Artificial Intelligence Techniques and Cloud Computing for Wind Turbine Pitch Bearing Fault Detection

Artificial Intelligence Techniques and Cloud Computing for Wind Turbine Pitch Bearing Fault Detection

Chao Zhang Fausto Pedro Garcia Marquez Long Zhang

***Abstract*—Blade bearings are critical rotating units for maximizing wind power yield. It is essential to detect the blade bearing faults at an early stage and prevent their catastrophic failure. One major challenge lies in the signal denoising under the time-varying operating conditions. This time-varying condition is often treated as a series of piece-wise time invariant conditions when filtering the collected signal. The duration of the time-invariant period, also referred to as the window length or comprehensive period, is often determined by trial-and-error, which could lead to improper separating the time varying signals and poor fault detection performance. In this paper, to find a suitable window length, a novel method called the Temporal Convolutional Augmented Bayesian Search (TCABS) algorithm is used to search for a 'comprehensive period' for the unknown signal. After estimating the window length, the Split Bayesian Augmented Lagrangian Algorithm (SBAL) was used based on split window techniques to construct time-varying models. The proposed TCABS and SBAL are validated by real signals collected from an industrial-scale wind turbine in operation for over 15 years.**

***Index Terms*—search algorithm, neural networks, system identification, cloud computing, blade bearing fault diagnosis, sequence data, non-destructive test**

I. Introduction

Wind power, as a major type of renewable energy, is a crucial part of modern energy mix to allow the public to construct a carbon emission-free society [1]. According to global statistics data on wind energy [2], the global total cumulative wind power capacity was only 94 GW in 2007, rising to 486.8 GW in 2016 and over 800 GW in 2021.

Wind turbines that extract wind power from natural wind flows, often work in an extreme environment that may be awash with various corrosion factors. Long-term exposure in harsh circumstances and mechanical wear damage wind turbine components and deteriorate their performance and shorten their lifespans.

The wind turbine blade bearing, a crucial but vulnerable inner component of wind turbines, is one of the most likely components to cause reliability problems. The small failures in blade bearings may result in poor pitching and aerodynamic of a wind turbine, so that the energy conversion efficiency may decrease, and they may also cause safety accidents when blades entirely lose control [3].

Human inspection costs are high [4], and slight internal damage may be difficult to detect, so it is necessary to design a condition monitoring and fault diagnosis (CMFD) method for wind turbine blade bearing to improve reliability and reduce economic losses and safety risks.

Acoustic emission (AE) sensors are chosen to collect data in this paper for their high sensitivity in the frequency range [5], after multidimensional considerations and comparisons among commonly-used sensors such as optical sensors and vibration sensors.

After the signal is collected using AE sensors, the raw noisy signal has to be filtered. There are very limited work of denoising signals on wind turbine blade bearing fault detection. This is because large-scale application of wind turbines happened 15 years ago, and damage in wind turbines often happens after 10 years. In general, it is a problem that was recently exposed to the industry and had little prior research.

As there are limited work on blade bearing CMFD, some typical work on CMFD used in other fields are briefly reviewed here. Self-adaptive noise cancellation method called discrete/random separation (DRS) [6], [7] can reduce the linear noise effects in raw signals by editing the cestrum of raw signals. Auto-regressive model [8] with reinforcement learning can well approximate the fault signal components by self-learning and evolution. Artificial neural network LSTM [9] can judge the healthy condition and predict the durability, namely remaining useful life. However, the aforementioned CMFD methods may not be able to be directly used for wind turbine blade bearing CMFD. For blade bearing CMFD, we need consider two challenges:

- The pitch bearing is operated under non-stationary condition due to unbalance wind load and start-stop. This time varying condition is often treated as a number of piecewise time invariant conditions with different windows lengths. In this case, the window lengths have to be determined properly to make sure the accurate separation of the signal.
- Two major denoising methods presently include linear filters and nonlinear filters. Linear filters often enjoy simple forms and fast computations but may perform poor for signal with strong nonlinear components. Nonlinear filters can deal with nonliear signal but may have more complex

This work was supported by the Engineering and Physical Sciences Research Council under Grant EP/S017224/1. (Corresponding author: Long Zhang)

Chao Zhang and Long Zhang are with the Department of Electrical and Electronic Engineering, The University of Manchester, Manchester M13 9PL, U.K. (e-mail: chao.zhang-14@student.manchester.ac.uk; long.zhang@manchester.ac.uk).

Fausto Pedro Garcia Marquez is with the ETSI Industrial, University of Castilla-La Mancha, Ciudad Real 13071, Spain, (e-mail: FaustoPedro.Garcia@uclm.es).

models. Proper model parameter estimation methods may be chosen to provide a good trade-off between model complexity and filtering performance.

In terms of the time-varying problem, Antoni [10] utilized fast kurtogram to lock corresponding frequency band so that the time-varying problem can be evaded. Liu [11] utilized a type of augmented lagrangian method for noise model construction, simplifying it as approximately time-invarying systems. These aforementioned methods may not solve the problems of time-varying conditions well. Regarding window length, Chen et al. [12] optimized the window size to predict protein structure based on a probabilistic approach. Ding et al. [13] used the optimal window length to detect anomalies for streaming data, considering the concept drift phenomenon. Shynkevich et al. [14] studied the impact of changing input window lengths on forecasting varied horizons, looking at the combination of a forecast horizon and input window length. Although they all showed the importance of appropriate window length, their determination process still depends on trial-and-error methods that are not universally applicable.

Most existing denoising methods can be classified as linear or nonlinear filters. discrete/random separation (DRS)is one typical linear filter. Liu et al. [15] used iterative nonlinear filter (INF) to reduce nonlinear noise and extract weak fault signals. Li et al. [16] applied the technique of bind vibration component separation to extract nonlinear mechanical fault features from various rotating components. In addition to the linear and nonlinear filter models, the model parameter optimization also effect the filtering performance. Huang et al. [17] proposed a novel multi-source sparse representation to realize fault diagnosis without prior knowledge. Liao et al. [18] improved a generalized infimal convolution smoothing (GICS) to enhance the estimation accuracy of sparse matrices. Sparse representation techniques will also be applied into our work.

Drawing on these concerns, this paper used the split window techinique to treat the time-varying condition with piecewise period. For the time-invarying system, the period is fixed, and we analyze and deal with it according to the fixed period. For a time-varying system, the period is changing, and we need to break the signal up into segments according to the period of change and then process it. It is difficult for us to do that, but we can find an approximate period (comprehensive period), which is a good representation of this little variable period. Then, the optimal window length contributing to excellent signal processing performance is always related to the comprehensive period in all window-based methods. To find appropriate piecewise time, referred to as winodw length, a novel method called Temporal Convolutional Augmented Bayesian Search (TCABS) was used. To provide a good trade-off between model complexity and filtering performance, a model parameter estimation methods named Split Bayesian Augmented Lagrangian Algorithm (SBAL) was employed to obtain sparse coefficients during the process of time-varying model construction.

It is worth mentioning that this TCABS method considers inner characteristics of an unknown signal for constructing a fault diagnosis system. The determination of these hyper-parameters may be related to inner characteristics of signals. If not consider inner characteristics of dynamic systems, this direct relationship of the black-box model between input and output may miss the nature of targeted objects, so understanding inner characteristics is crucial, especially for determining hyper-parameters here. Based on this idea, we proposed the parallel search algorithm TCABS, to obtain a 'comprehensive period' related inner characteristics of a signal. This 'comprehensive period' can be significant reference of window length so that a fault diagnosis system can be constructed without many trials.

For this time-varying signals, the window-based solutions are often applied to contrapose them in a segment-by-segment format. However, different window length will severely influence the final performance of signal processing. Therefore, our proposed TCABS, which can find a suitable window length, is essential to better signal processing.

In essence, the major work of this paper can be concluded as follows:

1) A search algorithm TCABS algorithm is proposed to determine hyper-parameter window length. TCN network structure is used as feature extraction of TCABS to obtain signal inner characteristics. Furthermore, a new acquisition function augmented expected improvement (AEI) is used to accelerate the convergence of the TCABS.

2) The Split Bayesian Augmented Lagrangian Algorithm (SBAL) was employed to estimate sparse coefficients of the nonlinear filter models. All proposed methods in this paper are validated with an industrial-scale and naturally damaged wind turbine which has been operated in a real wind farm for over 15 years.

II. Temporal Convolutional Augmented Bayesian Search (TCABS)

A. Research Subject – Sequence Data

Before clarification of our algorithm, variables have to be defined first. Regarding any unknown signal, it should be continuous and uninterrupted, but data acquired by sensors will be transformed into a discrete form. A series of discrete sequence pair from a sensor can be defined: $(x_0, y_0), ...(x_{n-1}, y_{n-1})$. The $x_0, ..., x_{n-1}$ is the time sequence, and the $y_0, .., y_{n-1}$ is the signal amplitude corresponding to the time sequence [19].

The purpose of our search algorithm is to calculate the comrephensive period for unknown signals, which can be referred to optimize subsequent sequence models. The input of our search structure is discrete sequence pair $(x_0, y_0), ...(x_{n-1}, y_{n-1})$, and the output is calculated comprehensive period. Formally, mapping relation of a search structure is function $\mathbf{s}$ to output comprehensive period T:

$$T = \mathbf{s}((x_0, y_0), ...(x_{n-1}, y_{n-1})) \tag{1}$$

Here discrete sequence pair $(x_0, y_0), ...(x_{n-1}, y_{n-1})$ is the collected signal that is definite, merely slightly influenced by sampling rate. The goal of our search algorithm is to calculate the best comprehensive period T_b during multiple iterations, where the discrete sequence pair should be related to some distribution.

B. Feature Extraction Module

The purpose of this search algorithm is to conveniently calculate comprehensive period, so that the hyper-parameter window length can be determined better to optimize subsequent sequence models. The computing time is crucial to a search algorithm. To save computing time, our feature extraction module is realized by a type of parallel method that refers to temporal convolutional networks (TCN) .

B1. Causal Convolutions

The causal convolution is a type of structure that refers to the fully-convolutional network (FCN) [20] to extract feature information from sequence data. The thought of causal convolutions is that compressed information obtained from convolution operation in low layer can be delivered into high layer, layer by layer. The content in higher layer is more comprehensive and concrete until the predicted value is output in top layer. The propagation process is shown in Fig. 1 where P_L is the window length needed to be adjusted.

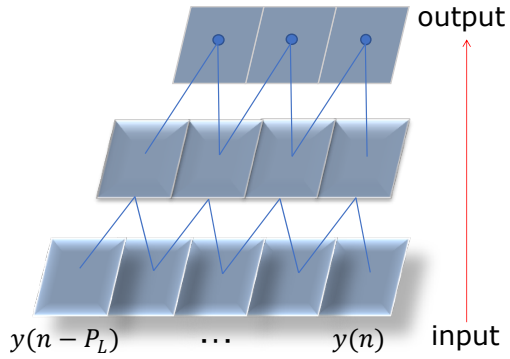

Fig. 1. Causal Convolution Architecture.

B2. Dilated Convolutions

As mentioned above, the causal convolutions are utilized to extract feature from sequence data that is delivered in form of progressive layer, but it will result in a very deep networks. Following the work [21], the dilated convolution can be employed to reduce depth of neural networks by enlarging receptive field. As can be seen in Fig. 2, the dilated convolution can be carried out with way of interval operation (here, convolution size $k = 2$ and interval $d = 2$). This dilated convolution can refer to longer history than regular causal convolution so that the number of convolution operations is few when input data length is definite, and these few operations indirectly reduce the required layers for a neural network. It is worth noting that, When interval $d = 1$, a dilated convolution reduces to a regular convolution.

B3. Residual Blocks

Although the dilated convolutions are applied, our neural network is still deep, resulting in significant training loss. A residual connection [22] proposed for image processing can also be applied for sequence data processing in order to avoid

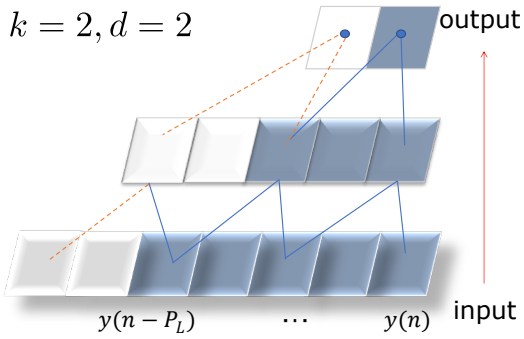

Fig. 2. Dilated Convolution Architecture.

performance degradation from deep networks. Compare with the normal output $\mathbf{y} = \mathcal{F}(\mathbf{x})$, the new output using a residual connection is defined as follows:

$$\mathbf{y} = \mathbf{x} + \mathcal{F}(\mathbf{x}) \tag{2}$$

Here $\mathbf{x}$ and $\mathbf{y}$ are the input and output of current layer. Observing (2), the residual connection is a flexible and pluggable type of connection block. When a deep layer is redundant, the (2) in this layer will become $\mathbf{y} = \mathbf{x}$, namely, identity mapping. This way has been repeatedly proved to benefit very deep networks because it is easy to obtain this kind of identity mapping when some layers are redundant. The Fig. 3 (convolution size $k = 3$, interval $d = 1$) describes this type of residual connection, the middle layer will be skipped when the networks find the identity mapping resulting in better performance during training process.

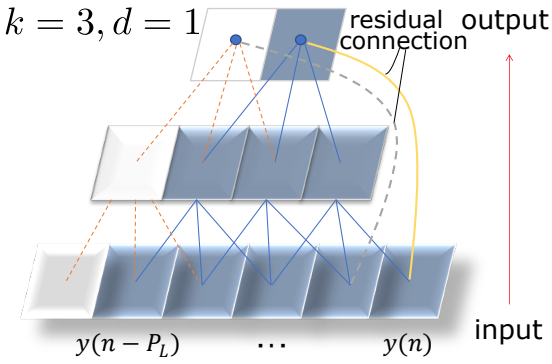

Fig. 3. Residual Block.

C. Search Module

The search module here is divided into initialization and exploration, referring to the Bayesian Optimization. Initialization can accelerate the convergency of search algorithm. Exploration can determine the search strategy of searching process.

C1. Initialization

In order to obtain an efficient search algorithm, we take a trick about the initial points to accelerate the searching process. We focus on the initial points because our proposed

search algorithm is a Bayesian-based structure that is severely influenced by initialization.

Regarding a continunous system, the fourier transform [23], [24] is defined:

$$\mathcal{F}[y(t)] = Y(w) = \int_{-\infty}^{\infty} y(t)e^{-iwt}dt \tag{3}$$

if we replace the w with $2\pi f$:

$$\mathcal{F}[y(t)] = Y(w) = \int_{-\infty}^{\infty} y(t)e^{-i2\pi ft}dt \tag{4}$$

Then, here we merely consider fixed sampling frequency, not variant sampling frequency, so the first assumption is that our collected signals possess the characteristics of the equal time interval. With regard to N data pairs $((x_0, y_0), ...(x_{n-1}, y_{n-1}))$, the discrete fourier transform (DFT) is obtained:

$$\mathcal{F}[y[n]] = Y(e^{iw}) = Y[k] = \sum_{n=0}^{N-1} y[n]e^{-i\frac{2\pi}{N}nk}(k = 0, ..., N-1) \tag{5}$$

The specific DFT can be calculated by some fast fourier transform algorithm (FFT), such as Cooley-Tuckey, which has been repeatedly proved efficient in many studies. The computational complexity in original DFT is $O(N^2)$, and that in Cooley-Tuckey is just $O(Nlog_2N)$.

Subsequently, we could transform the signal from time domain into frequency domain, using DFT methods. The frequency-domain signal from DFT is with interval $\Delta f = \frac{1}{N\Delta t}$. We can obtain that frequency axis are the folloing frequencies:

$$\begin{aligned} f = \frac{\omega}{2\pi} =& 0, \Delta f, 2\Delta f, \ldots, (N/2-1)\Delta f \\ & \pm (N/2)\Delta f, -(N/2-1)\Delta f, \ldots, -\Delta f \end{aligned} \tag{6}$$

It is noting that the front half is positive frequencies and the back half is negative frequencies, and here we only consider the positive frequencies.

After that, the m frequencies corresponding to top m large amplitudes in frequency domain will be selected into a set A:

$$A = \{a_1, ..., a_m\} \tag{7}$$

Note that this m should be selected according to the total number of valuable frequencies that have filtered the noisy frequencies. The noisy frequencies with tiny amplitudes are from noise that should be eliminated from our signals. Empirically, those frequencies with amplitudes smaller than one-tenth of median should be eliminated as noisy frequencies, and one-tenth of the total number of valuable frequencies can be determined as the m of A.

Afterward, We need to calculate the lowest common multiple for the maximum and the minimum element in set A and the reciprocal value $\mathbf{P}$ is the initial point, as shown in (8).

$$\mathbf{P} = \frac{1}{[A_{max}, A_{min}]} \tag{8}$$

C2. Exploration

Regarding the exploration, we refer to the Bayesian Optimization [25], which consists of two parts: probabilistic surrogate model and acquisition function. The probabilistic surrogate model here is selected as Gaussian Process (GP) due to its universality for nonlinearity and linearity. The acquisition function will use a trick to make exploration care more about our initial points that are exquisitely prepared. Four principles about exploration should be noticed:

- We should make the most of each exploration, including the initial point.
- The value of acquisition function (utility function) is small near the existing sample points because those points have already been explored, and then calculating the value of the function at those points is unhelpful to find the globally optimal solution.
- The value of the acquisition function is large at points with wide confidence intervals because these points are uncertain and worth exploring.
- The value of the function is larger at points where the mean of a surrogate model is larger because the mean is an estimate of the value from the surrogate model at that point, and these points are more likely to be near the extreme point.

It is assumed that $\mathbf{x}$ is the window length that needs to be searched and a GP is selected as surrogate model, as follows:

$$f(\mathbf{x}) \sim \mathcal{GP}\left(m(\mathbf{x}), k\left(\mathbf{x}, \mathbf{x}\right)\right) \tag{9}$$

where we can find that a GP is completely determined by mean function m and covariance function k. To be convenient, we simplify here that the mean value function m is zero, namely $m(\mathbf{x}) = 0$. Some other selection of priors for the mean can be referred to [26], [27]. The covariance function is the squared exponential function:

$$k\left(\mathbf{x}_i, \mathbf{x}_j\right) = \mathbf{exp}(-\frac{1}{2}\|\mathbf{x}_i - \mathbf{x}_j\|^2) \tag{10}$$

According to those definite points explored, the new surrogate model (GP) can be updated.

we define the prior set $\mathcal{D}_{1:t} = \{\mathbf{x}_{1:t}, f_{1:t}\}$ that consists of a series of exploration points($\mathbf{x}_{1:t}$ corresponds to x axis and $f_{1:t}$ corresponds to y axis)

Each point at function f can be described by normal distribution $\mathcal{N}(0, K)$

$$\mathbf{K} = \begin{bmatrix} k\left(\mathbf{x}_1, \mathbf{x}_1\right) & \ldots & k\left(\mathbf{x}_1, \mathbf{x}_t\right) \\ \vdots & \ddots & \vdots \\ k\left(\mathbf{x}_t, \mathbf{x}_1\right) & \ldots & k\left(\mathbf{x}_t, \mathbf{x}_t\right) \end{bmatrix} \tag{11}$$

By utilizing the properties of GP, $f_{1:t}$ and f_{t+1} observe joint Gaussian:

$$\begin{bmatrix} f_{1:t} \\ f_{t+1} \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \mathbf{K} & \mathbf{k} \\ \mathbf{k}^T & k\left(\mathbf{x}_{t+1}, \mathbf{x}_{t+1}\right) \end{bmatrix}\right) \tag{12}$$

where

$$\mathbf{k} = \begin{bmatrix} k\left(\mathbf{x}_{t+1}, \mathbf{x}_1\right) & k\left(\mathbf{x}_{t+1}, \mathbf{x}_2\right) & \cdots & k\left(\mathbf{x}_{t+1}, \mathbf{x}_t\right) \end{bmatrix} \tag{13}$$

Subsequently, a predictive distribution could be obtained through Sherman-Morrison-Woodbury formula [28]:

$$P\left(f_{t+1} \mid \mathcal{D}_{1:t}, \mathbf{x}_{t+1}\right)=\mathcal{N}\left(\mu_t\left(\mathbf{x}_{t+1}\right), \sigma_t^2\left(\mathbf{x}_{t+1}\right)\right) \tag{14}$$

$$\begin{aligned} \mu_t\left(\mathbf{x}_{t+1}\right) &= \mathbf{k}^T \mathbf{K}^{-1} f_{1:t} \\ \sigma_t^2\left(\mathbf{x}_{t+1}\right) &= k\left(\mathbf{x}_{t+1}, \mathbf{x}_{t+1}\right)-\mathbf{k}^T \mathbf{K}^{-1} \mathbf{k} \end{aligned} \tag{15}$$

The update process for next point $\mathbf{x}_{t+1}$ in GP is presented as from (11) to (15). However, we still need to make sense how to select the next point that is decided by acquisition function.

With regards to the exploration strategy, three types of acquisition functions are proved useful in many researches. They are upper confidence bound (UCB), probability of improvement (PI) and expected improvement (EI).

- UCB:

$$\mathrm{UCB}(\mathbf{x})=\mu(\mathbf{x})+\kappa \sigma(\mathbf{x}) \tag{16}$$

where $\kappa > 0$, a weight parameter.

- PI:

$$\mathrm{PI}(\mathbf{x})=\Phi\left(\frac{\mu(\mathbf{x})-f\left(\mathbf{x}^{+}\right)-\kappa}{\sigma(\mathbf{x})}\right) \tag{17}$$

where $\mathbf{x}^{+}=argmax f(\mathbf{x}_i)$. Φ is CDF of the standard normal distribution.

- EI:

$$\mathrm{EI}(\mathbf{x})= \begin{cases}\left(\mu(\mathbf{x})-f\left(\mathbf{x}^{+}\right)-\kappa\right) \Phi(Z)+\sigma(\mathbf{x}) \phi(Z) & \text{if } \sigma(\mathbf{x})>0 \\ 0 & \text{if } \sigma(\mathbf{x})=0\end{cases} \tag{18}$$

where $Z=(\mu(\mathbf{x})-f(\mathbf{x}^{+}))/\sigma(\mathbf{x})$. ϕ and Φ denote the PDF and CDF of the standard normal distribution, respectively.

The three acquisition functions have have their own merit and drawback:

- UCB: easy to find global search but convergance is slow.
- PI: easy to be trapped in locally optimal solution.
- EI: trade-off between global search and local optimization.

At last, we give priority to the EI due to its excellent comprehensive performance. In addition, to make full use of the initial point, we proposed EI-variant acquisition function–augmented expected improvement (AEI):

$$\mathrm{AEI}(\mathbf{x})= \begin{cases}\left(\mu(\mathbf{x})-f\left(\mathbf{x}^{+}\right)-\xi\right)\left(\Phi(Z)+G(\mathbf{x}) D(i)\right) & \\ +\sigma(\mathbf{x}) \phi(Z) & \text{if } \sigma(\mathbf{x})>0 \\ 0 & \text{if } \sigma(\mathbf{x})=0\end{cases} \tag{19}$$

Where i is current iteration number and ξ is hyper-parameter that is usually set at 0.01 [29]. $G(\mathbf{x})$ and $D(i)$ here are defined as a prior function and a decay function, respectively.

The $G(\mathbf{x})$ in (19) is shown below:

$$G(\mathbf{x})=\frac{1}{\sigma \sqrt{2 \pi}} \exp\left(-\frac{(\mathbf{x}-\mathbf{P})^2}{2 \sigma^2}\right) \tag{20}$$

Where $\mathbf{P}$ is the initial point.

The $D(i)$ in (19) is decay function, which could be the following three types: gauss decay, exponential decay and linear decay. To intuitively compare these three decay curves, their formulas are represented as (21) and corresponding graph can be seen as Fig. 4:

$$\begin{aligned} d_1 &= \frac{1}{k_1 \sqrt{2 \pi}} e^{(-\frac{x^2}{2 k_1})} \\ d_2 &= e^{(-k_2 x)} \\ d_3 &= -k_3 * x+1 \end{aligned} \tag{21}$$

where k_1, k_2, k_3 are coefficients that influence decay rate. By experiments, we found the gaussian decay d_1 may perform relatively well in our most experiments.

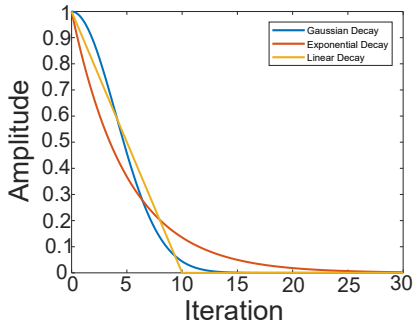

Fig. 4. Three Types of Decay Function.

D. Outline of TCABS

The general structure of TCABS is shown in Fig. 5. The general layers include Hierarchy I and Hierarchy II. Hierarchy I is aimed at feature extraction that refers to implementation details of TCN networks. In addition, Encoder1 and Encoder2, consisting of fully connected layers, process the forward and backward sequences, respectively. In terms of Hierarchy II, inspiration from Bayesian Optimization is integrated into it. The Decoder utilizes the training loss of Hierarchy I to calculate the goal so that the exploration function can know the next exploring point. The surrogate model here can be GP.

E. Application of Cloud Computing

Cloud computing and associated techniques are also applied in our researches. Cloud computing is a concept for providing on-demand network access to a shared pool of customizable computing resources (e.g., networks, servers, storage, applications, and services) that may be quickly supplied and released with no administration effort or service provider contact [30].

High-performance computing (HPC) is a common concept along with cloud computing, which tries to address sophisticated (scientific) computation issues by utilising supercomputers and computer clusters [31]. HPC is still a little different from cloud computing. Cluster computing capacity is frequently fixed, but operating an HPC programme often necessitates much human setting (e.g. tuning based on a particular cluster with a fixed number of homogenous computing nodes). Although the design thought of cloud computing and

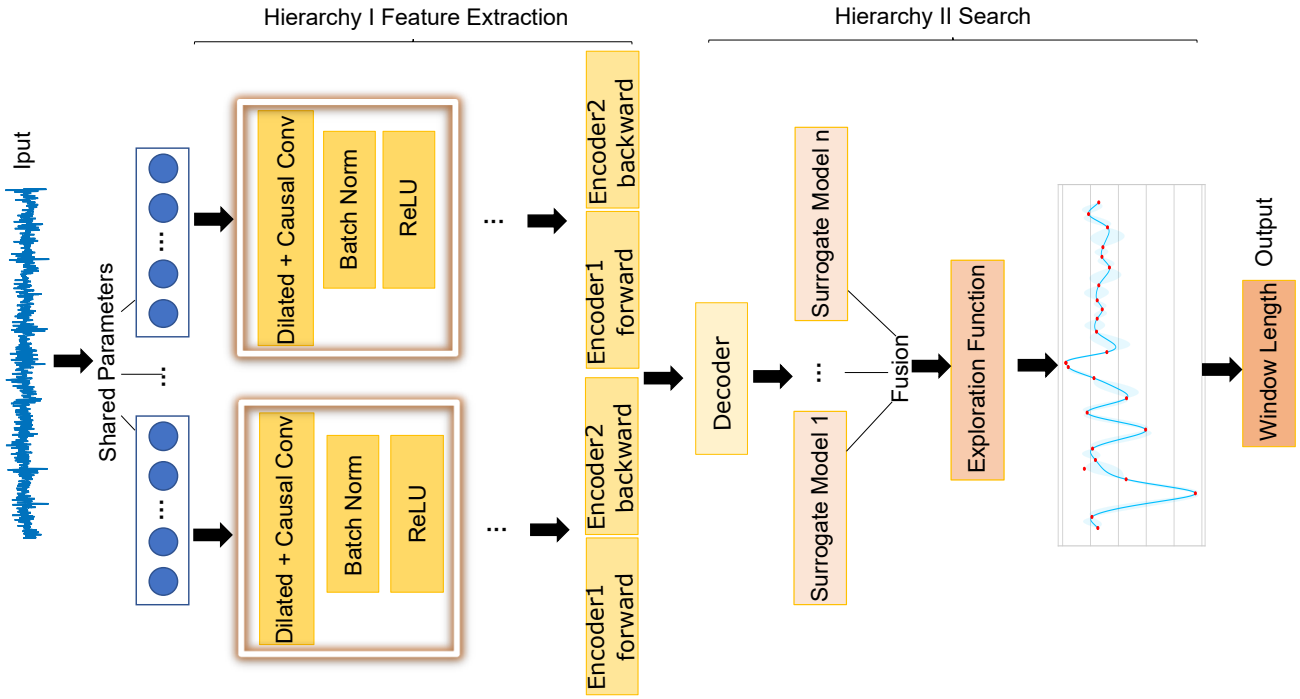

Fig. 5. TCABS Structure.

HPC may be distinct, the cloud computing platforms often realize HPC to accelerate computing processes.

In this paper, the AI studio developed by Baidu, Inc. was selected to realize cloud computing. AI Studio is an open cloud computing platform that can provide advanced HPC in the cloud. By using AI Studio, our proposed TCABS algorithm can be run efficiently. In addition, the data collected from sensors can be injected into cloud clusters to obtain calculation results directly. In traditional ways, the collected data may need to be analyzed by a personal computer, and then it takes a long calculation to get the processed results.

III. Discussion For TCABS

1. Why select TCN architecture as feature extraction module?

- TCN architecture could realize fine-grained control of input length for each step prediction. The input length of historical data could be manipulated according to demand at liberty, so each step's output could only utilize specific length information for forwarding prediction. Many models, such as RNN (Recurrent Neural Network), require all historical information during the whole prediction process, which is unavailing to adjust the window length in this case dynamically.
- Compared with other deep learning models, the training speed, memory occupation, and computation complexity all have advantages for the characteristics of parallel computing for each prediction step. At the same time, the TCN also possesses a strong capability of sequence prediction and information extraction, based on its causal convolutions, dilated convolutions, and residual block.

2. Why AEI acquisition function?

Even if we use the TCN structure, a type of neural network structure that can do parallel computation, it is still not fast enough and time-saving. In order to speed up the search speed and accelerate the convergence of searching process, we did a trick in the selection of initial points and modified the acquisition function so that the search algorithm could search carefully at points with high confidence.

IV. Fault Diagnosis

Our method for diagnosing wind turbine blade bearings consists of three sections: preliminary calculation, signal filter, and fault inference. Firstly, The preliminary calculation section will use TCABS to generate a window length for signal filter. Subsequently the signal filter section will utilized SBAL to obtain denoised signals excellently filtered. Finally, fault inference will resample the denoised signals in order domain to dertermine final fault type. The detailed procedures is shown in Fig. 6.

A. Order Domain Analysis

Frequency or order analysis is necessary to detect the fault frequency and establish the defect type. The frequency or order domain analysis is an important part of fault diagnosis because it may convert the original time domain signal into an intuitive signal in the frequency or order domain, allowing defects to be identified quickly.

For reliable fault detection and diagnosis of the specific fault type of a long-used and damaged bearing, the fault frequency must be calculated [32]. But fluctuated speed during data collection may generate a 'smearing problem' in the frequency

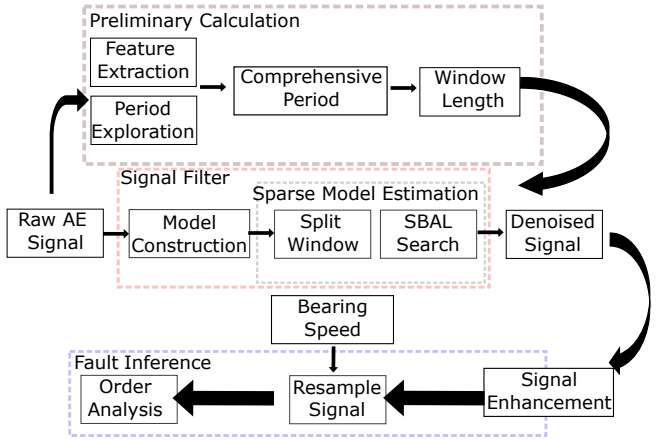

Fig. 6. Complete flow chart of fault diagnosis.

domain, resulting in many fault frequency components. Order analysis with resample techniques and Hilbert transform in the order domain is used to handle this dense challenge [33], [34]. Specific derivation on the fault order calculation can refer to [11].

Finally, Within a brand-new order domain, the denoising processing signal can be rebuilt or resampled, and the order domain is represented as follows:

$$\begin{aligned} O_{outer} &= \frac{N_b}{2} \cdot (1 - \frac{d_b}{d_p} cos\alpha)/R_r \\ O_{inner} &= \frac{N_b}{2} \cdot (1 + \frac{d_b}{d_p} cos\alpha)/R_r \\ O_{ball} &= \frac{d_p}{2d_b} \cdot (1 - (\frac{d_b}{d_p} cos\alpha)^2)/R_r \end{aligned} \tag{22}$$

Where O_{outer}, O_{inner} and O_{ball} denote the fault characteristic frequency (FCO) on inner race, outer race and balls respectively; d_b and d_p denote diameter of ball and pitch respectively; α means bearing contact angle; R_r represents gear ratio, proportional to the speed of bearing .

B. Model Construction

There are two types of signals in the raw acoustics emission (AE) signal obtained from industrial wind turbine blade bearings: deterministic signals and fault signals. [6], [7]. Deterministic signals, such as mechanical rotation motions and electric impulses, are routinely unwelcome components that must be discarded. Fault signals issued by damaged components are useful for further analysis and must be isolated from raw AE data. Generally, with the inclusion of two components, the raw signal from a wind turbine blade bearing may be complete, as shown in (23):

$$y(n) = d(n) + \xi(n), \quad n = 1, ..., M + N \tag{23}$$

$y(n)$ here represents the raw bearing signal. $d(n)$ and $\xi(n)$ denote deterministic signals and fault signals, respectively. The notations M and N refer to the dimensional size of the proposed dictionary matrix P, which will be clarified in the context that follows. Thereinto, the deterministic component can be built using historical recordings from a specific time period [15], with terms of both linearity and non-linearity [35].

Consequently, the deterministic signal component can be represented in matrix format to better express the linearity and non-linearity terms:

$$\hat{D} = P\Theta \tag{24}$$

Here $\hat{D} = [d(M+1), d(M+2), ..., d(M+N)]^T$ and $\Theta = [\theta_1, \theta_2, ..., \theta_M, \theta_{1,1}, \theta_{1,2}, ..., \theta_{M,M}, ..., \underbrace{\theta_{M,...,M}}_{\gamma}]^T$

The dictionary matrix P consistis of ρ elements. Only the first element is the linear term, and the remaining are high order nonlinear terms, as described below:

$$P = \begin{bmatrix} \rho_1 & \rho_2 & ... & \rho_\gamma \end{bmatrix} \tag{25}$$

All elements ρ_i in the matrix have a size of $N \times M$, hence the size of P is $N \times \gamma M$. The complete expression of ρ_γ is shown as follows:

$$\rho_\gamma = \begin{bmatrix} y(M)^\gamma & y(M)^{\gamma-1}y(M-1) & ... & y(1)^\gamma \\ y(M+1)^\gamma & y(M+1)^{\gamma-1}y(M) & ... & y(2)^\gamma \\ \vdots & \vdots & \vdots & \vdots \\ y(M+N-1)^\gamma & y(M+N-1)^{\gamma-1}y(M+N-2) & ... & y(N)^\gamma \end{bmatrix} \tag{26}$$

Finally, for the consistency of expression, the formula (23) here is also reconstructed as a matrix form as follows:

$$Y = P\Theta + \Xi \tag{27}$$

Here $Y = [y(M+1), y(M+2), ..., y(M+N)]^T$ and $\Xi = [\xi(M+1), \xi(M+2), ..., \xi(M+N)]^T$

C. Split Bayesian Augmented Lagrangian Algorithm (SBAL)

The split window idea is introduced for raw AE signals in order to achieve a more fine-grained control for the time-varying conditions. If the window length is defined as L_w and the raw AE signal length is represented as L_r, we can define the window set W:

$$W = [w_1, w_2, ..., w_i] \tag{28}$$

Combined with the aforementioned window length and signal length, the index of last element $i = L_w/L_s$.

Each window w_i in window set W is an independent state, approximate Hidden Markov Model (HMM). It is reasonable to assumed that the current window is just a continuation of the previous window. Then if the appropriate window length L_w is chosen, each window w_i will be relatively dependent because collected raw signals always comply with a certain period. Both of the aforementioned assumptions are valid HMM conditions.

Then, following [36], [37], regarding each window w_i, the l_1-$norm$ minimization can be used to solve the equation fitting problem $Y = P\Theta$, represented as follows:

$$\hat{\Theta}_{k+1} = \arg\min_\theta \frac{1}{2} \left\{ \left\| P\Theta - Y \right\|_2^2 + \lambda \left\| G\Theta \right\|_1 \right\} \tag{29}$$

Where $\hat{\Theta}_{k+1}$ is the alterable parameter in the kth iteration. To avoid over-fitting, $\lambda \in (0, 1)$ is the adjustable hyper-parameter related to penalty. G is a diagonal matrix that impacts each iteration directly. Subsequently, the original BAL algorithm

can be transferred into a window-based method SBAL. The complete steps of SBAL algorithm are shown as follows.

Algorithm 1 SBAL Search

Initialization:
Determine split window length L_w of each window w_i
$k = l = 0, c \in (0,1), \lambda_0 = \mu \in (0,1)$
$v_0 = d_0 = 0, G_0 = I$
while $l < L_w/L_s$ **do**
  while $sign(\hat{\Theta}_{k+1}) = sign(\hat{\Theta}_k)$; $|\hat{\Theta}_{k+1}| - |\hat{\Theta}_k| \to 0$ **do**
    $\hat{\Theta}_{k+1} = (P^T P + \mu G_k^T G_k)^{-1}(P^T Y + \mu G_k^T(v_k + d_k))$
    $v_{k+1} = max(0, (G_k\hat{\Theta}_{k+1} - d_k) - \mu/\lambda_k) - max(0, -(G_k\hat{\Theta}_{K+1} - d_k) - \mu/\lambda_k)$
    $d_{k+1} = d_k - (G_k\hat{\Theta}_{k+1} - v_{k+1})$
    $G_{k+1} = diag[P^T(\lambda_k I + P(diag[|(\hat{\Theta}_{k+1})|./(G_k)])P^T)^{-1}P]^{\frac{1}{2}}$
    $\varepsilon_{k+1} = \left\|P\hat{\Theta}_{k+1} - Y\right\|_2^2$
    $\lambda_{k+1} = \begin{cases} a\lambda_k, (a>1) & if \quad |\frac{\varepsilon_{k+1}}{\varepsilon_k} - 1| < c \\ b\lambda_k, (0<b<1) & if \quad |\frac{\varepsilon_{k+1}}{\varepsilon_k} - 1| \geq c \end{cases}$
    $k = k + 1$
  end while
end while

Ultimately, if the optimal Θ is determined, the denoised signal can also be acquired by the following formula:

$$Y_d = sign(Y)|Y - \Delta\hat{\Theta}_k| \tag{30}$$

where Δ is a constructed raw signal matrix, subset of matrix (26). Complete formula derivation can refer to [36], [37].

To intuitively know the denoising process with SBAL, the inner details for the proposed SBAL denoising process are shown in Fig. 7.

V. Simulation

As can be seen in Fig. 8, the signal is from time-invarying system with fixed period of one second. Fig. 8 (a) is pure fault signal and Fig. 8 (b) is fault signal adding noise. We cannot obtain the specific period from the simple observation on the Fig. 8 (b). According to the proposed TCABS method, the calculated period with largest score can be found at sequence length 513 in Fig. 9, namely 1.026 s, because sampling frequency here is 500.

From Fig. 9, We can observe that the blue shade area is not so large, denoting the exploration process is relatively successful. The blue shaded area around the GP curve represents the mean value minus variance to mean value plus variance, namely, $\mu(\mathbf{x}_t) - \sigma(\mathbf{x}_t)$ to $\mu(\mathbf{x}_t) + \sigma(\mathbf{x}_t)$. It can be viewed as a type of uncertainty degree, and the larger it is, the more unknown information is around the region. In other words, the blue shade area can guide exploration and exploitation.

Fig. 10 shows a signal from time-varying system with a increasing period. From Fig. 11, the sequence length with highest score exists at 3070, namely 6.14 s. This must be not the period for this time-invarying signal, but it can be regarded as a comprehensive period to describe characteristics for this piece of signal.

VI. Experiments

As illustrated in Fig. 12 (a), the industrial-scale wind turbine blade bearing for tests in this article is 261 kg with a pitch diameter of 1.1 m, and it has been operating in a genuine industrial wind power plant for more than 15 years. The industrial wind turbine is with a 7.75 m and 139 kg loading (blade) on the blade bearing, as illustrated in Fig. 12 (b), to represent the real maintenance scenario with loadings. In addition, unlike most research studies, the tests in this work employ naturally developed damage throughout the service term rather than purposely damaged or small-scale models of wind turbines.

Test 1 to Test 5 are experimental sequence data that can be used to further validate the performance of our proposed TCABS to hyper-parameters. The primary hyper-parameters in SBAL are delay M and window length N. To observe the effectiveness of hyper-parameters, exhaustive search with kurtosis indicator is executed in Test 1 when the search space is restricted to a fixed range (window length: 1-1000; delay: 1-100). Following work [37], the kurtosis value is used to measure denoising performance. The high kurtosis value often indicates that small noise exists in signal. Note that here only adopts one window without the split window technique to reduce calculation complexity, but the time to get search results of linear model Fig. 13 (a) and two order nonliear model Fig. 13 (b) are still very long, 4795.556094s and 13011.244237s, respectively. We can observe that the kurtosis will be significantly improved when delay reaches 30, no matter linear or nonliear models. It can be concluded that the delay is not the primary factors to influence model performance and delay 30 may be enough in almost all cases. In addition, the long search time here also validates the necessity to use TCABS for window length search because time consumption for exhaustive search will be more prolonged when using the SBAL algorithm and high model orders.

Subsequently, to better filter signal of Test 1, the nonlinear orders are increased to four, and the SBAL algorithm is used. Combined with TCABS algorithm, the final search length is obtained at 1039 and the time domain results are shown in Fig. 14. To emphasize the effectiveness of TCABS, four other controlled experiments were conducted. The experimental results are shown in Table I. In addition to kurtosis, three AE hit parameters, energy, Root-Mean-Square (RMS), and Average Signal Level (ASL), are utilized to compare these five experiments. The energy is often used to measure the fault signals that is extracted from raw AE signal. The RMS and ASL can be utilized to measure the noise levels in a signal [38]. From Table I, we can find that Test 1 with window length 1039 has the highest kurtosis and energy and lowest RMS and ASL, compared with the other four experiments, denoting that the model with window length 1039 possesses superior performance in noise signal filtering and fault signal extracting results.

Regarding analysis for filtered signals, Fault Characteristic Frequency (FCF) and Fault Characteristic Order (FCO) are useful tools for diagnosing the fault type of wind turbine blade bearings. But, the 'smearing problem' [11] of FCF will make

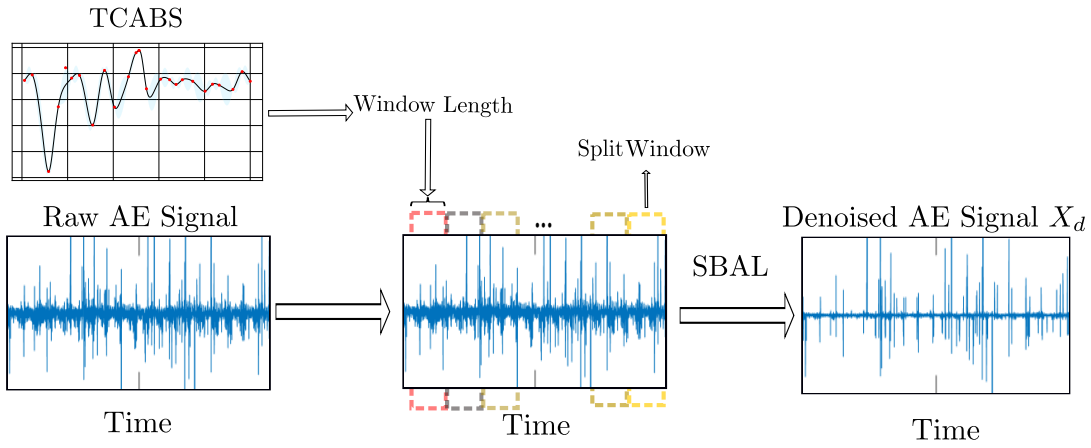

Fig. 7. SBAL denoising process.

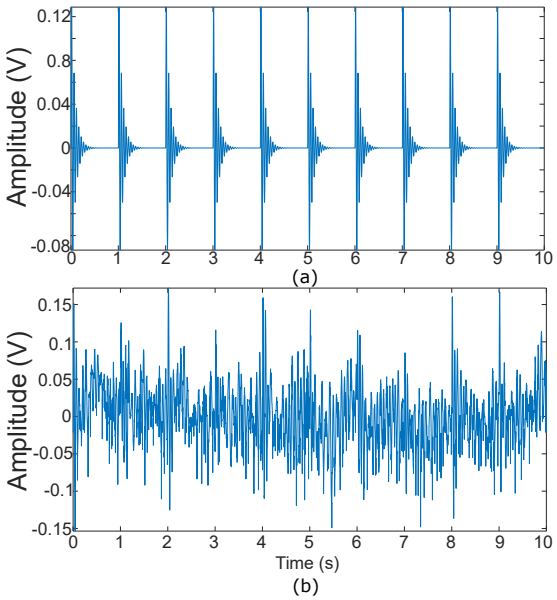

Fig. 8. Time-invarying signals in time Domain.

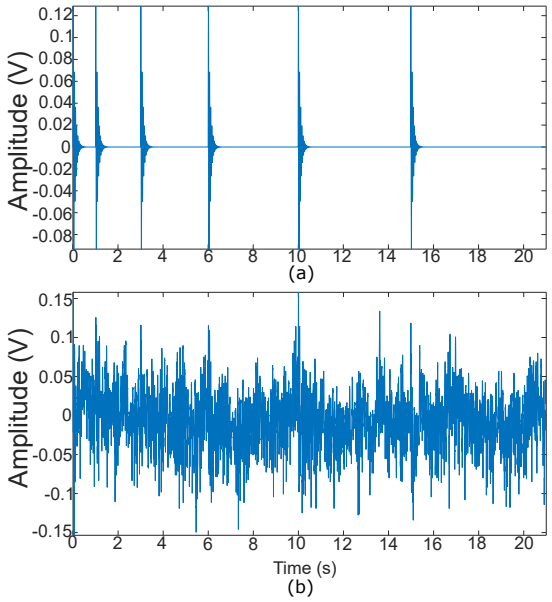

Fig. 10. Time-varying signals in time Domain.

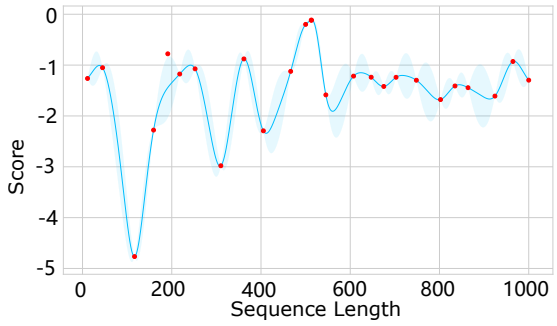

Fig. 9. Search Process of TCABS for time-invarying signals.

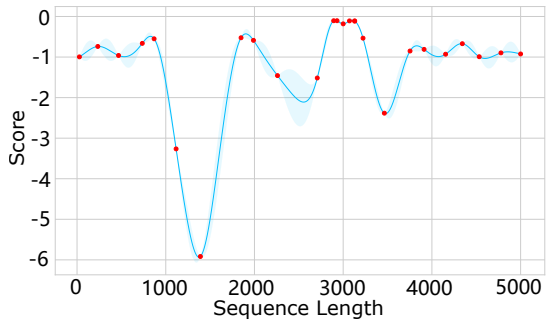

Fig. 11. Search Process of TCABS for time-varying signals.

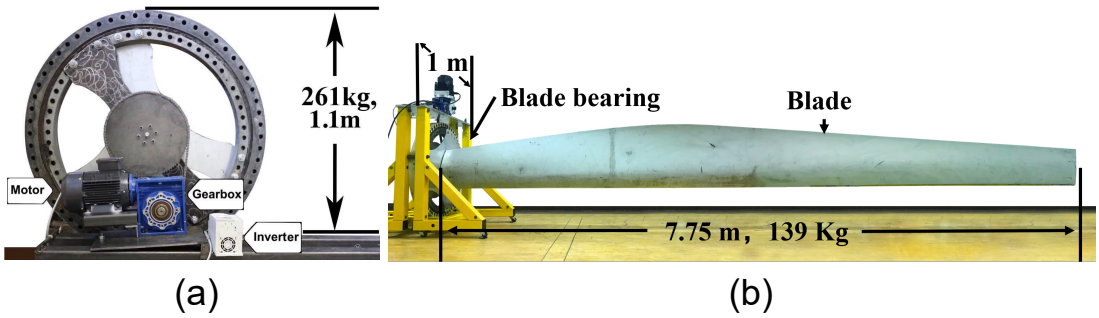

Fig. 12. Test rig of wind turbine blade bearing: (a) side view (b) front view.

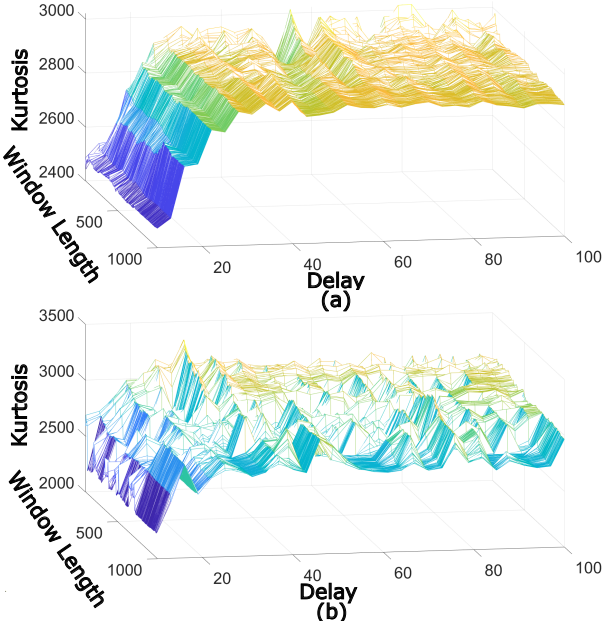

Fig. 13. Search by exhaustion method. (a) search results of linear model (b)search results of two order nonliear model

finding detect frequencies difficult when dealing with time-varying conditions, so FCO is often utilized. According to average bearing speed and inner parameters, the theoretical fault characteristic frequencies can be calculated. Furthermore, by computation of intrinsic parameters of blade bearings, the theoretical fault characteristic order (FCO) of this bearing can be calculated by (22), which is fixed for a specific blade bearing, as shown in Table II. To better represent fault diagnosis results, Fault Indicator (**FI**) is introduced as (31). **FI** value that approaches 100% means matching fault type. According to Fig. 15, the **FI** value for O_{outer} can be calculated: $|5.850 - 5.824|/5.824 \times 100\% = 99.6\%$, so fault type in this bearing may be the inner race fault.

$$\mathbf{FI} = 1 - (|O_I - O_T|/O_T) \times 100\% \tag{31}$$

where O_I is identified FCO, O_T is theoretical FCO.

In order to further verify the accuracy of diagnostic results and the effectiveness of our proposed TCABS and SBAL algorithm, another two experiments (Test a and Test b) in time-

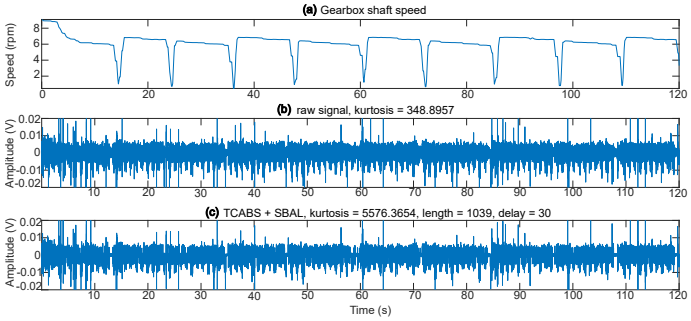

Fig. 14. Time domain results of Test 1.

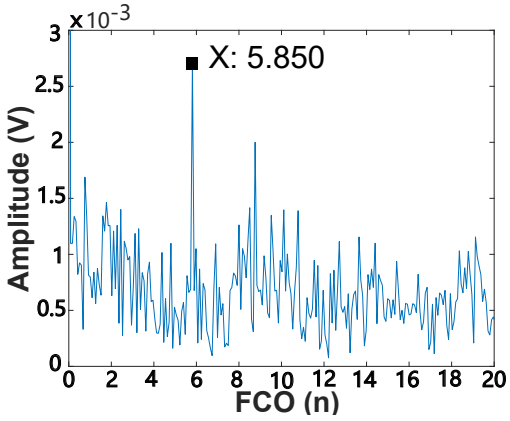

Fig. 15. Order results of Test 1 based on Hilbert envelope.

TABLE I
FIVE CONTROLLED EXPERIMENTS.

Test ID	Window Length	Kurtosis	Energy	RMS	ASL
Test 1	1039	**5576.4**	**1923.1**	**0.0027**	**-160.3**
Test 2	200	4317.6	1773.3	0.0037	-158.2
Test 3	500	3318.8	1752.8	0.0033	-155.6
Test 4	2000	3848.6	1659.7	0.0029	-158.9
Test 5	5000	3812.8	1890.1	0.0030	-157.5

TABLE II
THEORETICAL FCO OF THE WIND TURBINE BLADE BEARING.

R_r	O_{inner}	O_{outer}	O_{ball}
5.33	5.824	5.433	1.735

varying conditions are executed. In addition, the DRS method that has been repeatedly proved useful is used to contrast with our proposed method. Time domain results for Test a and Test b are shown as Fig. 16 and Fig. 17, respectively. Both kurtosis value from DRS method are smaller than our proposed method, denoting that our proposed method is superior to DRS method. Fig. 18 represents the order domain results of Test a and Test b that are obtained from resample signal, denoting identified FCO 5.834 and 5.798, respectively. Subsequently, the **FI** can be calculated as Table III shown. The inner race fault has **FI** value 99.7%, denoting that some damage existing in inner race.

TABLE III
FAULT INDICATOR CALCULATION.

Test	Inner race fault	Outer race fault	Ball fault
Test a	99.8%	92.6%	-136.3%
Test b	99.6%	93.3%	-134.2%
Average	99.7%	93.0%	-135.3%

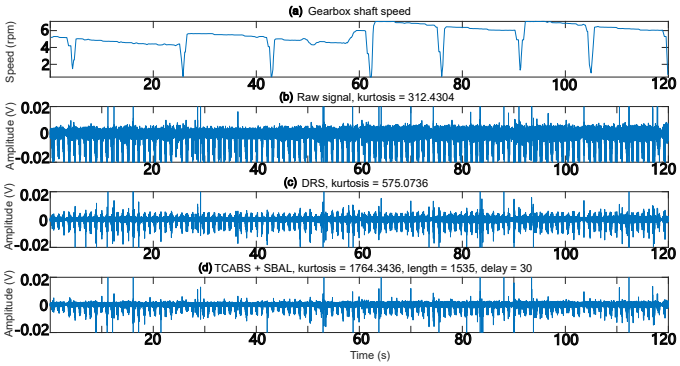

Fig. 16. Time domain results of Test a.

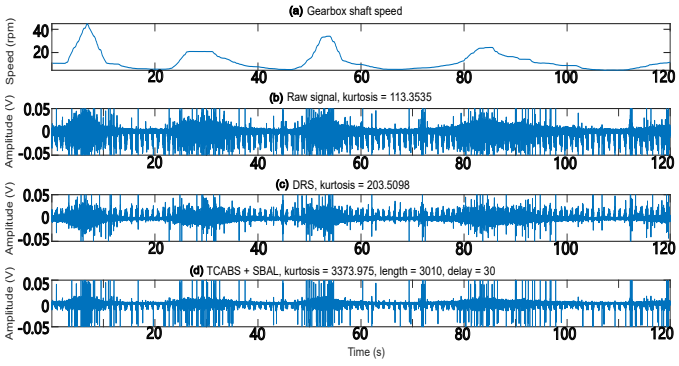

Fig. 17. Time domain results of Test b .

VII. Damage Validation

An electronic endoscope is used to check the industrial-scale wind turbine blade bearing in this work to confirm the results of defect detection. Fig.19 from endoscope denotes that the defect, measured size 9mm length and 5mm height, exists at the bearing inner race. This detection result validates

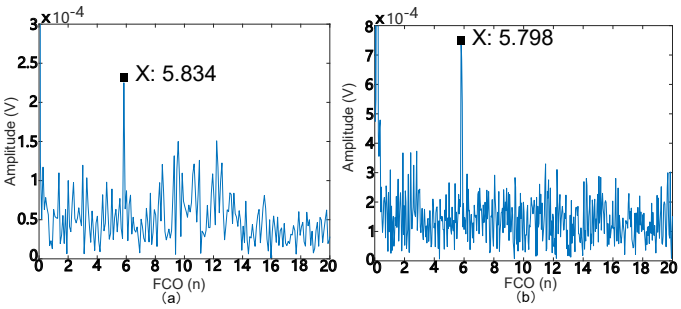

Fig. 18. Order domain results based on Hilbert envelope. (a) order domain results for Test a (b) order domain results for Test b

the accuracy of the previous diagnostic results. Therefore, the suggested TCABS and SBAL approaches in this work can be effective for defect detection of wind turbine blade bearings, indicating the wide application potential on a natural industrial occasion.

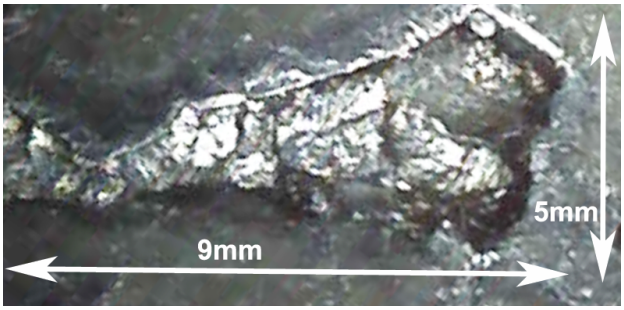

Fig. 19. Inner race defects detected by electronic endoscope.

VIII. Conclusion

In this paper, Temporal Convolutional Augmented Bayesian Search (TCABS) was proposed to determine the hyper-parameter window length for sequence models. The characteristic of TCABS is that it is a type of parallel searching algorithm to find suitable hyper-parameters, which can be executed in distributed systems to accelerate calculation. In addition, the SBAL used for nonlinear filter model construction can determine suitable parameters automatically and swiftly. The primary advantage of SBAL is that it can deal with the time-varying system because of the split window technique introduced to make batch processing for sequence data.

Both simulation and experiments of wind turbine blade bearing were executed to validate the efficacy of our proposed TCABS and SBAL methods. The experimental results show that the TCABS method is also useful in determining the hyper-parameter window length for the SBAL algorithm. Therefore, it is also believed that the TCABS method can be applied to similar window-based methods for automatic determination of the window length. In addition, by comparing kurtosis values, the SBAL method of parameter estimation for model construction can perform better in noise signal filtering and fault signal extraction than the traditional discrete/random separation (DRS) method.

REFERENCES

[1] K. K. Dhanda and L. P. Hartman, "The ethics of carbon neutrality: A critical examination of voluntary carbon offset providers," *Journal of Business Ethics*, vol. 100, no. 1, pp. 119–149, 2011.

[2] L. Tozer and N. Klenk, "Discourses of carbon neutrality and imaginaries of urban futures," *Energy research & social science*, vol. 35, pp. 174–181, 2018.

[3] Z. Liu, L. Zhang, and J. Carrasco, "Vibration analysis for large-scale wind turbine blade bearing fault detection with an empirical wavelet thresholding method," *Renewable Energy*, vol. 146, pp. 99–110, 2020.

[4] R. Poore, "Development of an operations and maintenance cost model to identify cost of energy savings for low wind speed turbines: July 2, 2004–june 30, 2008," National Renewable Energy Lab.(NREL), Golden, CO (United States), Tech. Rep., 2008.

[5] D. Mba and R. Rao, "Development of acoustic emission technology for condition monitoring and diagnosis of rotating machines: bearings, pumps, gearboxes, engines, and rotating structures," *Shock and Vibration Digest*, vol. 38, no. 1, pp. 3–18, 2006.

[6] J. Antoni and R. Randall, "Unsupervised noise cancellation for vibration signals: part i—evaluation of adaptive algorithms," *Mechanical Systems and Signal Processing*, vol. 18, no. 1, pp. 89–101, 2004.

[7] ——, "Unsupervised noise cancellation for vibration signals: part ii—a novel frequency-domain algorithm," *Mechanical Systems and Signal Processing*, vol. 18, no. 1, pp. 103–117, 2004.

[8] D. Zhang, Y. Fu, Z. Lin, and Z. Gao, "A reinforcement learning based fault diagnosis for autoregressive-moving-average model," in *IECON 2017-43rd Annual Conference of the IEEE Industrial Electronics Society*. IEEE, 2017, pp. 7067–7072.

[9] M. Yuan, Y. Wu, and L. Lin, "Fault diagnosis and remaining useful life estimation of aero engine using lstm neural network," in *2016 IEEE international conference on aircraft utility systems (AUS)*. IEEE, 2016, pp. 135–140.

[10] J. Antoni, "Fast computation of the kurtogram for the detection of transient faults," *Mechanical Systems and Signal Processing*, vol. 21, no. 1, pp. 108–124, 2007.

[11] Z. Liu, X. Tang, X. Wang, J. E. Mugica, and L. Zhang, "Wind turbine blade bearing fault diagnosis under fluctuating speed operations via bayesian augmented lagrangian analysis," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 7, pp. 4613–4623, 2020.

[12] K. Chen, L. Kurgan, and J. Ruan, "Optimization of the sliding window size for protein structure prediction," in *2006 IEEE Symposium on Computational Intelligence and Bioinformatics and Computational Biology*. IEEE, 2006, pp. 1–7.

[13] Z. Ding and M. Fei, "An anomaly detection approach based on isolation forest algorithm for streaming data using sliding window," *IFAC Proceedings Volumes*, vol. 46, no. 20, pp. 12–17, 2013.

[14] Y. Shynkevich, T. M. McGinnity, S. A. Coleman, A. Belatreche, and Y. Li, "Forecasting price movements using technical indicators: Investigating the impact of varying input window length," *Neurocomputing*, vol. 264, pp. 71–88, 2017.

[15] Z. Liu and L. Zhang, "Naturally damaged wind turbine blade bearing fault detection using novel iterative nonlinear filter and morphological analysis," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 10, pp. 8713–8722, 2019.

[16] Z. Li, X. Yan, Z. Tian, C. Yuan, Z. Peng, and L. Li, "Blind vibration component separation and nonlinear feature extraction applied to the nonstationary vibration signals for the gearbox multi-fault diagnosis," *Measurement*, vol. 46, no. 1, pp. 259–271, 2013.

[17] W. Huang, Z. Song, C. Zhang, J. Wang, J. Shi, X. Jiang, and Z. Zhu, "Multi-source fidelity sparse representation via convex optimization for gearbox compound fault diagnosis," *Journal of Sound and Vibration*, vol. 496, p. 115879, 2021.

[18] Y. Liao, W. Huang, C. Shen, Z. Zhu, J. Xuan, and L. Mao, "Enhanced sparse regularization based on logarithm penalty and its application to gearbox compound fault diagnosis," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–12, 2021.

[19] J. Lafferty, A. McCallum, and F. C. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," 2001.

[20] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[21] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," *arXiv preprint arXiv:1803.01271*, 2018.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[23] R. N. Bracewell and R. N. Bracewell, *The Fourier transform and its applications*. McGraw-Hill New York, 1986, vol. 31999.

[24] E. O. Brigham, *The fast Fourier transform and its applications*. Prentice-Hall, Inc., 1988.

[25] E. Brochu, V. M. Cora, and N. De Freitas, "A tutorial on bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning," *arXiv preprint arXiv:1012.2599*, 2010.

[26] R. Martinez-Cantin, N. De Freitas, E. Brochu, J. Castellanos, and A. Doucet, "A bayesian exploration-exploitation approach for optimal online sensing and planning with a visually guided mobile robot," *Autonomous Robots*, vol. 27, no. 2, pp. 93–103, 2009.

[27] E. Brochu, T. Brochu, and N. De Freitas, "A bayesian interactive optimization approach to procedural animation design," in *Proceedings of the 2010 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, 2010, pp. 103–112.

[28] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes with Source Code CD-ROM 3rd Edition: The Art of Scientific Computing*. Cambridge University Press, 2007.

[29] D. R. Jones, "A taxonomy of global optimization methods based on response surfaces," *Journal of global optimization*, vol. 21, no. 4, pp. 345–383, 2001.

[30] F. Liu, J. Tong, J. Mao, R. Bohn, J. Messina, L. Badger, D. Leaf *et al.*, "Nist cloud computing reference architecture," *NIST special publication*, vol. 500, no. 2011, pp. 1–28, 2011.

[31] T. Dillon, C. Wu, and E. Chang, "Cloud computing: issues and challenges," in *2010 24th IEEE international conference on advanced information networking and applications*. Ieee, 2010, pp. 27–33.

[32] W. Huang, G. Gao, N. Li, X. Jiang, and Z. Zhu, "Time-frequency squeezing and generalized demodulation combined for variable speed bearing fault diagnosis," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 8, pp. 2819–2829, 2018.

[33] F. Bonnardot, M. El Badaoui, R. Randall, J. Daniere, and F. Guillet, "Use of the acceleration signal of a gearbox in order to perform angular resampling (with limited speed fluctuation)," *Mechanical Systems and Signal Processing*, vol. 19, no. 4, pp. 766–785, 2005.

[34] L. Marple, "Computing the discrete-time" analytic" signal via fft," *IEEE Transactions on signal processing*, vol. 47, no. 9, pp. 2600–2603, 1999.

[35] R. Huang, F. Xu, and R. Chen, "General expression for linear and nonlinear time series models," *Frontiers of Mechanical Engineering in China*, vol. 4, no. 1, pp. 15–24, 2009.

[36] X. Tang, L. Zhang, and X. Li, "Bayesian augmented lagrangian algorithm for system identification," *Systems & Control Letters*, vol. 120, pp. 9–16, 2018.

[37] J. Wang, W. Qiao, and L. Qu, "Wind turbine bearing fault diagnosis based on sparse representation of condition monitoring signals," *IEEE Transactions on Industry Applications*, vol. 55, no. 2, pp. 1844–1852, 2018.

[38] Z. Liu, X. Wang, and L. Zhang, "Fault diagnosis of industrial wind turbine blade bearing using acoustic emission analysis," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 9, pp. 6630–6639, 2020.