Real-time Novelty Detection of an Industrial Gas Turbine using Performance Deviation Model and Extreme Function Theory

Xiwen Gu [a], Shixi Yang [a,*] , Yongfeng Sui [b], Evangelos Papatheou [c], Andrew D. Ball [d] and Fengshou Gu [d]

[a] *The State Key Laboratory of Fluid Power and Mechatronic Systems, School of Mechanical Engineering, Zhejiang University, Hangzhou, 310027, China*
[b] *The Hangzhou Steam Turbine Power Group Co. Ltd., Hangzhou, 310022, China*
[c] *The College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, EX4 4QF, UK*
[d] *The School of Computing and Engineering, University of Huddersfield, Huddersfield, HD1 3DH, UK*
[*] *Corresponding author email: yangsx@zju.edu.cn*

Abstract: Novelty detection is crucial to ensure the availability and reliability of an industrial gas turbine. With the application of modern health monitoring systems, there is an ample amount of data gathered from gas turbines, however they are usually from normal events with limited knowledge of any novelty. In current practice, the unknown event is detected by comparing with a model of normality through pointwise approaches, which is inefficient in terms of false alarms or missing alarms. This paper proposes an accurate novelty detection approach using performance deviation model and extreme function theory. The model is established from the multi-sensor real-time performance data. Outputs of the model, that is, the deviation curves, are considered as functions instead of individual data points to test the status of the system as 'normal' or 'abnormal' by the extreme value theory. The effectiveness of the proposed approach is demonstrated by the monitoring data from a single shaft gas turbine on site. Compared with other traditional methods, the proposed approach is superior in terms of high detection accuracy and high sensitivity with a good balance between the false alarm rate and missing alarm rate. This paper provides a reliable approach for the real-time health monitoring of the industrial gas turbines.

Keywords: Industrial gas turbines; Novelty detection; Performance deviation; Extreme functions; False alarms; Missing alarms.

1. Introduction

The industrial gas turbine (GT) is widely used in power plants and distributed energy systems. It is a complex thermodynamic system working under harsh environment and transient conditions. The improvement of availability and reliability of GTs are crucial to avoid unpredicted breakdowns and decrease the operation costs [1]. The performance-based health monitoring of GTs gathers real-time health-related information of the system and helps to make condition-based maintenance strategy, which has raised widely attention among manufacturers and operators [1,2].

The first step to describe the health state of the system is to decide whether any damage exists, usually through novelty detection approaches [3]. Novelty detection is defined as the task to identify new behaviours different from the normal expectations [4]. Usually, the available performance data of GTs are of low sampling rate so abstracting the abnormal features through traditional signal processing methods like spectrum analysis is difficult [5]. Some fault classification schemes require both normal data and abnormal data to train a good classifier [6]-[8]. They are unsuitable for application on site as the monitoring data with abnormal labels are insufficient or even unavailable for such a safety-critical system. Liu et al. [9] propose a Support Vector Machine (SVM) framework for fault detection with unbalanced data. One class learning methods including one-class SVM [10] and isolation forest [11] are also proved to be effective to deal with the datasets where there are none or very few examples of the minority class. Considering the realistic situation, the change of the ambient conditions and operation conditions have great influences on the monitoring parameters of the GT, which could increase the false alarm rate of the system if the classify is sensitive to the between-class separability of the features. Therefore, the real-time novelty detection of the GT with only normal data is challenging. For industrial GTs, the novelty detection is carried out based on a performance model of normality. The model should take into consideration the complicated nonlinearity of GT operations. The unknown event is then tested by the model, resulting in an indicator describing the novelty. The indicator is then compared with a threshold. Data lying outside the boundary may indicate an abnormal event [12].

The performance models of GTs are usually categorized into the physical-based and the data-driven ones. The physical-based model is constructed by complex thermodynamic and energy balance equations. Various physical-based models of GTs have been proposed on different configurations [13], loading conditions [14], transient behaviours [15] among others. Toolboxes [16]-[18] have also been developed for the modelling and thermodynamic simulation of GTs. However, accurate modelling is very challenging because of the highly complicated and nonlinear physics of GT operation. In addition, the nonlinear physics also demands high computational cost, which makes physical-based model unsuitable for real-time monitoring. With large datasets acquired from the health monitoring systems, data-driven models draw more attention recently and are commonly constructed through machine learning methods like autoassociative neural networks [19], nonlinear autoregressive exogenous [20], genetic programming [21], convolutional neural network [22] among

others. A data-driven model could learn the relationships between the variables of a system directly without a priori knowledge of the GT specification or struggling with the complicated dynamic equations. A data-driven model based on the real operation data could provide more reliable and accurate prediction of the GT performance [23].

The deviations between the model predictions and the measured values could be treated as fault indicators [6,24]. A threshold is established based on the statistics of normal operations to classify the deviations. For the GTs on site, adequate datasets of normality have been well gathered since the use of the monitoring systems, so the model of normal operations constructed by various performance parameters could be accurate and comprehensive. The calculated threshold is taken as the decision boundary to decide whether the deviation is novelty, that is to say, a potential fault. When the GT is running in an abnormal condition, the performance deviations are expected to exceed the threshold [25]. The threshold of the deviation values should be carefully selected. Small thresholds may increase the false alarms while too large thresholds may bring missing alarms and reduce the detection sensitivity. Traditional thresholds are calculated by the mean and variance values of the measurements based on Pauta Criterion, also known as 3σ principle [26]. Amirkhani et al. [27] set the thresholds through uncertainty analysis of the system based on Monte Carlo simulation. Toshkova et al. [28] fit the extreme value distribution of the data points and the extreme value is set as the threshold. Generally, the thresholds are set through a pointwise approach so the novelty detection is performed by assessing every single point in a time series. Individual points in time series are not independent and identically distributed [29]. When dealing with the data points from real and complex GT systems, the measurement error, the harsh environment and the variable working conditions could all affect the detection results, so pointwise comparison is not always accurate enough for health monitoring of industrial GTs on site.

Extreme Function Theory (EFT) proposed by Clifton is an effective method to assess functions instead of discrete points with improvement in classification results [30]. The functions are represented by series of points like the power curves of a wind turbine [29]. This paper proposes a novelty detection scheme combining a performance deviation model and the EFT. The model is established by the Backpropagation Neural Network (BPNN) to obtain the deviation curves. The decision threshold is set using the deviation curve as a function by the modified version of EFT proposed in [29]. The main contributions of the paper are summarized as follows.

1) A new BPNN-EFT approach is proposed for novelty detection of the industrial GTs for the first time. The proposed approach fuses the multi-sensor real-time performance data to establish the performance deviation model and detects the fault using only normal data.

2) A data-driven performance deviation model is established to predict the performance of the GT accurately. The inputs and outputs of the performance deviation model are carefully chosen considering the ambient conditions, operation conditions as well as the physical conditions of the GT.

3) The EFT is for the first time applied in the field of GT novelty detection and modified to be more intelligent to fit the proper extreme value distribution. A threshold is calculated by EFT using the performance deviation curves to realize novelty detection with high accuracy and sensitivity.

4) Real-time datasets from a Mitsubishi single shaft GT on site are used to demonstrate superiorities of the proposed approach for real-time novelty detection compared with other novelty detection methods.

The layout of the paper consists of 5 sections. The second section formulates the novelty detection approach combining a performance deviation model and EFT. The third section describes the background of the case study, including basic process of the GT, the problems, the available data, and the general methodology. The fourth section is the application of the approach with a discussion on results in comparison with other methods. Finally, a conclusion is shown in the fifth section.

2. Proposed novelty detection approach

2.1 Performance deviation model

Artificial Neural Network (ANN) has been considered a typical data-driven approach for system identification and performance prediction [31,32]. The nonlinear relationships between the inputs and outputs can be learnt through the iterative training process. The BPNN as one popular type of ANN is widely used for modelling and diagnosis of GTs [6,23,33,34]. It has been proved to be a powerful tool to predict the performance of GTs and capture the complicated nonlinearity of GT operations.

The performance deviation model based on BPNN is established as follows. Let $\mathbf{X}$ be the matrix of input parameters and $\mathbf{Y}$ denotes the output parameters. Both $\mathbf{X}$ and $\mathbf{Y}$ are of normal status. A BPNN is trained to learn the relationships between training samples $\mathbf{X}$ and $\mathbf{Y}$. The parameters of the BPNN are optimized by minimizing the loss function. Assuming $\mathbf{f}^*$ the predicted outputs based on the trained network, the performance deviation values are given by

$$\Delta = \mathbf{f}^* - \mathbf{Y}^* \tag{1}$$

where $\mathbf{Y}^*$ forms the measured outputs. When the testing inputs are of normal status, the deviations between the predicted outputs and the measured outputs are supposed to be small. With abnormal parameters inputting into the model, the corresponding deviation values may increase so the deviation curves imply the condition of the system.

2.2 Extreme function theory

2.2.1 Gaussian Process regression

EFT is a combination of Gaussian Process (GP) and extreme value distribution. A GP is defined as a collection of random variables, any finite number of which have a joint Gaussian distribution [35]. For a real process $f(\mathbf{x})$, a GP is specified by its mean function $m(\mathbf{x})$ and the covariance function $\mathrm{k}(\mathbf{x},\mathbf{x}')$.

$$\mathrm{f}(\mathbf{x}) \sim \mathrm{GP}(\mathrm{m}(\mathbf{x}), \mathrm{k}(\mathbf{x},\mathbf{x}')) \tag{2}$$

where

$$m(\mathbf{x}) = \mathrm{E}\left[f(\mathbf{x})\right] \tag{3}$$

$$k(\mathbf{x},\mathbf{x}') = \mathrm{E}\left[\left(f(\mathbf{x})-m(\mathbf{x})\right)\left(f(\mathbf{x}')-m(\mathbf{x}')\right)\right] \tag{4}$$

$\mathrm{m}(\mathbf{x})$ is the expected function value of input $\mathbf{x}$. $\mathrm{m}(\mathbf{x})$ is always set to be 0 to avoid expensive computations as in many other studies [35]-[37]. $\mathrm{k}(\mathbf{x},\mathbf{x}')$ describes the dependence between the function values at inputs $\mathbf{x}$ and $\mathbf{x}'$. The squared exponential covariance function is a popular choice as the covariance function defined as [35]

$$\mathrm{k}(\mathbf{x},\mathbf{x}') = \sigma_{\mathrm{f}}^2 \exp(-\left\|\mathbf{x}-\mathbf{x}'\right\|^2 / 2\mathrm{l}^2) \tag{5}$$

where σ_{f} and l are hyper parameters and the double bar '||' refers to the Euclidean norm.

For realistic modelling situations, the observations differ from the function values by a Gaussian noise ε with variance σ_{n}^2, which are $\mathbf{y} = \mathrm{f}(\mathbf{x}) + \varepsilon$. Assuming a training set $\{\mathbf{X},\mathbf{y}\}$ where $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_{\mathrm{n}}]^{\mathrm{T}}$, the covariance function becomes

$$\mathrm{cov}(\mathbf{y}) = \mathrm{K}(\mathbf{X},\mathbf{X}) + \sigma_{\mathrm{n}}^2\mathbf{I} \tag{6}$$

where $\mathbf{I}$ is the unit matrix. The prior of the noisy training observations $\mathbf{y}$ and the testing output $\mathbf{f}^*$ at the testing input matrix $\mathbf{X}^*$ is

$$\mathbf{y} \sim N(\mathbf{0}, \mathrm{K}(\mathbf{X},\mathbf{X}) + \sigma_{\mathrm{n}}^2\mathbf{I}) \text{ and} \tag{7}$$

$$\mathbf{f}^* \sim N(\mathbf{0}, \mathrm{K}(\mathbf{X}^*,\mathbf{X}^*)) \tag{8}$$

respectively. The joint distribution of the training and testing outputs under the prior is then defined as

$$\begin{bmatrix}\mathbf{y}\\ \mathbf{f}^*\end{bmatrix} \sim N\left(\mathbf{0}, \begin{bmatrix}\mathrm{K}(\mathbf{X},\mathbf{X}) + \sigma_{\mathrm{n}}^2\mathbf{I} & \mathrm{K}(\mathbf{X},\mathbf{X}^*)\\ \mathrm{K}(\mathbf{X}^*,\mathbf{X}) & \mathrm{K}(\mathbf{X}^*,\mathbf{X}^*)\end{bmatrix}\right) \tag{9}$$

where $\mathrm{K}(\mathbf{X},\mathbf{X}^*)$ is the covariance matrix between the training inputs and the testing inputs.

$$\mathrm{K}(\mathbf{X},\mathbf{X}^*) = \begin{bmatrix}\mathrm{k}(\mathbf{x}_1,\mathbf{x}_1^*) & \mathrm{k}(\mathbf{x}_1,\mathbf{x}_2^*) & \cdots & \mathrm{k}(\mathbf{x}_1,\mathbf{x}_{\mathrm{n}}^*)\\ \mathrm{k}(\mathbf{x}_2,\mathbf{x}_1^*) & \mathrm{k}(\mathbf{x}_2,\mathbf{x}_2^*) & \cdots & \mathrm{k}(\mathbf{x}_2,\mathbf{x}_{\mathrm{n}}^*)\\ \vdots & \vdots & \ddots & \vdots\\ \mathrm{k}(\mathbf{x}_{\mathrm{n}},\mathbf{x}_1^*) & \mathrm{k}(\mathbf{x}_{\mathrm{n}},\mathbf{x}_2^*) & \cdots & \mathrm{k}(\mathbf{x}_{\mathrm{n}},\mathbf{x}_{\mathrm{n}}^*)\end{bmatrix} \tag{10}$$

It is similar for the rest of the covariance matrices $\mathrm{K}(\mathbf{X},\mathbf{X})$, $\mathrm{K}(\mathbf{X}^*,\mathbf{X})$ and $\mathrm{K}(\mathbf{X}^*,\mathbf{X}^*)$.

The predictive distribution is defined by conditioning the joint prior on the observations as

$$\mathbf{f}^* \left| \mathbf{X},\mathbf{y},\mathbf{X}^* \right. \sim N\left(\mathrm{m}(\mathbf{f}^*), \mathrm{cov}(\mathbf{f}^*)\right) \tag{11}$$

where

$$\mathrm{m}(\mathbf{f}^*) = \mathrm{K}(\mathbf{X}^*,\mathbf{X})\left[\mathrm{K}(\mathbf{X},\mathbf{X}) + \sigma_{\mathrm{n}}^2\mathbf{I}\right]^{-1}\mathbf{y} \text{ and} \tag{12}$$

$$\mathrm{cov}(\mathbf{f}^*) = \mathrm{K}(\mathbf{X}^*,\mathbf{X}^*) - \mathrm{K}(\mathbf{X}^*,\mathbf{X})\left[\mathrm{K}(\mathbf{X},\mathbf{X}) + \sigma_{\mathrm{n}}^2\mathbf{I}\right]^{-1}\mathrm{K}(\mathbf{X},\mathbf{X}^*) . \tag{13}$$

The hyper parameters $\boldsymbol{\theta} = (\sigma_f, l, \sigma_n^2)$ are optimized by the maximum marginal likelihood, which is

$$\log \mathrm{p}(\mathbf{y}\left|\mathbf{X},\boldsymbol{\theta}\right.) = -0.5\mathbf{y}^{\mathrm{T}}\mathrm{cov}(\mathbf{y})^{-1}\mathbf{y} - 0.5\log\left|\mathrm{cov}(\mathbf{y})\right| - 0.5\mathrm{n}\log 2\pi \tag{14}$$

The marginal log likelihood is maximized by the gradient ascent-based optimization tool using the partial derivatives of Equation (14).

$$\frac{\partial}{\partial\theta_{\mathrm{j}}}\log \mathrm{p}(\mathbf{y}\left|\mathbf{X},\boldsymbol{\theta}\right.) = 0.5\mathbf{y}^{\mathrm{T}}\mathrm{cov}(\mathbf{y})^{-1}\mathbf{y} - 0.5\mathrm{tr}\left(\mathrm{cov}(\mathbf{y})^{-1}\frac{\partial \mathrm{cov}(\mathbf{y})}{\partial\theta_{\mathrm{j}}}\right) = 0.5\mathrm{tr}\left(\left(\boldsymbol{\alpha}\boldsymbol{\alpha}^{\mathrm{T}} - \mathrm{cov}(\mathbf{y})^{-1}\right)\frac{\partial \mathrm{cov}(\mathbf{y})}{\partial\theta_{\mathrm{j}}}\right) \tag{15}$$

where $\boldsymbol{\alpha} = \mathrm{cov}(\mathbf{y})^{-1}\mathbf{y}$ .The details and implementation of the GP algorithm can be seen in [35].

2.2.2 Extreme function theory based on Gaussian Process

Considering the time series of observations as functions, a GP model of normality is constructed based on the training samples of normal functions. The whole testing function $\mathbf{f}^* = \mathrm{f}(\mathbf{x}^*)$ should have a corresponding probability density $\mathbf{z} = \mathrm{f}_{\mathrm{n}}(\mathbf{f}^*)$, where $\mathrm{f}_{\mathrm{n}} = \mathrm{p}(\mathbf{f}^*\left|\mathbf{x},\mathbf{y},\mathbf{x}^*\right.)$ refers to the multivariate Gaussian distribution. According to the definition of a GP, f_{n} takes the form

$$\mathbf{z} = \left(1/\sqrt{(2\pi)^{\mathrm{d}}\left|\mathrm{cov}(\mathbf{f}^*)\right|}\right) \exp\left(-0.5\left(\mathbf{f}^* - \mathrm{m}(\mathbf{f}^*)\right)^{\mathrm{T}}\mathrm{cov}(\mathbf{f}^*)^{-1}\left(\mathbf{f}^* - \mathrm{m}(\mathbf{f}^*)\right)\right) \tag{16}$$

where d is the dimension of the testing input vector $\mathbf{x}^*$. $\mathrm{m}(\mathbf{f}^*)$ and $\mathrm{cov}(\mathbf{f}^*)$ are given by Equation (12) and Equation (13) respectively.

The basic idea of the EFT is to test whether a function is extreme or not compared to the model. The 'most extreme' is defined as the function with the lowest probability density z [30]. The function becomes more extreme when it moves away from the mean function. With considerable normal testing functions, the extremum of these functions could be estimated, which is the alarm threshold.

The low values of probability density z converge to the extreme value distribution with three possible forms: Weibull, Frechet and Gumbel [38]. The generalized extreme value (GEV) distribution is used in this paper as it is uncertain which distribution it will converge to. The minima distribution function of GEV is defined as

$$L(lnz,\mu,\sigma,\gamma) = 1 - \exp\left(-\left(1-\gamma\left(lnz-\mu\right)/\sigma\right)^{-1/\gamma}\right) \tag{17}$$

where lnz=ln(z) and μ, σ and γ are location, scale and shape parameters. The probability density z is treated on its logarithmic form for easier manipulation. The parameters of the cumulative distribution function (CDF) are optimized by the normalized mean squared error (NMSE) function, given by

$$\mathrm{NMSE}(\hat{\mu},\hat{\sigma},\hat{\gamma}) = \left(100/n\sigma_{\mathrm{p}}^2\right)\sum_{i=1}^{n}\left(L_{\mathrm{p}} - L\right)^2 \tag{18}$$

where L_p is the empirical CDF, L is the fitted CDF calculated by Equation (17), n is the number of points to fitting the CDF and σ_{p} is the standard deviation of L_{p}. The optimization

algorithm is differential evolution used in [29]. With the best parameters estimated on the lowest NMSE, a threshold is calculated with a confidence interval of 99%, which means that 1% of the normal functions could be wrongly identified.

3. Case study

3.1 Description of the GT

The case study in this paper considers a Mitsubishi single shaft GT M251S with around 29.8 MW generation capacity. It forms part of a Combined Cycle Power Plant (CCPP) located in a steelwork in northeast China. The GT consists of a 19-stage axial-flow compressor, a can-annular combustor, and a 3-stage turbine. As shown in Fig. 1, the air is drawn through an air filter to the compressor, where the air is compressed and passed to the combustor. Mixed with the fuel, the mixture gases are burnt in the combustor. The resulting hot gases expand when flowing through the turbine to rotate the shaft. The generator at the compressor side is then driven to produce power.

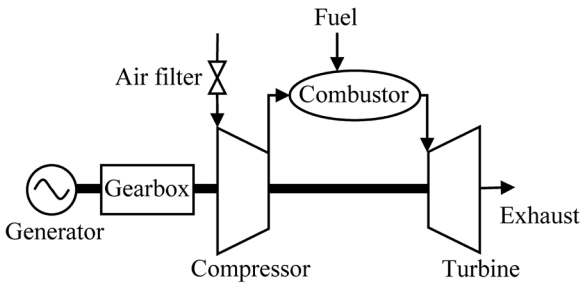

Fig. 1. Schematic of an industrial gas turbine M251S in the CCPP.

3.2 Problem statement and data

The case study presented includes recorded data from the GT on site in steady state operation. Table 1 displays 12 representative parameters. The compressor inlet pressure P_{in} and compressor inlet temperature T_{in} are the barometric pressure and the ambient temperature representing the ambient conditions. The load set L_{set} is the desired operation condition of the GT provided by the control system. By controlling the fuel gas flow m and the position of the inlet guide vane, which remains unchanged in this case, the generator output power W, the rotational speed N and the exhaust temperature T_{ex} of the turbine are kept in the safe area. The rotor vibrations at 4 different places and the exhaust pressure, that is, the turbine outlet pressure P_{ex} directly reflect the physical conditions of the GT. The T_{ex} is an important outlet parameter considerably affected by the ambient conditions, operation conditions as well as the physical conditions of the GT.

In a main overhaul, different cracks were found after inspection on the blades and the seals of the compressor as shown in Fig. 2. The data provided have been classified as normal and abnormal by the operators. The normal data available were gathered after the overhaul in a period of 26 days while 18 days of data from abnormal status were gathered before the overhaul. All data points recorded are the peak to peak values in 1-minute interval. Fig. 3 displays the values of the parameters in different status when taking the exhaust temperature T_{ex} and the rotor vibration on the compressor side V_{com} as examples.

Table 1 Available parameters provided by the operators

Parameter	Units	Notation
Compressor inlet pressure	kPa	P_{in}
Compressor inlet temperature	°C	T_{in}
Fuel gas flow	Nm^3 / h	m
Load set	MW	L_{set}
Generator output	MW	W
Speed	rpm	N
Rotor vibration on the compressor side	μm	V_{com}
Rotor vibration on the turbine side	μm	V_{turb}
Generator rotor vibration on the GT side	μm	V_{GTin}
Generator rotor vibration on the excitation side	μm	V_{GTex}
Exhaust pressure	kPa	P_{ex}
Exhaust temperature	°C	T_{ex}

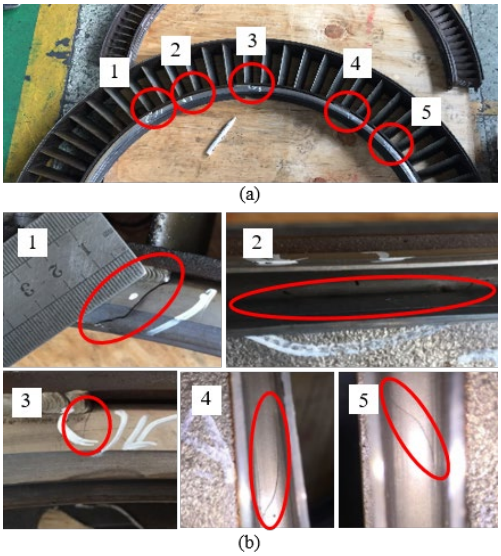

Fig. 2. Photos of the cracks found during the overhaul. (a) The overall photo of the cracks. (b)The partial enlarged details corresponding to the numbers 1-5 in (a).

It is worth noting that though the cracks are very serious, no alarm has been triggered in the actual operation process. Based on the classification results from the operators, it can be seen in Fig. 3(c) and (d) that though the GT has been operating in an abnormal status, the values of the monitoring parameters are relatively stable. The vibration of the compressor rotor turns larger in Fig. 3(b) when cracks exist, but the values are about 15 μm, which is far below the 70 μm of the current alarm threshold set by the operators. In order to detect the abnormality timely and precisely, it is crucial and also the demand of the GT operators to set the proper thresholds using the monitoring parameters.

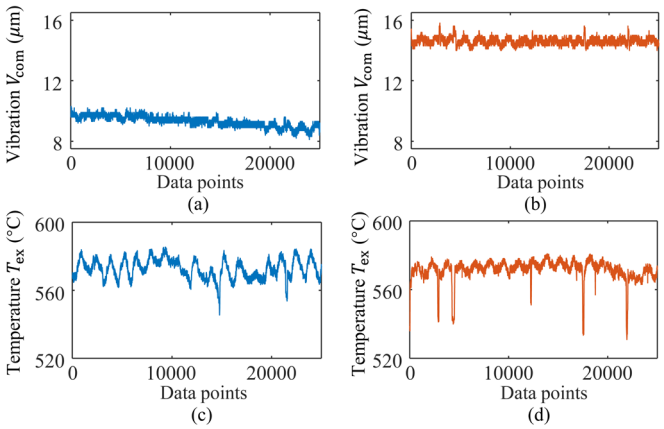

Fig. 3. Values of the performance parameters in different status. (a) The rotor vibration on the compressor side after the overhaul. (b) The rotor vibration on the compressor side with severe cracks. (c) The exhaust temperature of GT after the overhaul. (d) The exhaust temperature of GT with severe cracks.

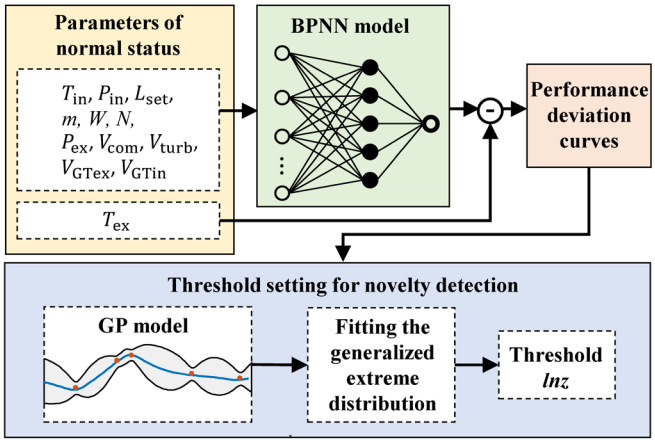

Fig. 4. Overall procedure of performance deviation modeling and threshold setting.

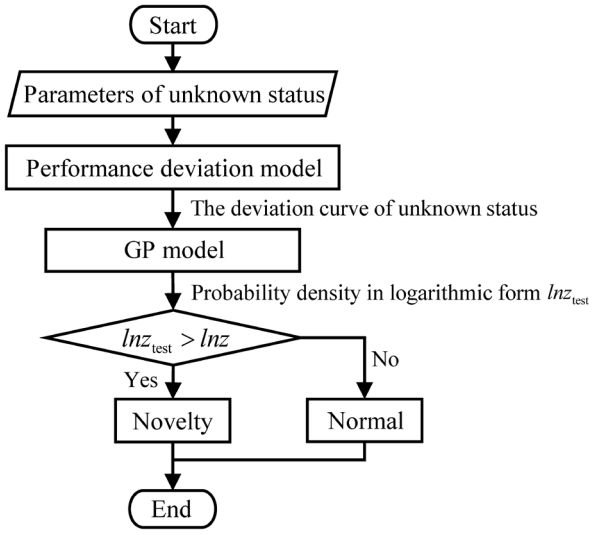

Fig. 5. Flow chart of real-time novelty detection for unknown events.

The proposed novelty detection approach pays attention on the change of the exhaust temperature T_{ex} as it is the key monitoring parameter of a GT to regulate the fuel supply and help control the GT [20]. When establishing the performance deviation model, T_{ex} is chosen as the output while the other parameters are used as the inputs to the network. Fig. 4 presents the overall procedure of the performance deviation modeling including the threshold setting based on EFT. Once the threshold is set, the real-time implementation is carried out for novelty detection according to the flow chart in Fig. 5.

4. Results and discussion

4.1 Training the performance deviation model

In order to get the deviation curves, a BPNN model of normality is developed to predict T_{ex} from P_{in}, T_{in}, m, L_{set}, W, N, V_{com}, V_{turb}, V_{GTin}, V_{GTex} and P_{ex}. 10000 points are sampled randomly from the 11 parameters respectively for training, so the size of the input matrix is 11×10000. All training samples are normalized to [-1,1] before training. The BPNN is developed in MATLAB. The training function is the 'traingdx' algorithm and the loss function is mean squared error. 85% of the samples are randomly picked for training while the rest 15% for validation. Number of hidden layer neurons is chosen as 5 with best fitting on the validation set. The weight and bias values of the network are optimized 5 times to ensure the global optimum. The trained network is tested by all the 26-day normal data. A comparison of the predicted values with the measured T_{ex} is shown in Fig. 6. The average prediction error is 0.16%, meaning that the BPNN model can well represent the stable operation performance of the GT. Fig. 7 shows the histogram of the average prediction errors after 1000 runs of different sampling from the normal data. Though the prediction errors vary, most of them are between 0.15% and 0.18% and they are all below 0.30%.

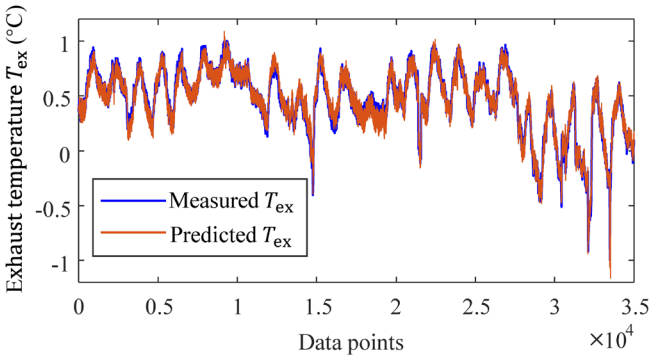

Fig. 6. Predicted GT exhaust temperature and measured exhaust temperature.

By computing the gap between the predicted and measured values, performance deviation values can be obtained as shown in Fig. 8. It can be seen that the deviation values of the normal data are relatively stable around zero while the residuals deviate noticeably from zero when there are cracks in the system, which coincides with the comments in [25]. The deviation curves are

constructed at the resolution of 8 hours including 480 data points because an 8-hour interval is a work shift of the GT operators. It is desired by the operators to monitor the health status of the GT when a work shift ends. Considering all the available data, 132 deviation curves are obtained including 78 deviation curves identified as 'normal' and the rest 54 abnormal curves. Each deviation curve is treated as a sample to train or test as a function in terms of the extremity and novelty.

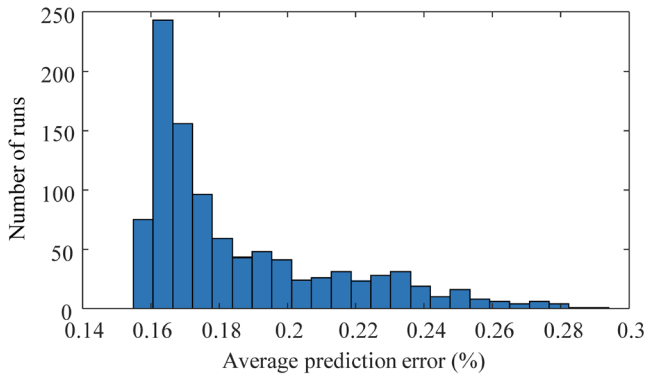

Fig. 7. Histogram of the prediction errors after 1000 runs of different sampling from the normal data.

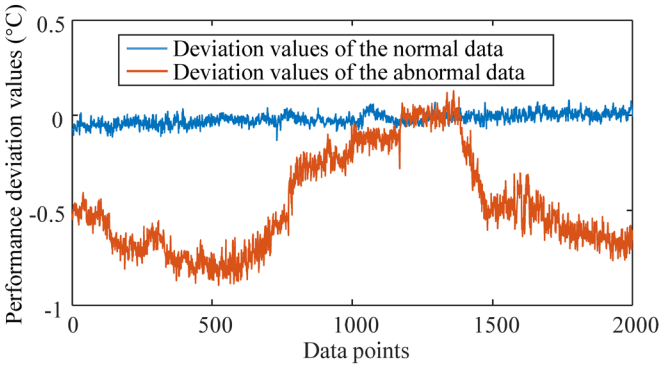

Fig. 8. The performance deviation values of the data in normal or abnormal status.

4.2 Threshold setting and novelty detection

The deviation curves are split into three sets including training, validation, and testing for threshold setting and novelty detection. 100 data points are randomly picked from each curve to reduce the computation cost as the GP is quite efficient in small sample classification [35]. The training set consists of 3 deviation curves in normal status. 40 deviation curves, also in normal status, are used to a create validation set but a bootstrap method is used here to obtain 120 samples. The original deviation curves for validation are randomly and repetitively sampled for 3 times. The rest 35 deviation curves identified as normal and all the 54 abnormal curves compose the testing set.

The GP model is trained by the training set and then the validation samples are put into the model to obtain the corresponding probability density values in the logarithmic form *lnz*. The GEV distribution is fitted and optimized based on the left tail of the probability density values, which refer to 10% of the low values of *lnz*. Fig. 9 shows the fitting results of the GEV distribution.

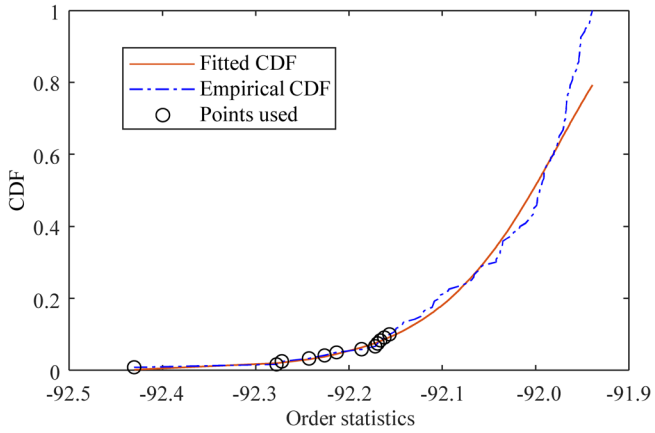

Fig. 9. The GEV distribution fitted on 10% of the *lnz* data.

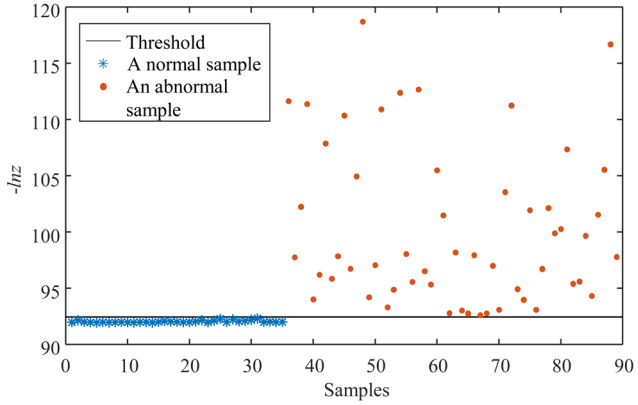

Fig. 10. Novelty detection results using the performance deviation model and EFT in which a sample corresponds to a performance deviation curve in 8 hours and the threshold is calculated by fitting the GEV.

With the optimized parameters, a threshold is calculated to be -92.33 with a confidence interval of 99%. Applying the testing samples to the GP model, the corresponding *lnz* values and the novelty detection results are shown in Fig. 10. For better visualization, *lnz* is plotted as *–lnz* in Fig. 10. and the negative value of the threshold is taken. As seen in Fig. 10, all the normal testing samples are below the threshold and identified as normal correctly while the abnormal testing samples all exceed the threshold as expected. There are neither false positives (FP), meaning 'normal' wrongly identified as 'abnormal', nor false negatives (FN), meaning 'abnormal' wrongly identified as 'normal'. 1000 runs of different sampling from the deviation curves are also constructed to investigate the effect of random sampling when creating the training, validation, and testing sets. The average classification accuracy of the testing set after 1000 runs is 99.68%.

4.3 Comparison with other methods

4.3.1 Comparison between BPNN and other networks

The BPNN is compared with other common-used networks including the one dimensional convolutional neural network (1D-CNN) and the long short-term memory (LSTM) network to verify the superiority of BPNN in constructing the performance deviation model of the GT. The CNN is a typical network with the 'deep' layer structure including combinations of convolutional layers and pooling layers [39]. The LSTM network is widely used in sequence learning for its ability to track the relationships and dependencies among time steps [40]. In the comparative experiment, the BPNN, the 1D-CNN and the LSTM network are implemented by Python using the same computer (Intel Core i5-7300HQ CPU @ 2.5 GHz and 8GB RAM). The same training, validation and testing data are used when training and testing the models.

For a fixed index of the training points, the comparisons of the prediction accuracy and computation complexity are shown in Table 2. The results in Table 2 are average prediction accuracy and average time for training after running the corresponding models 10 times. The optimized BPNN consists of an input layer, a hidden layer with 5 neurons and an output layer. The structure and parameters of the 1D-CNN and the LSTM network are optimized to obtain the similar prediction errors as the BPNN. The structure of the 1D-CNN is designed to be an input layer, 2 convolutional layers, a pooling layer, and a fully-connected layer (the output layer). The LSTM network is 3 layers including an input layer, a LSTM hidden layer with 10 neurons and an output layer. The prediction accuracy of the 1D-CNN and LSTM network could be higher with more layers or neurons but the time for training will correspondingly increase. In this paper, the BPNN is the most suitable method for building the performance deviation model of the GT as it could obtain comparable prediction accuracy using less time.

Table 2 Comparison results of different neural networks

Method	BPNN	1D-CNN	LSTM
Prediction accuracy (%)	99.85	99.84	99.86
Time for training (s)	28.05	42.91	91.69

4.3.2 Comparison between EFT and other threshold setting methods

Using the performance deviation model and EFT, the results shown in Fig. 10 seem to be convincing as data from a real GT could be well classified. A GP-based pointwise approach is compared with EFT in this paper. The same deviation curves are used here, and thresholds of deviation values are set based on GP. There are in total 37440 normal data points corresponding to 26 days and 25920 abnormal data points corresponding to 18 days. A GP model is trained based on the normal deviation values of one day. The output of the GP model displays a predictive distribution described by a mean prediction and covariance intervals. The upper and lower thresholds are set by the plus and minus three standard deviation values of the mean prediction. Fig. 11 shows the novelty detection results by the pointwise approach. Fig. 11(a) and Fig. 11(b) are examples of the novelty detection results in one day. Fig. 11(c) is the bar plot of the number of FP in a single day while the number of FN in one day is shown in Fig. 11(d). It can be seen that false alarms and missing alarms occur much more frequently compared with the EFT approach.

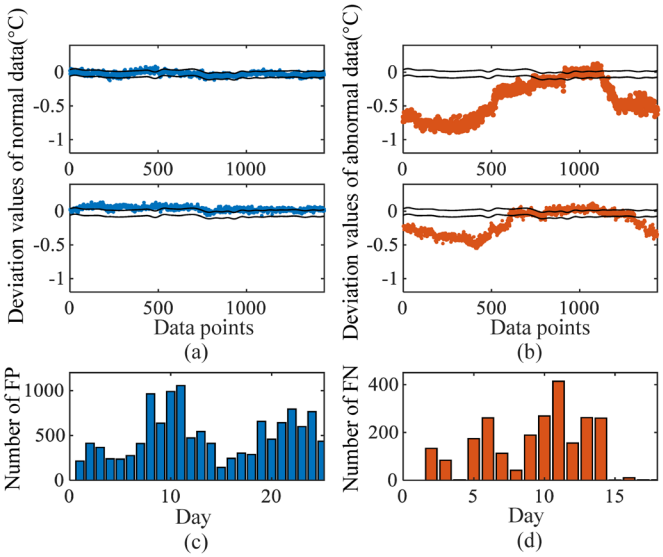

Fig. 11. Novelty detection with the pointwise GP approach. (a) Two examples of the outlier analysis of the one-day data points in normal status. (b) Two examples of the outlier analysis of the one-day data points with cracks in the system. (c) Bar plot of the number of FP each day. (d) Bar plot of the number of FN each day.

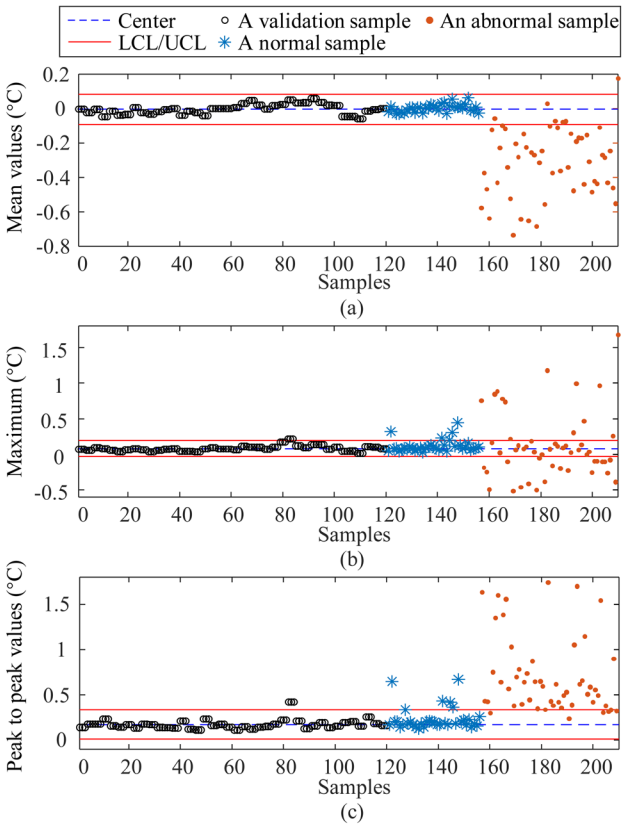

Fig. 12. Novelty detection based on thresholds set by time-domain features extracted every 8 hours. (a) Mean values. (b) Maximum values. (c) Peak and peak values.

Another comparison is made by extracting some statistic features in time domain instead of single data points for threshold setting. The same validation set, which includes data points in normal status, is used here to obtain the series of features and calculate the thresholds. The threshold is set as $\mu \pm 3\sigma$, where μ is the mean value of the feature series and σ is the standard deviation. The novelty detection results based on thresholds set by time-domain features is shown in Fig. 12 in the way of control charts. The charts have lower control limit (LCL) and upper control limit (UCL), which are corresponding to $\mu+3\sigma$ and $\mu-3\sigma$ respectively. The center line is corresponding to μ. The mean, maximum and peak to peak values of the deviation curves are chosen as the features respectively. Samples plotted outside the limits may trigger an alarm.

Different approaches are compared through the classification accuracy rate (CAR), false alarm rate (FAR), missing alarm rate (MAR), and sensitivity (S).

$$\mathrm{CAR} = (\mathrm{TP}+\mathrm{TN})/(\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}) \quad (19)$$

$$\mathrm{FAR} = \mathrm{FP}/(\mathrm{TN}+\mathrm{FP}) \quad (20)$$

$$\mathrm{MAR} = \mathrm{FN}/(\mathrm{TP}+\mathrm{FN}) \quad (21)$$

$$\mathrm{S} = \mathrm{TP}/(\mathrm{TP}+\mathrm{FN}) \quad (22)$$

where TP are true positives meaning abnormal samples correctly identified as 'abnormal' and TN are true negatives meaning the correctly identified normal samples. Based on the same performance deviation model and different threshold setting approaches, the novelty detection performance can be seen in Table 3. The values of CAR, FAR, MAR and S for different threshold setting methods in Table 3 are all average values based on 1000 times of running. It is clearly shown that the EFT has higher CAR than all the other approaches. Though the FAR of mean feature threshold approach is very low, many abnormal samples fail to trigger the alarm with the MAR of 7.41%. The EFT can effectively reduce the FAR without increasing the MAR and it is quite sensitive when detecting the novelty.

Table 3 Comparison of different threshold setting methods

	EFT	Pointwise GP	Mean feature	Maximum feature	Peak to peak feature
CAR (%)	99.68	79.56	95.50	75.28	88.76
FAR (%)	0.76	27.52	0	14.29	14.29
MAR (%)	0.03	10.61	7.41	31.48	9.26
S (%)	99.97	89.39	92.59	68.52	90.74

4.3.3 Comparison with one-class learning methods

The proposed novelty detection method is further compared with other typical one-class learning methods including local outlier factor (LOF)[41], one-class support vector machine (OCSVM)[42] and isolation forest (iForest)[43]. In the comparative experiment, the same real-time monitoring data are used. The one-class learning models are trained by the fixed 10000 normal data points and then tested with all the 26-day normal data as well as the 18 days of abnormal data. The 12 monitoring parameters of the GT are considered as input features of the models. The LOF, OCSVM and iForest are implemented by the scikit-learn library of Python. Radial basis function kernel is used for OCSVM. To decrease the stochastic nature of the training procedure, results of iForest are average outcomes by running the model 10 times.

Table 4 Comparison of different novelty detection methods

Method	CAR (%)	FAR (%)	MAR (%)	S (%)
LOF	87.69	2.13	27.03	72.97
OCSVM	83.80	0.89	38.31	61.69
iForest	98.79	1.88	0.24	99.76
Proposed method	99.68	0.76	0.03	99.97

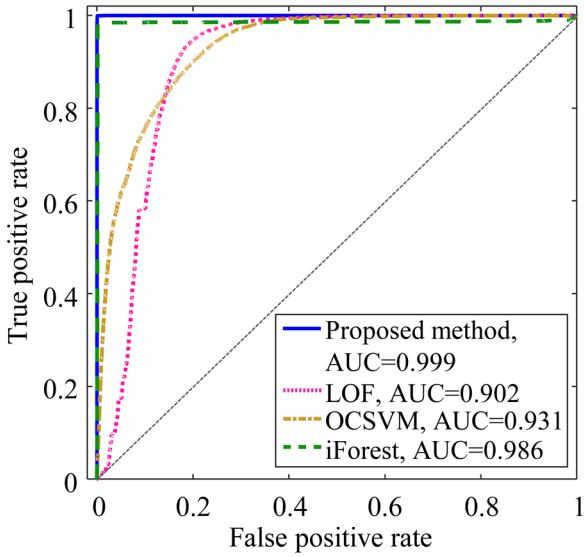

Fig. 13 ROC curve and AUC of different novelty detection methods

The comparison results are listed in Table 4. The LOF classifier performs relatively poorer in this case maybe because the dimension of the features is not small. The FAR of OCSVM is low, which means that 99.11% of the normal patterns are correctly classified. However, OCSVM is less sensitive to the abnormality. The reason may be that for most of the features used in this paper, the separabilities between the normal and abnormal conditions were poor. When selecting the parameters that are well-separated between classes, for example, vibration of the rotor on the compressor side, the sensitivity of OCSVM to detect the abnormality could rise up to almost 100%. However, the prior knowledge of the between-class separability is not always obtainable when there are no abnormal data available for training. The FAR of iForest is unstable and not low enough. It may be because the change of the ambient

conditions and the operation conditions could mislead the model to classify the normal data as a fault. The proposed method obtains the best detection performance for the real-life GT monitoring data among these methods. It realizes novelty detection using only normal data. The performance deviation model by BPNN fuses different monitoring parameters based on the basic principles of the GT. The threshold set by EFT based on the deviation curves shows good balance between the false alarms and missing alarms.

The superiorities of the proposed method are also proved by the receiver operating characteristic (ROC) curve and area under the curve (AUC) displayed in Fig. 13. The ROC curve is obtained by the false positive rates (FPr) and the corresponding true positive rates (TPr) calculated for all the possible threshold values, where FPr=FAR=FP/(FP+TN), TPr=TP/(TP+FN) [4]. The curve of the proposed method is closest to the left and the top border of the ROC space with the highest AUC, which implies that the classifier is the most accurate one.

5. Conclusion

This paper focuses on the novelty detection approach of an industrial GT using performance deviation model and extreme function theory. The aim is to detect the abnormality timely and precisely, using the real-life monitoring parameters of GTs. A case study of a Mitsubishi single shaft GT M251S is carried out, where the GT suffers a fault during the operation but fails to trigger the current alarm. A performance deviation model of normality is established to fuse the multi-type monitoring parameters and emphasize the differences of parameters between the normal and abnormal status. Deviation curves, implying the status of the GT, are obtained based on BPNN. The normality of the deviation curves is tested as functions instead of the pointwise approach. A GP model is trained using the normal deviation curves. The threshold representing the extremity of the functions compared to the training functions is then calculated based on EFT. The average novelty detection accuracy of the proposed method is 99.68%.

Several comparative experiments are carried out. The BPNN is compared with other neural networks to verify that it can obtain comparable prediction accuracy using less time. The EFT is compared with the pointwise approach, which uses GP for modeling and threshold setting. Comparisons are also made between EFT and feature based approach, for which the threshold is set based on the 3σ rule. The EFT is clearly accurate with much lower rate of false alarms (0.76%) and missing alarms (0.03%). The proposed approach is further compared with common-used one-class leaning methods. The proposed one shows quite high sensitivity to detect the abnormal samples with good balance between the false alarms and missing alarms. This paper shows a successful case of novelty detection on real-time data from the GT. The use of performance deviation model and EFT could help the health monitoring of the industrial GTs with high detection accuracy and sensitivity.

Future work will focus on the online threshold update and optimization with real-time normal data of GT from various working conditions. It is also interesting to develop real-time prototype based on a programmable digital device in application field.

Acknowledgements

The authors acknowledge the support from the National Natural Science Foundation of China (Grant No. U1809219, No. 51705302) and the Key Research and Development Project of Zhejiang Province (Grant No. 2020C01088).

References

[1] M. Tahan, E. Tsoutsanis, M. Muhammad, and Z. A. A. Karim, Performance-based health monitoring, diagnostics and prognostics for condition-based maintenance of gas turbines: A review, Appl. Energy, 198 (2017) 122-144. https://doi.org/10.1016/j.apenergy.2017.04.048.

[2] H. Hanachi, C. Mechefske, J. Liu, A. Banerjee, and Y. Chen, Performance-based gas turbine health monitoring, diagnostics, and prognostics: A Survey, IEEE Trans. Reliab., 67 (2018) 1340-1363. https://doi.org/10.1109/TR.2018.2822702.

[3] C. R. Farrar, and K. Worden, Structural health monitoring: A machine learning perspective, 3rd ed., John Wiley & Sons, UK, 2013.

[4] X. M. Ding, Y. H. Li, A. Belatreche, L. P. Maguirea, An experimental evaluation of novelty detection methods, Neurocomputing, 135 (2014) 313-327. https://doi.org/10.1016/j.neucom.2013.12.002.

[5] A. D. Kenyon, V. M. Catterson, S. D. J. McArthur, and J. Twiddle, An agent-based implementation of hidden markov models for gas turbine condition monitoring, IEEE Trans. Syst. Man Cybern., 44 (2014) 186-195. https://doi.org/10.1109/TSMC.2013.2251539.

[6] Z. Li, S. S. Zhong, and L. Lin, Novel Gas Turbine Fault Diagnosis Method Based on Performance Deviation Model, J. Propul. Power, 33 (2017) 730-739. https://doi.org/10.2514/1.B36267.

[7] J. He, S. X. Yang, E. Papatheou, X. Xiong, H. B. Wan and X. W. Gu, Investigation of a multi-sensor data fusion technique for the fault diagnosis of gearboxes, Proc. Inst. Mech. Eng. Part C-J. Eng. Mech. Eng. Sci., 233 (2019) 4764-4775. https://doi.org/10.1177/0954406219834048.

[8] L.P. Yan, X. Z. Dong, T. Wang, Q. Gao, C. Q. Tan, D. T. Zeng et al., A fault diagnosis method for gas turbines based on improved data preprocessing and an optimization deep belief network, Meas. Sci. Technol., 31 (2020) 015015. https://doi.org/10.1088/1361-6501/ab3862.

[9] J. Liu, Y. F. Li, E. Zio, A SVM framework for fault detection of the braking system in a high speed train, Mech. Syst. Signal Proc., 87 (2017) 401-409. https://doi.org/10.1016/j.ymssp.2016.10.034.

[10] S.M. Erfani, S. Rajasegarar, S. Karunasekera, C. Leckie, High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning, Pattern Recogni., 58 (2016) 121-134. https://doi.org/10.1016/j.patcog.2016.03.028.

[11] R. B. de Santis, M. A. Costa, Extended Isolation Forests for fault detection in small hydroelectric plants, Sustainability, 12 (2020) 6421. https://doi.org/10.3390/su12166421.

[12] M. A. F. Pimentel, D. A. Clifton, L. Clifton, and L. Tarassenko, A review of novelty detection, Signal Process., 99 (2014) 215-245. https://doi.org/10.1016/j.sigpro.2013.12.026.

[13] C. W. Gu, H. Wang, X. X. Ji, and X. S. Li, Development and application of a thermodynamic-cycle performance analysis method of a three-shaft gas turbine, Energy, 112 (2016) 307-321. https://doi.org/10.1016/j.energy.2016.06.094.

[14] A. Mehrpanahi, G. Payganeh, and M. Arbabtafti, Dynamic modeling of an industrial gas turbine in loading and unloading conditions using a gray box method, Energy, 120 (2017) 1012-1024. https://doi.org/10.1016/j.energy.2016.12.012.

[15] H. Asgari, M. Venturini, X. Q. Chen, and R. Sainudiin, Modeling and simulation of the transient behavior of an industrial power plant gas turbine, J. Eng. Gas. Turbines Power-Trans. ASME, 136 (2014) 061601. https://doi.org/10.1115/1.4026215.

[16] S. M. Jones, Steady-state modeling of gas turbine engines using the numerical propulsion system simulation code, In Proceedings of the ASME Turbo Expo: Power for Land, Sea and Air, 1 (2010) 89–116. https://doi.org/10.1115/GT2010-22350.

[17] V. Panov, GasturboLib: Simulink library for gas turbine engine modelling, in Proceedings of the ASME Turbo Expo: Power for Land, Sea and Air, 1 (2009) 555–565. https://doi.org/10.1115/GT2009-59389.

[18] J. W. Chapman, T. M. Lavelle, R. May, J. S. Litt, and T. H. Guo, Propulsion System Simulation Using the Toolbox for the Modeling and Analysis of Thermodynamic Systems (T MATS), in 50th AIAA Joint Propulsion Conference, (2014). https://doi.org/10.2514/6.2014-3929.

[19] D. F. Amare, T. B. Aklilu and S. I. Gilani, Gas path fault diagnostics using a hybrid intelligent method for industrial gas turbine engines, J. Braz. Soc. Mech. Sci. Eng., 40 (2018) 548. https://doi.org/10.1007/s40430-018-1497-6.

[20] H. Asgari, X. Q. Chen, M. Morini, M. Pinelli, R. Sainudiin, P. R. Spina, and M. Venturini, NARX models for simulation of the start-up operation of a single-shaft gas turbine, Appl. Therm. Eng., 93 (2016) 368-376. https://doi.org/10.1016/j.applthermaleng.2015.09.074.

[21] J. Enríquez-Zaratea, L. Trujillo, S. de Lara, M. Castelli, E. Z-Flores, L. Munoz, and A. Popovic, Automatic modeling of a gas turbine using genetic programming: An experimental study, Appl. Soft. Comput., 50 (2017) 212-222. https://doi.org/10.1016/j.asoc.2016.11.019.

[22] S. S. Zhong, S. Fu, and L. Lin, A novel gas turbine fault diagnosis method based on transfer learning with CNN, Measurement, 137 (2019) 435-453. https://doi.org/10.1016/j.measurement.2019.01.022.

[23] Z. M. Liu, and I. A. Karimi, Gas turbine performance prediction via machine learning, Energy, 192 (2020) 116627. https://doi.org/10.1016/j.energy.2019.116627.

[24] A. D. Fentaye, S. I. U. Gilani, A. T. Baheta, and Y. G. Li, Performance-based fault diagnosis of a gas turbine engine using an integrated support vector machine and artificial neural network method, Proc. Inst. Mech. Eng. Part A-J. Power Energy, 233 (2018) 786-802. https://doi.org/10.1177/0957650918812510.

[25] A. D. Fentaye, A. T. Baheta, S. I. Gilani, and K. G. Kyprianidis, A review on gas turbine gas-path diagnostics: State-of-the-art methods, challenges and opportunities, Aerospace, 6 (2019) 83. https://doi.org/10.3390/aerospace6070083.

[26] M. Bai, J. Liu, J. Chai, X. Zhao, D. Yu, Anomaly detection of gas turbines based on normal pattern extraction, Appl. Therm. Eng., 166 (2020) 114664. https://doi.org/10.1016/j.applthermaleng.2019.114664.

[27] S. Amirkhani, A. Chaibakhsh, and A. Ghaffari. Nonlinear robust fault diagnosis of power plant gas turbine using Monte Carlo-based adaptive threshold approach, ISA Trans., 100 (2020) 171-184. https://doi.org/10.1016/j.isatra.2019.11.035.

[28] D. Toshkova, M. Asher, P. Hutchinson, and N. Lieven, Automatic alarm setup using extreme value theory, Mech. Syst. Signal Proc., 139 (2019) 106417. https://doi.org/10.1016/j.ymssp.2019.106417.

[29] E. Papatheou, N. Dervilis, A. E. Maguire, C. Campos, L. Antoniadou, and K. Worden, Performance monitoring of a wind turbine using extreme function theory, Renew. Energy, 113 (2017) 1490-1502. https://doi.org/10.1016/j.renene.2017.07.013.

[30] D. A. Clifton, L. Clifton, S. Hugueny, D. Wong, and L. Tarassenko. An extreme function theory for novelty detection, IEEE J. Sel. Top. Signal Process., 7 (2013) 28-37. https://doi.org/10.1109/JSTSP.2012.2234081.

[31] J. He, S. X. Yang, and C. B. Gan, Unsupervised fault diagnosis of a gear transmission chain using a deep belief network, Sensors, 17 (2017) 1564-1564. https://doi.org/10.3390/s17071564.

[32] S. Khan, T. Yairi, A review on the application of deep learning in system health management, Mech. Syst. Signal Proc., 107 (2018) 241–265. https://doi.org/10.1016/j.ymssp.2017.11.024.

[33] H. Nikpey, M. Assadi, P. Breuhaus, Development of an optimized artificial neural network model for combined heat and power micro gas turbines, Appl. Energy, 108 (2013) 137–148. http://dx.doi.org/10.1016/j.apenergy.2013.03.016.

[34] R. Bettocchi, M. Pinelli, P. R. Spina, M. Venturini, Artificial Intelligence for the Diagnostics of Gas Turbines—Part I: Neural Network Approach, in 50th ASME Turbo-Expo, (2005). https://doi.org/10.1115/1.2431391.

[35] C. E. Rasmussen and C. K. I. Williams, Gaussian Processes for Machine Learning, The MIT Press, USA, 2006.

[36] E. Schulz, M. Speekenbrink, and A. Krause, A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions, J. Math. Psychol., 85 (2018) 1-16. https://doi.org/10.1016/j.jmp.2018.03.001.

[37] J. Park, M. Hamadache, J. M. Ha, Y. Kim, K. Na, and B. D. Youn, A positive energy residual (PER) based planetary

gear fault detection method under variable speed conditions, Mech. Syst. Signal Proc., 117 (2019) 347-360. https://doi.org/10.1016/j.ymssp.2018.08.010.

[38] R. A. Fisher and L. H. C. Tippett, Limiting forms of the frequency distribution of the largest or smallest member of a sample, Math. Proc. Camb. Philos. Soc., 24 (1928) 180–190. https://doi.org/10.1017/S0305004100015681.

[39] L. Jing, M. Zhao, P. Li, X. Xu, A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox, Measurement, 111 (2017) 1-10. https://doi.org/10.1016/j.measurement.2017.07.017.

[40] D. Wu, Z. Jiang, X. Xie, X. Wei, W. Yu, R. Li, LSTM learning with Bayesian and Gaussian Processing for anomaly detection in industrial IoT, IEEE Trans. Industr. Inform., 16 (2020) 5244-5253. https://doi.org/10.1109/TII.2019.2952917.

[41] M. M. Breunig, H. P. Kriegel, R. T. Ng, J. Sander, LOF: identifying density-based local outliers, in Acm Sigmod International Conference on Management of Data, 29 (2000) 93-104. https://doi.org/10.1145/342009.335388.

[42] F. T. Liu, K. M. Ting, Z. H. Zhou, Isolation forest, in 8th IEEE International Conference on Data Mining, 2008, pp. 413-422. https://doi.org/10.1109/ICDM.2008.17.

[43] J. Ma, S. Perkins, Time-series novelty detection using one-class support vector machines, in Proceedings of the International Joint Conference on Neural Networks, 3 (2003) 1741-1745. https://doi.org/10.1109/IJCNN.2003.1223670.