}))$. The second term in (10) represents the cross-correlations between the observations from different modes where a shrinkage is introduced based on the similarity of modes. When more monitoring data of Mode-2 are available, the similarity between two modes and the additional factor λ keep updating. Thereby, the transfer is also automatically adjusted.

Since GP-1 and its parameter $\boldsymbol{\theta}_1$ are prior information in transfer learning which can be determined by using the method in Section 3.1, only the parameter $\boldsymbol{\theta}_2$ of GP-2 need to be estimated here. Considering that the data size of $\boldsymbol{y}_2$ from Mode-2 may be small, we proposed to estimate $\boldsymbol{\theta}_2$ based on the conditional distribution $p(\boldsymbol{y}_2|\boldsymbol{y}_1)$. According to (6), we know that this distribution is also a Gaussian, i.e.,

$$p(\boldsymbol{y}_2|\boldsymbol{y}_1) = \mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2), \tag{11}$$

where the mean $\boldsymbol{\mu}_2 = m(\boldsymbol{x}_{2,n}) + \widetilde{\boldsymbol{K}}(\boldsymbol{x}_2, \boldsymbol{x}_1)\boldsymbol{K}(\boldsymbol{x}_1, \boldsymbol{x}_1)^{-1}(\boldsymbol{y}_1 - m(\boldsymbol{x}_1))$ and the covariance matrix

$\boldsymbol{\Sigma}_2 = \boldsymbol{K}(\boldsymbol{x}_2, \boldsymbol{x}_2) - \widetilde{\boldsymbol{K}}(\boldsymbol{x}_2, \boldsymbol{x}_1)\boldsymbol{K}(\boldsymbol{x}_1, \boldsymbol{x}_1)^{-1}\widetilde{\boldsymbol{K}}(\boldsymbol{x}_1, \boldsymbol{x}_2)$.

The log-likelihood function is expressed by

$$\ell(\theta_2) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_2| - \frac{1}{2}(y_2 - \mu_2)^T \Sigma_2^{-1}(y_2 - \mu_2). \tag{12}$$

The parameter $\boldsymbol{\theta}_2$ can be estimated by maximizing (Eq. (12)).

The proposed ASTL-GP shown in Fig. 6 combines the data from multiple normal modes, and captures both autocorrelations between different observations as well as the cross-correlations between different modes. This allows GP-2 to be more flexible and credible at the early stage of Mode-2, if strong cross-correlations between these two modes exist.

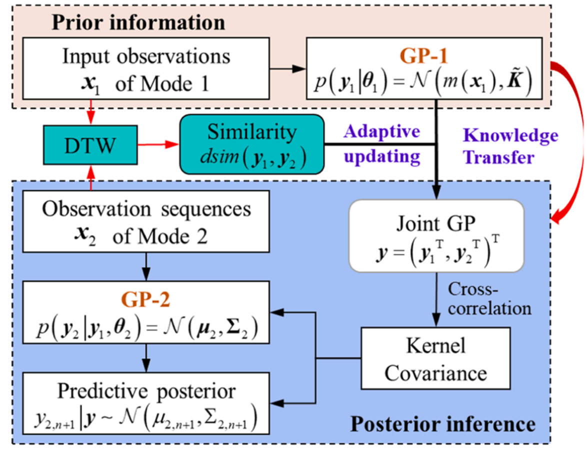

Fig. 6. The ASTL-GP framework for multimode process monitoring.

4. The online multimode process monitoring and anomaly detection framework

4.1. Data preprocessing

Due to the external interferences and the instability of sensors, the monitoring data from TSI are usually nonstationary and noisy. Some abrupt spikes or missing data occur. Therefore, to avoid false alarms and ensure continuous CM, smoothness and data interpolation can be used for the preprocessing of the raw data. Generally, exponential smoothing, singular spectrum analysis, spline-based non-parametric regression technique, or deep learning methods can be selected for data smoothing. Mean imputation, regression imputation and machine learning-based imputation are commonly used data completion methods.

Considering the requirements of real-time monitoring, only the simpler methods, such as exponential smoothing spline-based regression and mean imputation, are used for data preprocessing in this paper, which can be conducted efficiently and conveniently online. After pretreatment of raw observation data, the obtained smooth data sequences are input into the online process monitoring and anomaly detection.

4.2. Two-hierarchical identification criterion

The key part of monitoring a multimode process without labeled historical data is how to distinguish different operating modes in an unsupervised framework. That is, we should develop criteria which can simultaneously identify different normal operating conditions (expected) and detect abnormal conditions (unexpected). When a normal mode turns to another normal mode or to the abnormal mode, there exist change points in the monitoring data in both cases. The occurrence of change points can represent both normal mode transition and anomaly. Therefore, we develop a two-hierarchical identification criterion, where the change point is first identified, and then a decision whether it is an anomaly or a normal mode transition is made automatically.

Suppose the system is currently in the r-th normal operation mode and the input observations up to time $t_{r,n}$ is denoted by $\boldsymbol{x}_r = [x_{r,1}, \cdots, x_{r,n}]^{\mathrm{T}}$. Then, according to (3), (8) and (9), the predictive posterior of the target output $y_{r,n+1}$ at time $t_{r,n+1}$ is Gaussian-distributed, i.e.,

$$p(y_{r,n+1}|\boldsymbol{y}_{r-1}, \boldsymbol{y}_r) = \mathcal{N}(\mu_{r,n+1}, \Sigma_{r,n+1}). \tag{13}$$

When $r = 1$, the predictive posterior is $p(y_{r,n+1}|\boldsymbol{y}_r)$. Given a confidence level $1-\alpha$, the posterior interval estimates of $y_{r,n+1}$ are $[\mu_{r,n+1} - z_{\alpha/2}\sigma_{r,n+1}, \mu_{r,n+1} + z_{\alpha/2}\sigma_{r,n+1}]$, where $z_{\alpha/2}$ is the quantile of standard normal distribution and $\sigma^2_{r,n+1} = \Sigma_{r,n+1}$.

Then, a condition index (CI) is constructed based on the posterior intervals:

$$h_{r,n+1} = \max\left[\frac{y_{r,n+1} - PI^U_{r,n+1}}{PI^U_{r,n+1} - PI^L_{r,n+1}}, \frac{PI^L_{r,n+1} - y_{r,n+1}}{PI^U_{r,n+1} - PI^L_{r,n+1}}\right], \tag{14}$$

where $PI^U_{r,n+1} = \mu_{r,n+1} + z_{\alpha/2}\sigma_{r,n+1}$ and $PI^L_{r,n+1} = \mu_{r,n+1} - z_{\alpha/2}\sigma_{r,n+1}$. A positive $h_{r,n+1}$ means the observation $y_{r,n+1}$ is out of the intervals at time $t_{r,n+1}$.

It is reasonable to identify $y_{r,n+1}$ as a change point, if $h_{r,n+1}$ is positive. Furthermore, we take two or more consecutive observations to determine if there are change points. Multiple consecutive outliers can highlight the occurrence of a new mode which differs from the current mode. The specific criterion for identifying a change point in the multimode process is summarized below.

Criterion-1: Given a significance level α, if γ_0 consecutive observations from time $t_{r,n+1}$ to $t_{r,n+\gamma_0}$ satisfy $h_{r,n+1} > 0, h_{r,n+2} > 0, \cdots, h_{r,n+\gamma_0} > 0$, then $y_{r,n+1}$ is flagged as a change point.

Generally, CM is conducted every few minutes. Under this resolution, false alarms can result in unexpected operation or maintenance decisions. To eliminate false alarms, only when γ consecutive observations fall on the same side out of the confidence intervals, a new operating mode (normal or abnormal conditions) is defined to occur.

After a change point has been identified, we need to make the second-hierarchy criterion for distinguishing between the normal mode transition and anomaly. Generally, when the operating mode of the equipment is switched from a normal one to another normal one, the monitoring data of the new mode will be still in control after relatively short-term fluctuations. As the abnormality is usually an early sign of a potential fault, if the system enters abnormal conditions, the corresponding monitoring data are expected to fluctuate dramatically and be out of control continuously. Therefore, Criterion-2 for detecting the anomaly is given below.

Criterion-2: For consecutive change points $y_{r,n+1}, \cdots, y_{r,n+\gamma_2}$ and the corresponding sequence of CIs, if there is an excessive number of $h_{r,n+\gamma_1}, \ldots, h_{r,n+\gamma_2}$ out of the given control interval, then an anomaly is identified. Otherwise, a normal mode transition occurs.

The control limits of the given interval can be calculated based on the idea of control charts. During $[t_{r,n+1}, t_{r,n+\gamma_1}]$, the system is assumed to be in a short-term fluctuation. Hence, the judgement of Criterion-2 is made from $t_{r,n+\gamma_1}$.

4.3. The Online unsupervised monitoring framework

Since the value ranges of the raw monitoring data are unfixed, it would be better if the data points were standardized as the inputs for the monitoring models. This can ensure the mean of the inputs is zero. Considering that there is no prior information (mean and variance) about the raw monitoring data, an online adaptive standardization method is necessary. In addition, as the length of the monitoring data increases over time, the computational burden will increase. To improve the operating efficiency of the proposed monitoring framework, an early-stopping rule for the training of ASTL-GP is designed. Then, the pseudo-code of a joint real-time data standardization and early-stopping procedure is presented in Algorithm 1:

The updating of the mean and standard deviation is early stopped at time $t_{L+\xi-1}$. The number ξ of iterations is smaller than the duration of one normal operating mode. That is, the standardization of the observations for one mode depends on the first $\xi + L$ observations. In practice, if several consecutive μ^v_{i+1} and σ^v_i are rather stable, the iteration can be early stopped and thus the value of ξ is determined. Combined with Algorithm 1, the pseudo-code of the online unsupervised monitoring and anomaly detection framework for the multimode process of steam turbine is summarized in Algorithm 2 as follows:

Algorithm 1

Real Time Data Standardization and Early Stopping.

Input: Raw monitoring data $v_1, v_2, \cdots$ for one operating mode, number L of lags, number ξ of iterations.
Output: The standardized data and the estimates of the parameters $\boldsymbol{\theta}^*$.
1: **Initialization**: Set $i = L$ and the observations $v_1, \cdots, v_i$ are z-score standardized.
2: **while** $(i - L + 1) \le \xi$ **do**
3: Calculate the mean μ^v_i and standard deviation σ^v_i of $v_1, \cdots, v_i$.
4: Standardize the observation v_{i+1} to $\widetilde{v}_{i+1} = (v_{i+1} - \mu^v_i)/\sigma^v_i$.
5: Update the mean μ^v_{i+1} and standard deviation σ^v_{i+1} of $v_1, \cdots, v_i, v_{i+1}$.
6: Construct the input $x_{i+1} = [v_i, \cdots, v_{i-L+1}]^{\mathrm{T}}$ and the output $y_{i+1} = v_{i+1}$. Estimate the unknown parameters of GP or ASTL-GP.
7: Set $i = i + 1$.
8: **end while**
9: The final mean and standard deviation of the observations are determined as $\mu^v_* = \mu^v_{L+\xi-1}$ and $\sigma^v_* = \sigma^v_{L+\xi-1}$. The unknown parameters $\boldsymbol{\theta}^*$ of GP or ASTL-GP are estimated based on $x_{L+1}, \cdots, x_{L+\xi}$ and $y_{L+1}, \cdots, y_{L+\xi}$.
10: **For** all the observations after time $t_{L+\xi-1}$, **do**
Implement the data standardization based on $\mu^v_*$and σ^v_*as in Step 3;
11: **end for**
12: Predict the posteriors based on the trained model with $\boldsymbol{\theta}^*$.

To make the proposed framework more intuitive, its schematic is shown in Fig. 7.

5. Case study

This section presents the case study of a steam turbine in a 610-MW thermal power plant, which belongs to Shanghai Electric Group and is located in Iraq. Moreover, the proposed method is also suitable to other steam turbines of other companies or country.

5.1. Condition monitoring and anomaly detection

The TSI data has various kinds of monitoring variables. Considering that the bearings are key components with higher frequent faults in steam turbines, only the metal temperatures of Bearing #1 are selected here as described in Section 2. The dataset covers the period 15/01/2021 to 15/04/2021. The observations were sampled every 10 min. As the steam turbine's health states would change over time, the ambient temperature rises and the electricity consumption fluctuates, the operating modes of the steam turbine can be different.

Fig. 8 shows the raw data of the metal temperature. The actual alarm temperature in the plant is set to 107 °C. From Fig. 8, we can see that the bearing is subject to multiple operating modes before failure alarm.

Algorithm 2

An Unsupervised Framework for Adaptive Multimode Process Monitoring and Anomaly Detection

Input: The data sequence from the CM of a steam turbine, $x_{r,1}, \cdots, x_{r,n}$.
Output: The anomaly y^*.
1: **Initialization**: Set r =1 and n =1. Construct a GP for mode-r.
2: **Preprocessing**: Smooth and complete the raw monitoring data up to $t_{r,n+1}$. Standardize the preprocessed monitoring data as in Algorithm 1.
3: **Modeling**: Estimate the unknown parameters of GP-r, and calculate the predictive posterior at time $t_{r,n+1}$ as in Eq. (13).
4: Implement the early-stopping rule as in Algorithm 1.
5: **while** the anomaly $y^*$is not identified, **do**
6: **if** $y_{r,n+1}$ is flagged as a change point according to Criterion-1 **then**
7: Implement Criterion-2 with last $\gamma_2$observations.
8: **if** the sequence of CIs is in control, **then**
9: A new normal operation mode occurs and set $r = r + 1$, $n = 1$.
Calculate the similarity between the monitoring data from mode r and mode r-1. Construct a new GP-r. Estimate the parameters and compute the predictive posteriors by using the adaptive transfer learning as in Subsection 3.3.
10: **else**
11: The change point is identified as an anomaly.
12: **end if**
13: **end if**
14: Set $n = n + 1$ and continue the CM.
15: **end while**

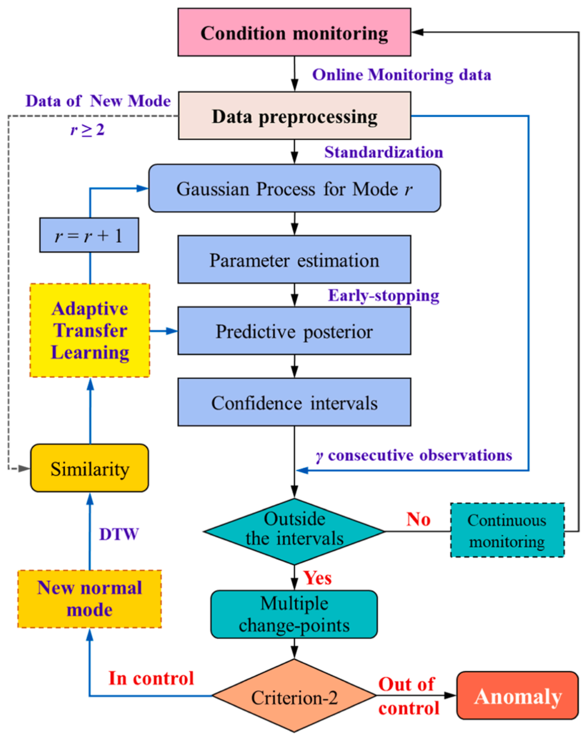

Fig. 7. Schematic of the online unsupervised monitoring framework for a steam turbine with multiple operating modes. (Initial value of r is 1.)

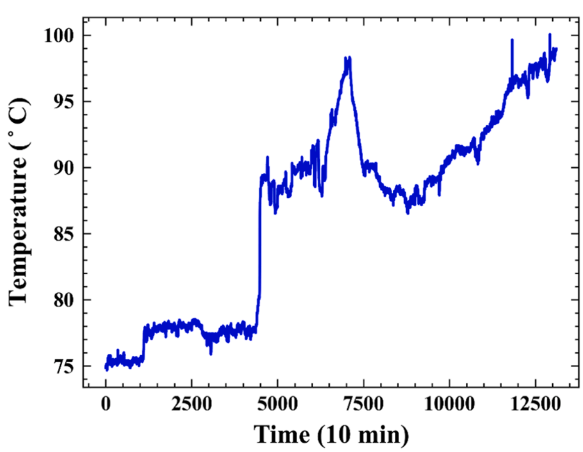

Fig. 8. Raw monitoring data of the metal temperature of the #1 bearing bushes, which were sampled every 10 minutes.

When approaching the actual alarm threshold, the monitoring temperature fluctuates heavily. Therefore, an appropriate online multimode process monitoring method is necessary to detect the faults at early stage. Then, the maintenance decisions can be made in time.

The number of lags is set as $L = 12$, which means the data from the past two hours is used to predict the observation at the next time. As the power plant had annual overhaul at the end of 2020, we can assume the system is in a healthy state during the first three days. Then, the parameter ξ in the early stopping rule is set to 432. According to this assumption, most observations in the first three days of the CM should fall within the confidence intervals. Thus, $z_{\alpha/2}$ can be obtained by

$$z_{\alpha/2} = \max\left\{ \left(\mu_{1,n+1} - y_{1,n+1}\right) / \sigma_{1,n+1} \right\}. \tag{15}$$

In Eq. (15), the quantile is determined as the largest $z_{\alpha/2}$ among the first three days' observations to avoid false alarms.

The kernel function $k(x_i, x_j)$ of GP should be carefully determined as it represents the underlying correlation between the observations. The candidate kernel functions are shown in Table 1. The monitoring data from Mode-1 are used to train the GPs with different kernels and the corresponding log-likelihood values are also listed in Table 1. It can be found that the squared exponential kernel which has the largest log-likelihood value should be selected as the kernel function of GP. Combined with the Euclidean distance $d_{ED}(\cdot)$ between x_i and x_j, the formula of the squared exponential kernel is given by

$$k\left(x_i, x_j\right) = \sigma_f^2 \exp\left(- d_{ED}\left(x_i - x_j\right) / \left(2\sigma_e^2\right)\right), \tag{16}$$

As the observations are standardized, σ_f is set to one.

Since there is no data of other modes for the first normal operating mode (Mode-1) at the beginning of CM, the GP is used to fit the observations collected on-line for Mode-1. When the system switches to other operating modes from Mode-1, the ASTL-GP model is applied. In practice, the engineers or the on-site workers usually determine the anomalies when a specific monitoring signal exceeds the prefixed threshold for two or three consecutive hours. Therefore, we assume that if 18 consecutive observations (during three hours) are beyond the posterior intervals, a change point is flagged. Then, according to Criterion-2, the X-charts are used to distinguish the anomaly or the transition point of normal modes from the consecutive change-points.

Fig. 9 shows the results of multimode process monitoring and anomaly detection by using Algorithms 1 and 2, where $z_{\alpha/2}$ is 3.8 by calculating (15). The whole operating process of the steam turbine can be divided into four normal modes. The red hollow circles in the left panels of Fig. 9 denote change points, and the observations after the red hollow circle are also identified as change points. The right panels of Fig. 9 are the X-charts whose control limits are determined by 95% confidence intervals. From Fig. 9(a), (b) and (c), we can find that the part of the CI sequence which exceeds zero becomes in control again after a very short time fluctuation. This indicates that these change points are the normal mode transition points. On the contrary, after a change point is identified in the CM of Mode-4, the sequence of identified change-points presents a significant deterioration trend and is finally out of control. Therefore, an anomaly occurs. Although there are a few outliers in the early monitoring time of some modes, this phenomenon is not a continuous process and can be regarded as accidental events due to random factors and lack of monitoring data rather than change points. Note that the computational efficiency is a very important factor for an online CM task. When a new incoming observation is available, the proposed monitoring procedure is only iterated once for outputting a prediction and then making the identification result. Therefore, the CPU time of the proposed method for one iteration of each mode are reported in Table 2. The operating system of the test computer is Windows 10, the CPU is AMD Ryzen 7 5800H with Radeon Graphics 3.20 GHz, and the RAM memory is 16.0 GB. The calculating software is Python 3.7. It can be seen from Table 2 that when a new observation is available, the CPU time for running one iteration to outputting one prediction in any operation mode is below 7 s. That is, the anomaly detection result at each inspection can be obtained in a few seconds. Compared with the 10 min observed interval, the proposed method is obviously very efficient with less computational burden in the online application before the next observation reaching and thereby the

Table 1
Candidate kernel functions and their estimation results.

Kernel	Formula	Log-likelihood
Laplacian	$k(x_i, x_j) = \sigma_f^2 \exp(-\tau/\sigma_e)$	-284.664
Exponential	$k(x_i, x_j) = \sigma_f^2 \exp(-\tau/(2\sigma_e^2))$	-147.945
Squared exponential	$k(x_i, x_j) = \sigma_f^2 \exp(-\tau^2/(2\sigma_e^2))$	-98.9056
Matern 3	$k(x_i, x_j) = \sigma_f^2 \left(1 + \frac{\sqrt{3}\tau}{\sigma_e}\right) \exp\left(-\frac{\sqrt{3}\tau}{\sigma_e}\right)$	-109.618

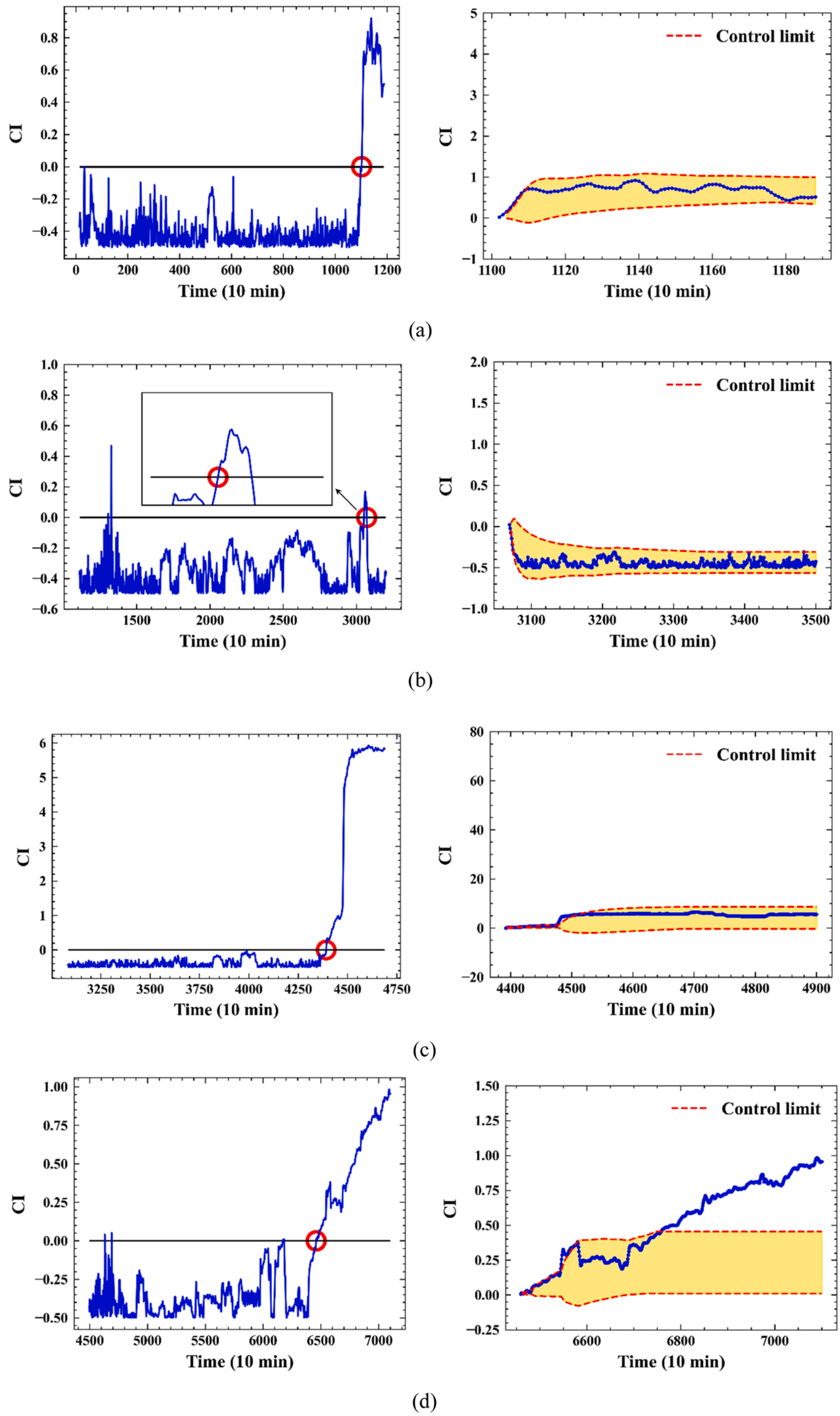

Fig. 9. Results of the adaptive multimode process monitoring and anomaly detection. (a) Mode-1; (b) Mode-2; (c) Mode-3; (d) Mode-4 and anomaly. (The left panel of each subfigure is the monitoring result of the CI sequence, and the right panel is the identification result of the proposed two-hierarchical criterion).

Table 2
CPU time of the proposed method for one iteration during different operational modes.

Mode-1	Mode-2	Mode-3	Mode-4
1.592 sec	5.773 sec	5.680 sec	6.980 sec

reliability management decisions can be made timely.

The identification results of the multimode process monitoring compared with the alarm raised by TSI are shown in Fig. 10. The alarm raised by the oil pressure gauge of TSI was on April 8, 2021, and the steam turbine was shut down for repair on May 4, 2021. From Fig. 10, the proposed monitoring framework can accurately and adaptively identify different operating modes, and the anomaly is thereby detected at the 6477th samples (February 28, 2021). That is, the proposed approach can detect the steam turbine's anomaly before its failure and shut down. Fig. 11 shows that the bearing bushes were severely wearing out and generated visual cracks. Considering that wearing out and crack are types of soft failure, the actual fault may occur earlier and gradually propagate to an alarm finally. Therefore, our approach can release a timely alarm of the potential fault. Then, the operators of the power plant have time to adjust the generation plans and arrange maintenance activities.

To analyze the robustness of the proposed anomaly detection method, we assume different values of consecutive observations for Criterion-2, i.e., 6, 12, 24 and 30 consecutive observations (one, two, four and five hours). The identification results of the multimode process monitoring under these settings are shown in Fig. 12. We can find that the operating process is adaptively divided into four normal operation modes when the values of consecutive observations are set as 6, 12 and 18. While the operating process is separated into three normal modes under Criterion-2 of 24 and 30 consecutive observations. This indicates the identification sensitivity of the proposed method becomes lower. However, the detected anomaly time under these settings is very close and all of them are before the steam turbine's failure time. It can conclude that the proposed unsupervised anomaly detection method is robust and credible under identification criteria of different settings.

5.2. Comparative study

In this section, the proposed ASTL-GP model is compared with GP. Fig. 13 shows the comparison between ASTL-GP and GP. Only the monitoring data of the early stage of Mode-2 is selected for this illustration. From Fig. 13, the predictive posterior intervals of ASTL-GP are narrower than those of GP, which indicates the performance of ASTL-GP is more robust especially at the beginning of CM. Furthermore, the monitoring results of the whole process by these two models are given in

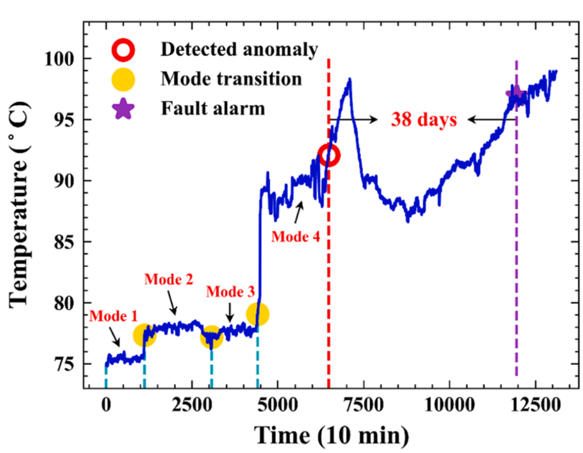

Fig. 10. Anomaly detection results compared with the fault alarm by TSI.

Table 2. The FAR (false accept rate) [10] is used to measure the performance of CIs, which is the ratio of the number of outliers to the total number of CIs for Mode-r, i.e.,

$$\mathrm{FAR}_r = \frac{1}{n_r}\sum_{i=0}^{n_r}\varepsilon(h_{r,i}), \tag{17}$$

where $\varepsilon(h_{r,i}) = 1$, if $h_{r,i} > 0$; otherwise, $\varepsilon(h_{r,i}) = 0$. n_r is the total number of CIs for Mode-r.

From Table 3, we can see that ASTL-GP can accurately identify all four normal operation modes, whereas GP cannot identify the third mode. This indicates that ASTL-GP is more sensitive to the mode transition than GP. Even when the difference between the operating mechanisms of Mode-2 and Mode-3 is not obvious, ASTL-GP still can make precise identification. This may be because the transfer learning in ASTL-GP can obtain more operation information from other modes. Also, the transition time of other modes and anomaly time detected by GP are closer to those by ASTL-GP. The reason is the differences among different modes (except between Mode-2 and Mode-3) are significant, and thus the mode transitions are identified easily. Furthermore, the FARs of ASTL-GP are also much smaller than those of GP. This further validates the effectiveness of ASTL-GP.

Furthermore, the proposed unsupervised monitoring method is compared with three other state-of-the-art monitoring methods: deep autoencoder (DAE), long short-term memory (LSTM) and wavelet spectrum-based control chart (WS-CC). The first and the second were applied to wind turbines [28,29], where the residuals between the true observations and the predictions by DAE and LSTM were used as the monitoring indices. The last was applied to a steam turbine generator based on temperature signals [38]. The configurations of these models were selected as follows.

DAE: The numbers of neurons in the input and output layers were 12 and 1, respectively. The encoder was composed of 3 hidden layers. The number of neurons in each hidden layer were 64, 32 and 16, respectively. The batch size was 32 and the epochs were 100.

LSTM: The network consisted of three layers. The dimensions of the input and output were 12 and 1, respectively. The number of neurons in the hidden layer of LSTM was 16. The batch size was 16 and the epochs were 100. The dropout was set to 0.5.

WS-CC: Four-leveled decomposition was conducted by the Haar wavelet transform. The T^2 statistic was used to construct the condition index.

Moreover, the mse (mean squared error) and the Adam optimizer were respectively selected as the loss function and the optimization method for both DAE and LSTM models.

The comparative monitoring results by the different methods are shown in Table 3. Obviously, the existing methods were designed for process monitoring with only one normal operation mode and one abnormal mode, whereas the proposed method can adaptively monitor the multimode processes without priors. From Table 4, DAE and LSTM can accurately identify the transition point between Mode-1 and Mode-2. This is because deep learning methods have excellent nonlinear fitting abilities. However, they misidentify the transition point as an anomaly and ignore the multiple normal operating modes. Moreover, the it is difficult for deep learning methods to quantify the uncertainties in the steam turbine operation. The anomaly detected by WS-CC is earlier than those of other methods. This suggests that WS-CC is most sensitive to the changes in observations and gives earlier alarms. Hence, the wavelet decomposition is not suitable for low frequency data. Even in the CM of Mode-1, the proposed method still has the lowest FAR. This is because that both the uncertainty and correlations in the monitoring data are considered in the proposed method. As more observations are available, the characteristics of the operating process will be captured accurately, which finally results in precise predictions and robust detection performance.

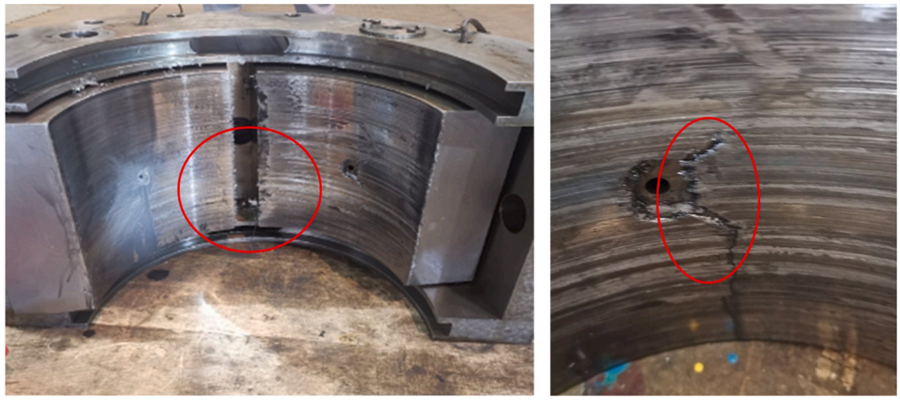

Fig. 11. Bearing bushes with severe wearing out and generated cracks.

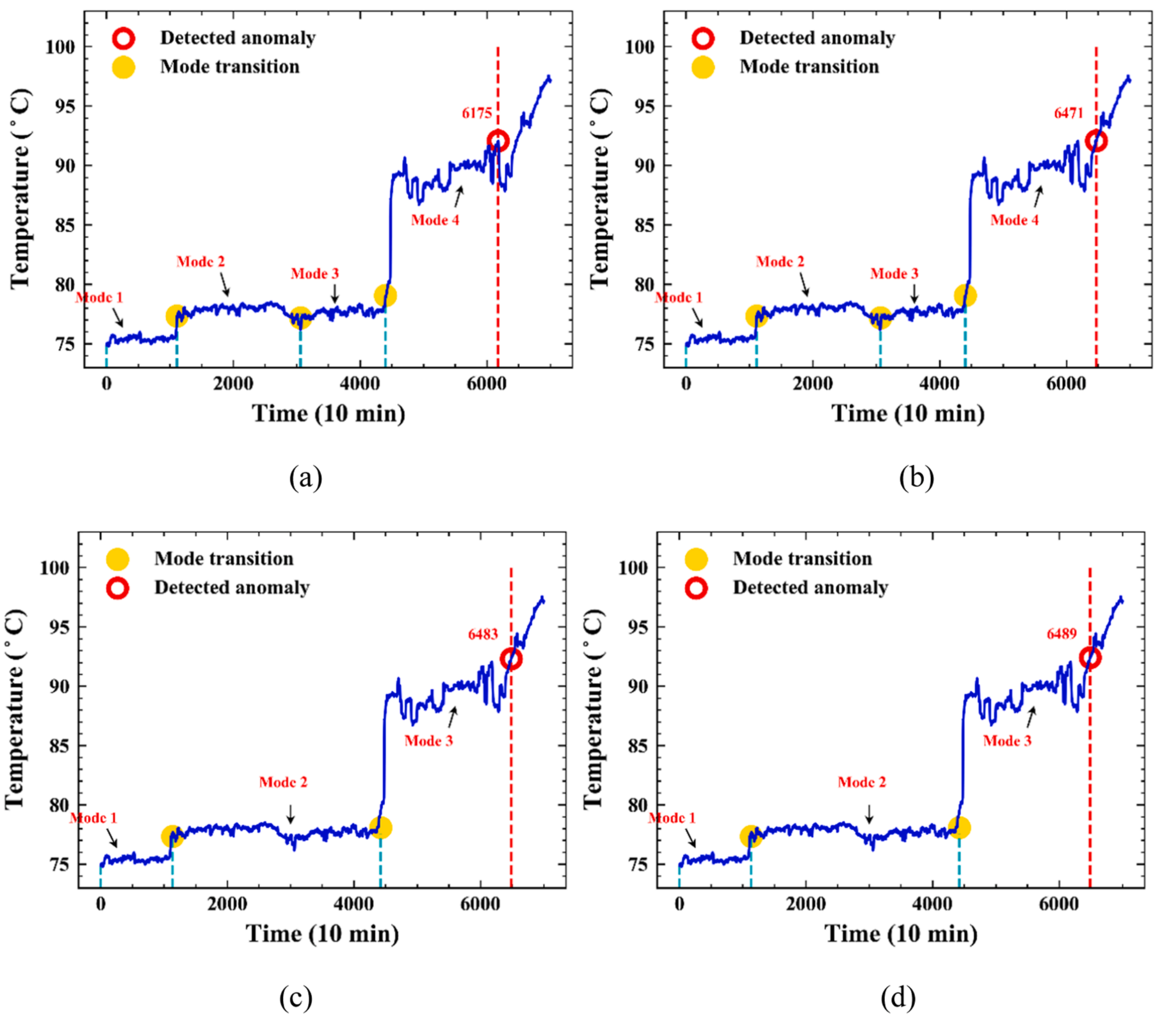

Fig. 12. Anomaly detection results under different criteria. (a) 6 consecutive observations; (b) 12 consecutive observations; (c) 24 consecutive observations; (d) 30 consecutive observations.

5.3. Discussions

CM and fault detection for equipment plays an important role in reliability engineering, which can provide useful information for condition-based maintenance and health management. According to the comparative results above, the proposed method can identify different operational modes and detect the abnormality accurately and effectively. When the anomaly was detected earlier as in DAE, LSTM and WS-CC, wrong false alarms occurred. Then, the shutdown and maintenance actions triggered in advance could make a great waste of the production potential in steam turbine. On the other hand, if the anomaly was detected later than the occurrence of true faults, it would lead to extra maintenance cost and economic loss. Therefore, a proper and reasonable CM method can not only make full use of the steam turbine's performance, but also significantly reduce management costs. Furthermore, both the availability and the economic efficiency of steam turbines can

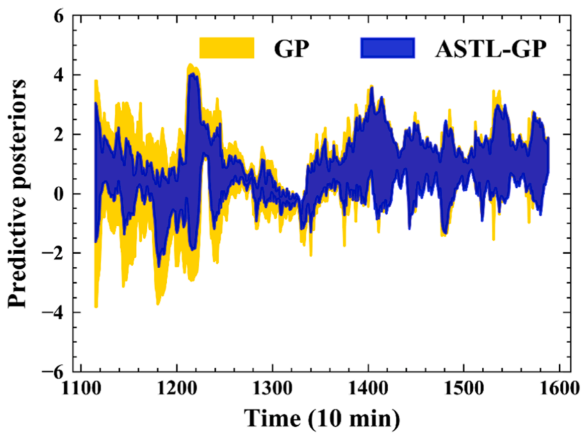

Fig. 13. Comparison between ASTL-GP and GP for the CM of Mode-2.

Table 3
Comparative monitoring results by GP and ASTL-GP.

Metric	Model	Mode-2	Mode-3	Mode-4	Anomaly
Starting time	GP	1120	–	4413	6477
	ASTL-GP	1120	3069	4410	6477
FAR	GP	0.52%	–	0.61%	–
	ASTL-GP	0.10%	0.0%	0.51%	–

Table 4
Comparative monitoring results by different methods.

Method	Multimode	Anomaly time	FAR
DAE	No	1115	0.37%
LSTM	No	1120	0.28%
WS-CC	No	529	8.70%
The proposed method	Yes	6477	0.09%

The FAR of the proposed method is calculated based on the CM of Mode-1.

be guaranteed. With the help of the proposed method, the engineers can make more informed reliability and maintenance decisions. Additionally, the proposed CM framework can be extended to other industrial applications if the operational conditions and modes are time-varying or non-deterministic.

6. Conclusions

This paper proposed an online unsupervised multimode process monitoring and anomaly detection framework for an application to steam turbines with the TSI data. The prior fault information including expert knowledge or labelled historical data was not needed. An adaptive self-transfer learning algorithm based on multiple GPs was developed to deal with the paucity of monitoring data at the beginning of a new incoming operating mode. Dynamic time warping was used to measure the similarity between the past mode and the current mode in real time. The similarities which were used to define the kernel covariance matrices of GPs can capture the uncertainties and cross-correlations between different normal modes were captured. A two-hierarchical identification criterion was constructed to first identify the transition points between normal modes and second detect the anomaly. The proposed framework was validated on a real steam turbine and its superiority were fully illustrated. The results shown that the normal mode transition points and the anomaly were distinguished accurately. Four normal operating modes were identified and the detection time of the anomaly gave an early warning before the actual fault occurred. In addition, compared with other state-of-the-art condition monitoring methods, the proposed method can adaptively and exactly monitor the multimode processes, while the existing methods were only capable of identifying one normal mode before the anomaly.

Although the proposed method has better flexibility and applicability in condition monitoring, only the temperature signals were considered. It is worthy investigating how to construct multimode process monitoring models to detect the steam turbine fault based on multivariate TSI data in the future.

CRediT authorship contribution statement

Zhen Chen: Methodology, Validation, Data curation, Writing – original draft. **Di Zhou:** Validation, Software, Formal analysis. **Enrico Zio:** Investigation, Writing – review & editing. **Tangbin Xia:** Visualization. **Ershun Pan:** Conceptualization, Methodology, Supervision.

Declaration of Competing Interest

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

Data availability

Data will be made available on request.

Acknowledgments

The authors would like to thank the Editor and anonymous referees for their careful work and remarkable comments which considerably help to improve this manuscript substantially. This work is supported in part by the National Key Research and Development Program of China under Grant 2020YFB1711100, and in part by National Natural Science Foundation of China under grant number 72001138, grant number 72071127 and grant number 52005327.

References

[1] Qu F, Liu J, Zhu H, Zang D. Wind turbine condition monitoring based on assembled multidimensional membership functions using fuzzy inference system. IEEE Trans Ind Informat 2020;16(6):4028–37.
[2] Gu H, Si F, Cui Y, Zhu H, Cui X. Information entropy theory for steam turbine system monitoring study. Eng Rep 2020;11(2):1–10.
[3] Zio E. Prognostics and health management (PHM): where are we and where do we (need to) go in theory and practice. Reliab Eng Syst Safe 2022;218:108119.
[4] Tolo S, Tian X, Bausch N, Becerra V, Santhosh TV, Vinod G, Patelli E. Robust on-line diagnosis tool for the early accident detection in nuclear power plants. Reliab Eng Syst Safe 2019;1(186):110–9.
[5] Fallahi F, Bakir I, Yildirim M, Ye Z. A chance-constrained optimization framework for wind farms to manage fleet-level availability in condition-based maintenance and operations. Renw Sust Energy Rev 2022;168:112789.
[6] Zhao X, Ru D, Wang P, Gan L, Wu H, Zhong Z. Fatigue life prediction of a supercritical steam turbine rotor based on neural networks. Eng Fail Anal 2021;127:105435.
[7] Quintanar-Gago DA, Nelson PF, Diaz-Sanchez A, Boldrick MS. Assessment of steam turbine blade failure and damage mechanisms using a Bayesian network. Reliab Eng Syst Safe 2021;207:107329.
[8] Zeng XJ, Yang M, Bo YF. Gearbox oil temperature anomaly detection for wind turbine based on sparse Bayesian probability estimation. Int J Elec Power 2020;123:106233.
[9] Quintanilha IM, Elias VR, da Silva FB, et al. A fault detector/classifier for closed-ring power generators using machine learning. Reliab Eng Syst Safe 2021;212:107614.
[10] Zhang Y, Dong ZY, Kong W, Meng K. A composite anomaly detection system for data- driven power plant condition monitoring. IEEE Trans Ind Informat 2019;16(7):4390–402.
[11] Zhou H, Genez TA, Brintrup A, Parlikad AK. A hybrid-learning decomposition algorithm for competing risk identification within fleets of complex engineering systems. Reliab Eng Syst Safe 2022;1(217):107992.
[12] Yang W, Liu C, Jiang D. An unsupervised spatiotemporal graphical modeling approach for wind turbine condition monitoring. Renew Energy 2018;127:230–41.
[13] Andrade Melani AH, Carvalho Michalski MA, Silva RF, Souza GF. A framework to automate fault detection and diagnosis based on moving window principal

component analysis and Bayesian network. Reliab Eng Syst Safe 2021;1(215):107837.
[14] Hendrickx K, Meert W, Mollet Y, Gyselinck J, Cornelis B, Gryllias K, Davis J. A general anomaly detection framework for fleet-based condition monitoring of machines. Mech Syst Signal Process 2020;139:106585.
[15] Jin X, Xu Z, Qiao W. Condition monitoring of wind turbine generators using SCADA data analysis. IEEE Trans Sustain Energy 2021;12(1):202–10.
[16] Zhang L, Lin J, Ramin K. An angle-based subspace anomaly detection approach to high- dimensional data: with an application to industrial fault detection. Reliab Eng Syst Safe 2015;142:482–97.
[17] Cattaneo L, Polenghi A, Macchi M. A framework to integrate novelty detection and remaining useful life prediction in industry 4.0-based manufacturing systems. Int J Comput Integ Manuf 2021;16:1–21.
[18] Lee H, Kim NW, Lee JG, Lee BT. An approach for utilizing correlation among sensors for unsupervised anomaly detection of wind turbine system. In: 2021 International conference on information and communication technology convergence (ICTC). IEEE; 2021. p. 104–9. October.
[19] Fu S, Zhong S, Lin L, Zhao M. A re-optimized deep auto-encoder for gas turbine unsupervised anomaly detection. Eng Appl Artif Intel 2021;101:104199.
[20] Dao PB, Staszewski WJ, Barszcz T, Uhl T. Condition monitoring and fault detection in wind turbines based on cointegration analysis of SCADA data. Renew Energy 2018;116:107–22.
[21] Kenyon AD, Catterson VM, McArthur SD, Twiddle J. An agent-based implementation of hidden Markov models for gas turbine condition monitoring. IEEE Trans Syst Man Cybern 2013;44(2):186–95.
[22] Wang L, Zhang Z, Long H, Xu J, Liu R. Wind turbine gearbox failure identification with deep neural networks. IEEE Trans Ind Informat 2016;13(3):1360–8.
[23] González-Muñiz A, Díaz I, Cuadrado AA. García-Pérez D. Health indicator for machine condition monitoring built in the latent space of a deep autoencoder. Reliab Eng Syst Safe 2022;224:108482.
[24] Lyu D, Niu G, Liu E, Zhang B, Chen G, Yang T, et al. Time space modelling for fault diagnosis and prognosis with uncertainty management: a general theoretical formulation. Reliab Eng Syst Safe 2022;226:108686.
[25] Lyu D, Niu G, Liu E, Yang T, Chen G, Zhang B. Uncertainty management and differential model decomposition for fault diagnosis and prognosis. IEEE Trans Ind Electron 2021;69(5):5235–46.
[26] Zhang Y, Li M, Dong ZY, Meng K. Probabilistic anomaly detection approach for data-driven wind turbine condition monitoring. CSEE J Power Energy Syst 2019;5(2):2019.
[27] Song N, Hu X, Li N. Anomaly detection of wind turbine generator based on temporal information. In: Proceedings of the 2019 7th international conference on information technology: IoT and smart city; 2019. p. 477–82. Dec.
[28] Qiao W, Lu D. A survey on wind turbine condition monitoring and fault diagnosis-Part II: signals and signal processing methods. IEEE Trans Ind Electron 2015;62(10):6546–655.
[29] Janssens O, Slavkovikj V, Vervisch B, Stockman K, Loccufier M, Verstockt S, Walle RV, Van Hoecke S. Convolutional neural network-based fault detection for rotating machinery. J Sound Vib 2016;377:331–45.
[30] Ince T, Kiranyaz S, Eren L, Askar M, Gabbouj M. Real-time motor fault detection by 1-D convolutional neural networks. IEEE Trans Ind Electron 2016;63(11):7067–75.
[31] Pandit RK, Infield D. SCADA-based wind turbine anomaly detection using Gaussian process models for wind turbine condition monitoring purposes. IET Ren Power Gen 2018;12(11):1249–55.
[32] Chen J, Li J, Chen W, Wang Y, Jiang T. Anomaly detection for wind turbines based on the reconstruction of condition parameters using stacked denoising autoencoders. Renew Energy 2020;147:1469–80.
[33] Renström N, Bangalore P, Highcock E. System-wide anomaly detection in wind turbines using deep autoencoders. Renew Energy 2020;157:647–59.
[34] Chen H, Liu H, Chu X, Liu Q, Xue D. Anomaly detection and critical SCADA parameters identification for wind turbines based on LSTM-AE neural network. Renew Energy 2021;172:829–40.
[35] Zhang C, Hu D, Yang T. Anomaly detection and diagnosis for wind turbines using long short-term memory-based stacked denoising autoencoders and XGBoost. Reliab Eng Syst Safe 2022;222:108445.
[36] Zhang L, Lin J, Shao H, Zhang Z, Yan X, Long J. End-to-end unsupervised fault detection using a flow-based model. Reliab Eng Syst Safe 2021;1(215):107805.
[37] Yu J, Liu P, Li Z. Hybrid modelling and digital twin development of a steam turbine control stage for online performance monitoring. Renew Sustain Energy Rev 2020;133:110077.
[38] Huang CG, Huang HZ, Li YF. A bidirectional LSTM prognostics method under multiple operational conditions. IEEE Trans Ind Electron 2019;66(11):8792–802.
[39] Jiang Q, Yan X. Monitoring multi-mode plant-wide processes by using mutual information based multi-block PCA, joint probability, and Bayesian inference. Chemometr Intell Lab 2014;136:121–37.
[40] Peng K, Zhang K, You B, Dong J. Quality-related prediction and monitoring of multi- mode processes using multiple PLS with application to an industrial hot strip mill. Neurocomputing 2015;168:1094–103.
[41] Quatrini E, Costantino F, Gravio G, Patriarca R. Machine learning for anomaly detection and process phase classification to improve safety and maintenance activities. J Manuf Syst 2020;56:117–32.
[42] Pan SJ, Yang Q. A survey on transfer learning. IEEE Trans Knowl Data Eng 2021;22:1345–59.
[43] Zhuang F, Qi Z, Duan K, Xi D, Zhu Y, Zhu H, et al. A comprehensive survey on transfer learning. Proc IEEE 2020;109(1):43–76.
[44] Liu X, Du J, Ye ZS. A condition monitoring and fault isolation system for wind turbine based on SCADA data. IEEE Trans Ind Informat 2022;18(2):986–95.
[45] Bae SJ, Mun BM, Chang W, Vidakovic B. Condition monitoring of a steam turbine generator using wavelet spectrum-based control chart. Reliab Eng Syst Safe 2019;184:13–20.
[46] Yu Y, Xiong Q, Ye ZS, Liu X, Li Q, Wang K. A review on acoustic reconstruction of temperature profiles: from time measurement to reconstruction algorithm. IEEE Trans Instrum Meas 2022;77:1–24.
[47] Zhu Y, Zhu C, Tan J, Tan Y, Rao L. Anomaly detection and condition monitoring of wind turbine gearbox based on LSTM-FS and transfer learning. Renw Energy 2022;189:90–103.
[48] Cao B, Pan SJ, Zhang Y, Yeung DY, Yang Q. Adaptive transfer learning. In: Proceedings of the AAAI conference on artificial intelligence. Vol. 24; 2010. p. 1.
[49] Wagle N, Frew EW. Forward adaptive transfer of Gaussian process regression. J Aerosp Informat Syst 2017;14(4):214–31.
[50] Yang K, Lu J, Wan W, Zhang G, Hou L. Transfer learning based on sparse Gaussian process for regression. Inform Sci 2022;605:286–300.
[51] Tighineanu P, Skubch K, Baireuther P, Reiss A, Berkenkamp F, Vinogradska J. Transfer learning with gaussian processes for Bayesian optimization. In: International conference on artificial intelligence and statistics; 2022. p. 6152–81. May.