Wind turbine pitch bearing fault detection with Bayesian augmented temporal convolutional neural networks

Journal Title
XX(X):2–32

DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

Wind turbine pitch bearing fault detection with Bayesian augmented temporal convolutional neural networks

Chao Zhang[1] and Long Zhang[2]

Abstract

There are few studies on the fault diagnosis of deep learning in real large-scale bearings, such as wind turbine pitch bearings. We present a novel fault diagnosis method, Bayesian augmented temporal convolutional neural networks (BATCN), to filter the raw signal in wind turbine pitch bearing defect detection. This method, which employs temporal convolutional neural networks, is designed to capture the temporal dependencies of the signal, with such a focus on non-stationary relationships in the collected signals. By referring to the thoughts of Bayesian optimization, our approach can spontaneously find the best patch length that influences fault signal extraction during the filtering process, avoiding manual tuning of this hyper-parameter. This BATCN method is first performed on simulation signals and an open-source dataset of general bearings, and then validated on industrial wind turbine pitch bearings both in the lab and in the real wind farm, where the bearings have been operated for over 15 years. The results show that our method can work well for large-scale slow-speed wind turbine pitch bearings.

Keywords

Fault diagnosis, wind turbine pitch bearing, slow speed, non-stationary, convolutional neural networks

Introduction

Wind power, as a sustainable and reliable energy source [1,2], its installation capacity has expanded at a predominant pace in recent years throughout the world, and this trend is expected to continue constantly in order to achieve carbon emission target in 2050 and provide sustainable electric energy [3]. Pitch bearings, also referred to as blade bearings, one of the most significant components of wind turbines, can improve the generating efficiency and ensure the safe operation of wind turbine systems [4,5]. However, the extreme operating conditions in industrial occasions may result in serious faults, along with catastrophic accidents and significant financial losses [6,7]. Therefore, Condition Monitoring and Fault Diagnosis (CMFD) [8] of wind turbine pitch bearings driven by measured data is one of the feasible solutions for guaranteeing reliability and controlling maintenance costs [9]. In the CMFD, the vibration or acoustic signal can be used as the fundamental unit to diagnose the faults existing in bearings [10,11].

When faults occur in bearings, they often generate periodic or quasi-periodic fault signals [12]. However, fault signals are weak under slow speed conditions because low rotation results in little kinetic energy, owing to Newton's Law. In addition, weak fault signals are often masked by background noise (from natural distractions) and harmonic interference (from bearing rotation, gearbox, and motor driving) [13]. For the reasons mentioned above, fault signals in wind turbine pitch bearings are usually challenging to be extracted from collected raw signals.

Over the last few decades, some classical denoising methods have been proposed for filtering signals to extract fault signals. The discrete/random separation (DRS) method [15,16] filters raw signals by eliminating the periodic property in the time domain. The spectral kurtosis (SK) method [17] directly extracts fault signals in a narrow frequency band by finding the best central frequency and band width with the aim of determining the best inverse filter [18]. Some researchers recently have also focused on other potential approaches in bearing defect diagnosis, like dictionary methods

Department of Electrical and Electronic Engineering, University of Manchester, UK

Corresponding author:
Long Zhang, Department of Electrical and Electronic Engineering, University of Manchester, UK
Email: long.zhang@manchester.ac.uk

and statistical methods. Regarding dictionary methods, Bayesian and Lagrangian based method was proposed to construct a noise dictionary for collected raw signals[19]. The correlation filtering approach was used to create an adaptive over-complete dictionary using the unit impulse response function[20]. As for statistical methods, Guo et al.[21] used the matrix decomposition method k-means singular value decomposition (K-SVD) for defect detection of wind turbine bearings. Aye et al.[22] used the Principal Component Analysis (PCA) approach to gently identify bearing degradation under varied slow-speed situations.

In recent years, the deep learning techniques, especially convolutional neural networks (CNNs)[23,24], have attracted significant research attention on rolling bearing fault detection. Zhang et al.[25] integrated the function of the principal component analysis (PCA) into CNN to process the fault signals from multi-dimensional bearing vibration signals. Chen et al.[26] researched a transfer learning to augment CNN so that the experience in bearing and gear faults can be reused. Li et al.[27] investigated a novel wavelet driven deep neural network of using the wavelet kernels in the first network layer so that it can offer the CNN the ability to discover weak fault signals of gearbox. Kumar et al.[28] apply a sparsity CNN to diagnose bearing faults under the condition of limited training samples.

Currently, most of the deep learning methods are designed for direct fault classification and judgement. Only few work is related to direct signal denoising of vibration signals from bearings. Wang et al.[29] proposed a novel joint learning CNN for bearing condition monitoring. This method is divided into fault diagnosis and signal denoising module which can obtain good noise robustness through the decoupling functional modules. In the signal denoising module, the encoder-decoder network structure is utilized, attention-based encoder for useful fault signals and decoder for recovering signal details. This method is validated with excellent signal denoising performance. Although this work studied the signal denoising for mechanical bearings with deep learning techniques, large-scale mechanical rotating bearings (such as wind turbine bearings) with special noise levels due to friction and collision between different components has not been studied. This part of the research gap needs to be filled.

Additionally, a seemingly insurmountable technical problem of determining the patch length may exist in the signal processing during

the fault diagnosis process. The patch length during the processing of vibration signals is a key factor influencing the processing performance. The deep learning method considering the input patch length was used for intelligent fault diagnosis of rolling bearings in[30], but suitable patch length was only given by some trials. The one-dimensional signals is split with fixed patch length and the determination of patch length can be ignored by the subsequent processing of transformer encoder in[31]. However, this is just another way indirectly thinking about patch length. The fault pluse extraction can be realized by dictionary learning method in[32], but the patch length in this work was only given by empirical setting.

On top of the aforementioned challenges, there are some open challenges in the field of wind turbine pitch bearing fault diagnosis. More specifically, wind turbine pitch bearing is operated under oscillating operating speed, leading to non-stationary fault signals. Although the existing filtering methods for constant rotating speed can still be applied, the performance may be limited when the fault signal is weak or harmonic interference (e.g. introduced by driving rotor and gear) strongly couples with fault frequency band. Furthermore, vibration based fault detection often requires high sampling frequency, producing large amount of data and redundant information. This introduces high computational loading when using complex filtering methods, such as deep neural networks.

This paper proposed a robust denoising method that can be applied for collected signals to realize fault diagnosis of wind turbine pitch bearing. It is the first time to use CNN for wind turbine pitch bearing fault extraction. This method utilizes the strong feature extraction capability of convolutional networks to process collected discrete signals consisting of much redundant information. We proposed a new fault diagnosis method, called Bayesian augmented temporal convolutional networks (BATCN), to detect defects in wind turbine pitch bearings. The BATCN method, using deep learning techniques, can learn the intrinsic features of fault signals to guarantee the extraction of non-stationary fault signals. The proposed BATCN method learned not only the fault signal and period information but also the dependency relationship among different time intervals, so that it can deal with the non-stationary signals well. By employing global characteristics of convolutional networks, the BATCN also has the ability of harmonic interference suppression. In addition, the dilated convolution in BATCN can realize the sparsity so that the computation complexity

can be reduced, to a large extent. Referring to Bayesian optimization, our approach can spontaneously find the best patch length in the BATCN method so that the proposed method can perform well without complex adjusting parameters.

Regarding the research design, it can explain the work of this paper from another perspective. Specifically, the classic SK method has been repeatedly demonstrated to be very effective in general bearing signal denoising, so it will be used as the primary comparing method in this paper. Unlike the processing complexity for image and video streaming, the one-dimensional signals are relatively straightforward, so the CNN may be sufficiently effective and hence are the focus of this paper.

Combined with the foregoing design analysis, the primary contributions of this paper can therefore be summarized as follows:

- This paper proposes a novel method called BATCN of using the CNN technique to denoise the signals of wind turbine pitch bearings, which can achieve excellent filtering performance without adjusting patch length.
- This is presumably the first study to verify that deep learning can improve the denoising performance of wind turbine pitch bearings. Compared with the classical SK method, the deep learning method is proved more effectively in some conditions of the wind turbine pitch bearings.

The rest of this work is laid out as follows: Section II is dedicated to the complete fault diagnosis procedure for wind turbine pitch bearings and some prior knowledge on fault diagnosis. In Section III, the theoretical process of proposed bayesian augmented temporal convolutional networks (BATCN) is explained in detail. Section IV uses two simulations to validate the noise suppression and harmonic interference suppression, respectively. Section V subsequently presents experimental comparisons between the proposed BATCN approach and the classical SK approach. Section VI shows the meta-analysis of the proposed method, and Section VII concludes the work in this paper.

Fault Diagnosis

This section would present the complete fault diagnosis procedure for wind turbine pitch bearings, which could be divided into two procedures:

signal filter and fault inference. Firstly, the signal filter procedure would utilize BATCN to filter raw acoustic emission (AE) signals and then get denoised signals. Subsequently, fault inference would calculate the fault frequency to identify corresponding fault type by envelope signals. Fig. 1 shows the concrete procedures.

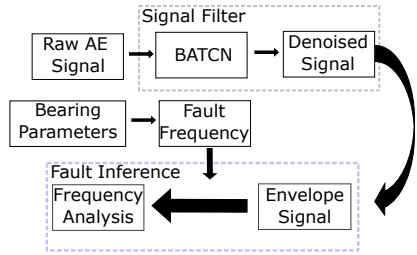

Figure 1. Procedures of fault diagnosis

The following section would briefly explain the fault frequency and envelope techniques used in this paper. The fault frequency (FF) is a key parameter used to identify the type of fault during the frequency analysis process. Through frequency analysis, characteristic information such as the FF can be obtained, allowing defects to be accurately identified based on the calculated FF[39]. The FF could be obtained by (1).

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

where f_{inner}, f_{outer}, and f_{ball} respectively signify the fault frequency (FF) of the inner race, outer race, and balls. The ball number is N_b, the ball and pitch diameters are d_b and d_p, respectively. The bearing contact angle is α, and the rotating frequency of wind turbine pitch bearings is f_r.

With regards to the envelope processing, this paper focuses exclusively on the use of the Hilbert transform. By utilizing the complex signal, the envelope and phase can be calculated conveniently. However, it should be noted that all signals collected from natural sources are real signals,

and the complex signals could not be obtained directly. Therefore, the Hilbert transform is used to construct complex signals for the purposes of envelope processing. This could also be called Hilbert envelope.

Denoising BATCN method

The architecture of proposed BATCN is illustrated in Fig. 2. It is worth noting that the feature extraction module in BATCN refers to a type of parallel method called temporal convolutional neural networks. In addition, this proposed BATCN method utilizes Bayesian optimization to find the optimal patch length to filter signals better. Meanwhile, the forward and backward sequences are processed by Encoder 1 and Encoder 2, respectively, both of which are made up of fully connected layers. The Decoder calculates the goal using the training loss output from the Encoders and the kurtosis information of signals so that the exploration function may determine the next exploring location. The Gaussian Process can be used as a surrogate model in this case.

Feature Extraction Module

This feature extraction module in BATCN method includes three segments: causal convolutions, dilated convolutions, and residual connections. Fig. 3 demonstrates the architecture of feature extraction, which processes the discrete sequences of input.

Causal Convolutions To obtain characteristic information from unidentified sequence data, causal convolution is applied, inspired by the fully-convolutional networks (FCN)[33]. Utilizing causal convolutions, the realistic data derived from low-layer convolution operations can be transmitted to the high-layer with abstract form. The information in the higher layers becomes increasingly intensive until the projected value is output in the top layer, as illustrated in Fig. 3.

The causal convolutions describe the relationship between two adjacent layers. Assume input vector of the previous layer $\boldsymbol{x} = [x_0, ..., x_{n-1}]$ and input vector of the following layer $\boldsymbol{y} = [y_0, ..., y_{n-1}]$, and x_i is input sequence data at location i in a previous layer, and y_i is input data at location i for the following layer. The nominal relationship of causal convolutions could be represented as follows:

$$\boldsymbol{y}_i = F_c\left(\{\boldsymbol{x}_{si+\delta i}\}_{0\leq\delta i\leq k}\right) \tag{2}$$

where k is the kernel size, s is the stride, and F_c means the multiplication for convolution. Note that δi here is continuous positive integer.

Dilated Convolutions Although causal convolutions can extract feature information well, this may still generate deep networks with many layers. According to study[34], dilated convolution may be used to minimise the depth of neural networks by extending the receptive field. When the input size is known, dilated convolution refers to a longer history than ordinary causal convolution, resulting in fewer convolution operations and, a reduction in the number of layers necessary for a neural network.

For the sequence data input vector $\boldsymbol{x} = [x_0, ..., x_{n-1}]$, the dilated convolution operation F_d is the special realization of F_c, defined as:

$$F_d(s) = (\boldsymbol{x} * f_d)(s) = \sum_{i=0}^{k-1} f(i) \cdot x_{s-d \cdot i} \tag{3}$$

where f_d is symbolism of dilated filter with a filter $f : \{0, \ldots, k-1\} \rightarrow \mathbb{R}$. k is the filter size and d is the dilation factor. $s - d \cdot i$ is responsible for the direction of past data. It is worth mentioning that a dilated convolution becomes a normal convolution when interval $d = 1$.

Residual Connection The neural networks are still deep, despite the use of dilated convolutions, resulting in awful training loss curve. To minimise performance degradation from deep networks, a residual connection[35] developed for image processing may also be used for sequence data processing. When compared to the standard output $\boldsymbol{y} = F_c(\boldsymbol{x})$, the new output employing a residual connection can be defined as follows:

$$\boldsymbol{y} = \boldsymbol{x} + F_c(\boldsymbol{x}) \tag{4}$$

Here the input and output of the current layer are $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively. Observing (4), the residual connection is a sort of connection block that is both flexible and pluggable. When a deep layer becomes redundant, the (4) in this layer becomes $\boldsymbol{y} = \boldsymbol{x}$, indicating identity mapping. This method has consistently been shown to benefit extremely deep networks, because it may be easier to achieve this form of identity mapping when certain layers are redundant. This sort of residual connection is depicted in Fig. 3, where the intermediate layer is bypassed when the networks think that the identity mapping may result in superior performance throughout the training phase.

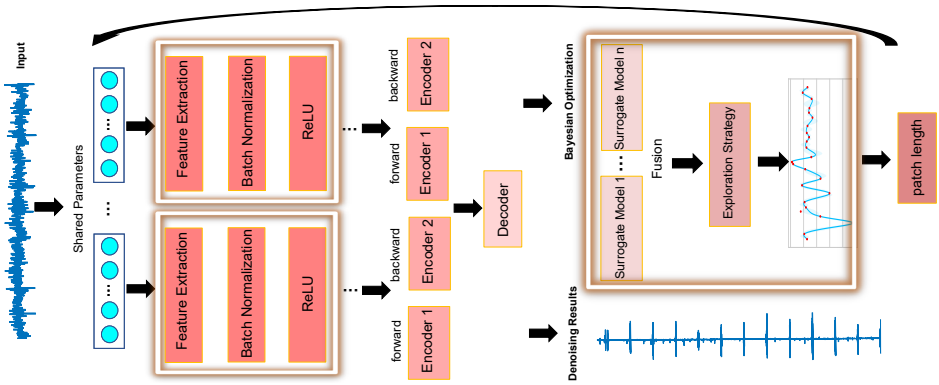

Figure 2. BATCN structure

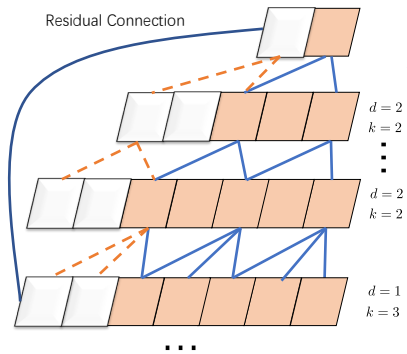

Figure 3. Feature extraction module

Determination of Patch Length with Bayesian Optimization

Before training the neural networks, the patch length, also referred to as the input size for the training dataset, is crucial for denoising performance. For signal sequence data $\boldsymbol{x} = (x_0, y_0), ...(x_{n-1}, y_{n-1})$, if the input size (patch length) is l and the sequence length is just an integer multiple of l, train dataset consists of multiple patches: $\boldsymbol{x} = \{\{(x_0, y_0), ...(x_{l-1}, y_{l-1})\}, \{(x_l, y_l), ...(x_{2l-1}, y_{2l-1})\}$ $, ..., \{(x_{n-l-2}, y_{n-l-2}), ...(x_{n-1}, y_{n-1})\}\}$. If not an integer multiple, the last patch can not be filled with the number of input size$s$, and the truncation or padding techniques can be used. This paper will apply truncation to this special edge case. In other words, each patch in training dataset is independent that means the whole training process can be parallel to speed

up the training. This is one of the reasons why we refer to this TCN network structure.

To some extent, the patch length during model construction is a factor that will influence the performance of denoising models. However, trial-and-error methods, such as enumeration method, are unrealistic because every single training session is very time-consuming. The aimless search may not be suitable for this occasion, so a Bayesian optimization based search strategy is used for finding the best patch length automatically.

Search process for patch length It is assumed that l is a random variable, representing the patch length to be determined. $f(l)$ represents the real objection function about l, which is related to loss function of neural networks and has high computational complexity. The patch length vector is $\boldsymbol{l} = [l_0, l_1, ..., l_t, ..., l_T]$ and T is the maximum iteration. Regarding the surrogate model in Bayesian optimization, a gaussian process (GP) is chosen as follows:

$$f(l) \sim \mathcal{GP}\left(m(l_t), k\left(l_t, l_{t'}\right)\right) \tag{5}$$

where the $t, t' \in (0, T)$, and then we can see that the mean function m and the covariance function k totally determine the GP. If we can approximate the real GP in (5), the $f(l)$ can also be obtained (one optional way is to regard the mean value as the $f(l)$). To be simplified, it can be assumed here that the prior mean $m(l_t) = 0$.[36,37] can also be used to refer to a different mean value. The covariance here $k(\cdot)$ can be represented with exponential function and L_2-norm, as follows:

$$k\left(l_t, l_{t'}\right) = \mathbf{exp}(-\frac{1}{2}\|l_t - l_{t'}\|^2) \tag{6}$$

Subsequently, the Sherman-Morrison-Woodbury formula[38] can be used to generate a predictive distribution P, namely, normal distribution $\mathcal{N}$ with mean $\mu(l)$ and variance $\sigma(l)$, at tth iteration:

$$P\left(f_t \mid \mathcal{D}_{0:t-1}, l_t\right) = \mathcal{N}\left(\mu_t\left(l_t\right), \sigma_t^2\left(l_t\right)\right) \tag{7}$$

where $f_t = f(l_t)$ and $\mathcal{D}_{0:t-1}$ is the historical observation from iteration 0 to $t-1$. The $\mu_t(\cdot)$ and $\sigma_t(\cdot)$ are represented as follows:

$$\begin{aligned} \mu_t\left(l_t\right) &= \mathbf{k}^T \mathbf{K}^{-1} f_{0:t} \\ \sigma_t^2\left(l_t\right) &= k\left(l_t, l_t\right) - \mathbf{k}^T \mathbf{K}^{-1} \mathbf{k} \end{aligned} \tag{8}$$

where $f_{0:t}$ is a column vector, $f_{0:t} = f(l_{0:t})$. $\mathbf{k}$ and $\mathbf{K}$ are coefficient matrixes obtained according to formula (6) and are described as follows:

$$\begin{aligned} \mathbf{k} &= \begin{bmatrix} k(l_t, l_0) & k(l_t, l_1) & \cdots & k(l_t, l_t) \end{bmatrix}^T \\ \mathbf{K} &= \begin{bmatrix} k(l_0, l_0) & \dots & k(l_0, l_t) \\ \vdots & \ddots & \vdots \\ k(l_t, l_0) & \dots & k(l_t, l_t) \end{bmatrix} \end{aligned} \tag{9}$$

According to (7), the GP can be updated:

$$f(l) \sim \mathcal{GP}(\mu_t, \sigma_t^2) \tag{10}$$

From (5) to (10), we can know the procedures of updating the GP. Furthermore, the acquisition function for choosing the next point needs to be concretely explained.

Some acquisition functions related to the exploration approach have been empirically identified as effective on most occasions. The three primary acquisition functions are explained as follows:

- Upper confidence bound (*UCB*):

$$UCB(l) = \mu(l) + \rho\sigma(l) \tag{11}$$

where ρ is a weight parameter, larger than 0, adjustable parameter.

- Probability of improvement (*PI*):

$$PI(l) = \Phi\left(\frac{\mu(l) - f(l^+) - \lambda}{\sigma(l)}\right) \tag{12}$$

where $l^+ = \arg\max_{l_i \in l_{0:t}} f(l_i)$, the λ is larger than 0, left to the user. Φ is cumulative distribution function (CDF) of standard normal distribution

- Expected improvement (*EI*):

$$EI(l) = \left(\mu(l) - f\left(l^+\right) - \xi\right)\Phi(Z) + \sigma(l)\phi(Z) \tag{13}$$

where $Z = (\mu(l) - f(l^+))/\sigma(l)$; ϕ and Φ are probability density function (PDF) and cumulative distribution function (CDF) of standard normal distribution, respectively. The ξ is larger than 0, left to the user.

Regarding selection of acquisition function, global search for *UCB* is simple, but convergence speed is slow. *PI* can quickly converge, but it is possible to become stuck in a locally optimum solution. *EI* is able to find a balance between global and local optimization.

Finally, we select *EI* acquisition function as the exploration strategy, because it will make trade-off between global optimization and local search.

Simulation Validation

Noise and harmonic interference are significant factors that may influence the fault diagnosis of wind turbine pitch bearings. This section firstly introduce simulation signals, including: fault signals, harmonic signals, and noise signals. Subsequently, noise and harmonic interference suppression of the proposed BATCN method will be examined, respectively. In addition, spectral kurtosis (SK)[17], a denoising approach that has been proven to be effective in a variety of general bearing applications, was used for comparison.

Simulated Signals

The simulation signal is comprised of three parts: fault signals $\mathbf{x} = [x(1), ..., x(t)]$, harmonic signals $\mathbf{h} = [h(1), ..., h(t)]$ and noise signals $\mathbf{n} = [n(1), ..., n(t)]$. The details for each part are given as follows.

Fault Signals: At time t, the fault signal $x(t)$, consisting of sinusoidal impulse signal $I(t)$, are shown as follows:

$$x(t) = \sum_{k} a_k I((t - kM - \tau_k)) \tag{14}$$

The standard form of $I(t)$ is shown as follows:

$$I(t) = \mathrm{e}^{-2\xi\pi f_n t} \sin\left(2\pi f_n \sqrt{1-\xi^2} t\right) \tag{15}$$

where different a_k on simulation model setting indicates various amplitude of impulse components in fault signals. M is the period it takes for an impulse to arise. τ_k denotes the fine random interval bias among impulses. f_n is resonance frequency. ξ is the damping ratio that influences the shape of impulse components.

Harmonic Signals: To effectively test the ability of harmonic interference suppression, the harmonic signals h here are combination of two harmonic interference and specified as follows:

$$\begin{cases} h(t) = h_1(t) + h_2(t) \\ h_1 = h_a(1 + 0.5\cos(2\pi * 180t)) * \cos(2\pi * 1000t + 0.5\cos(2\pi * 35t)) \\ h_2 = h_b(1 + 0.5\cos(2\pi * 360t)) * \cos(2\pi * 2000t + 0.5\cos(2\pi * 70t)) \end{cases} \tag{16}$$

where h_1 and h_2 are two types of harmonic interference. h_a and h_b represent the amplitude of h_1 and h_2, respectively.

Noise Signals: The noise signals are also significant to guarantee the integrity of simulated signals, shown as follows:

$$n(t) \sim \mathcal{N}\left(0, \sigma^2\right) \tag{17}$$

where the nature of $\mathbf{n}(t)$ is Gaussian white noise, with mean value 0 and standard deviation value σ.

Indicators To measure various interferences, three indicators, signal-to-noise ratio (SNR), signal-to-harmonic ratio (SHR), and signal-to-interference ratio (SIR), were introduced, respectively:

$$\mathrm{SNR} = 10\log_{10}\left(\frac{\|\mathbf{x}\|_2^2}{\|\mathbf{n}\|_2^2}\right) \tag{18}$$

$$\mathrm{SHR} = 10\log_{10}\left(\frac{\|\mathbf{x}\|_2^2}{\|\mathbf{h}\|_2^2}\right) \tag{19}$$

$$\mathrm{SIR} = 10\log_{10}\left(\frac{\|\mathbf{x}\|_2^2}{\|\mathbf{y} - \mathbf{x}\|_2^2}\right) \tag{20}$$

where $\mathbf{x}$, $\mathbf{h}$, $\mathbf{n}$ and $\mathbf{y}$ represent fault siganls, noise signals, harmonic signals, and composited signals, respectively. The relationship among them is $\mathbf{y} = \mathbf{x} + \mathbf{h} + \mathbf{n}$.

Additionally, to evaluate the denoising performance, kurtosis value is introduced by following[40]. Kurtosis measures the flatness of data distribution, and it is the statistics of the degree of steepness of data distribution morphology. The large kurtosis values often denote good

denoising results.

$$kur = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(x(i)-\bar{x}\right)^4}{\left(\frac{1}{n}\sum_{i=1}^{n}\left(x(i)-\bar{x}\right)^2\right)^2} \tag{21}$$

where kur is kurtosis value. $\bar{x}$ is mean value and n is the number of sampling points.

Noise Suppression Performance

The collected raw signals from real world are often mingled with random noise signals, indicating that noise suppression is necessary for a fault diagnosis method.

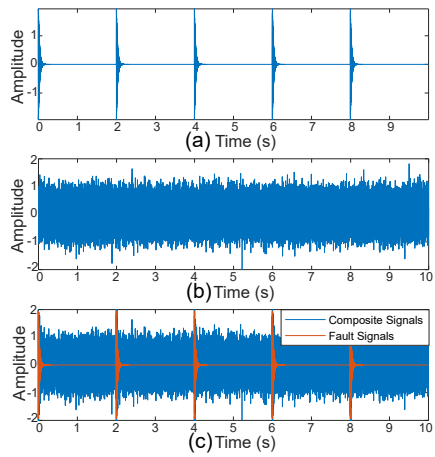

Figure 4. Constructing simulation signals for noise suppression test(a) simulated fault signals with kurtosis of 101.78, (b) simulated noise signals with kurtosis of 2.98, (c) composited signals with SNR -10.01 dB, kurtosis of 3.77.

To evaluate noise suppression, a comparison is made between the classical SK method and the proposed BATCN method. The simulated signal to be tested comprises both noise and fault signals, with the following parameter settings. M, as the impulse interval, is set at 2. τ_k, the fine random interval bias among impulses, is fixed at 0. f_n is resonance frequency, set as 1800. The damping ratio ξ is set at 0.003. Fig. 4 (a) depicts the simulated signal, where $a_k = 2$, and the values of h_a and h_b are fixed at 0. Fig. 4 (b) shows the simulated noise signals with Gaussian white noises. Fig. 4 (c) presents the synthetic relationship of the newly generated simulation signal that needs to be denoised.

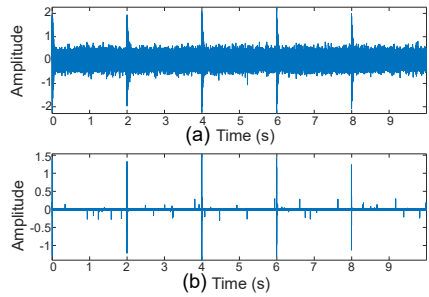

Figure 5. Noise suppression evaluation (a) SK filtered signals with kurtosis of 10.25, (b) BATCN filtered signals with kurtosis of 1363.46.

Fig. 5 shows the comparison results of noise suppression. The SK denoised results in Fig. 5 (a) show that the profile of the fault signals can be seen in the time domain. When compared to composite signals, the kurtosis increased from 3.77 to 10.25, indicating that the classical SK method is effective in eliminating general noise but low amplitude noise redundancy persists. The BATCN denoised results in Fig. 5 (b) show that the noise can be almost fully reduced and the filtered signals have a large kurtosis of 1363.46. The results from this simulated example shows that the BATCN method has better noise suppression ability than SK method.

Harmonic Interference Suppression Performance

Harmonic interference from rotor rotating and gear meshing may make difficult to recognize fault characteristic frequency. To examine harmonic interference suppression, SK method is also used to compare the proposed BATCN method.

The simulation parameters are set as follows: $M = 2$, $\tau_k = 0$, $f_n = 1800$, $\xi = 0.003$, $a_k = 2$, $h_a = h_b = 1$. Fig. 6 illustrates that the raw simulated signal, without any processing methods, is significantly affected by harmonic interference frequencies (HF) and their modulated frequencies (MF). Based on Equation (16), it can be inferred that h_1 generates harmonic interference frequencies (HF1) at 1000 Hz and their modulated frequencies (MF1) at 180 Hz, while h_2 generates HF2 at 2000 Hz and MF2 at 360 Hz. Fig. 6 (b) clearly shows that the two harmonic interference frequencies, HF1 and HF2, are prominently dominant in the frequency spectrum. By using the Hilbert envelope technique, it can be observed from Fig. 6 (c) that the HF2 frequency is attenuated, while the

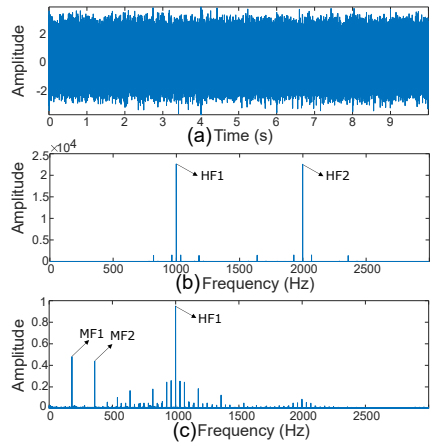

Figure 6. Harmonic interference suppression evaluation without any denoising method (a) time domain, (b) frequency domain, (c) Hilbert envelope spectrum.

primary frequencies become MF1, MF2, and HF1. These phenomena can significantly contaminate the frequency spectrum, making it extremely challenging to extract the real fault frequency.

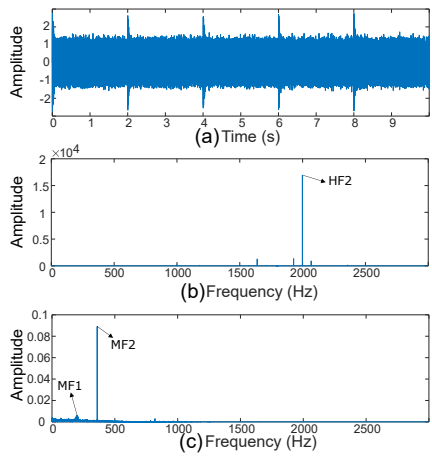

Figure 7. Harmonic interference suppression evaluation with SK method (a) time domain, (b) frequency domain, (c) Hilbert envelope spectrum.

Fig. 7 actually shows three sub-figures, labeled as (a), (b), and (c), respectively. Fig. 7 (a) shows the time-domain signal after the processing of the SK method, where the profile of fault signals can be observed.

Fig. 7 (b) shows the frequency spectrum after the SK method, where the dominant frequency component is still the harmonic interference frequency HF2. Fig. 7 (c) shows the frequency spectrum after the Hilbert envelope, where the modulation frequencies MF1 and MF2 are still obvious.

In Fig. 8 (b), the RF can be observed in frequency domain after using the proposed BATCN method. Additionally, all the HF, MF and RF are submerged in Hilbert envelop spectrum, as shown in Fig. 8 (c). Meanwhile, the fault frequency (FF) can be observed clearly through enlargement. Through comparisons between SK and BATCN results, it can be concluded that BATCN can effectively extract fault signals and detect fault frequency due to the significant advantage of harmonic interference suppression.

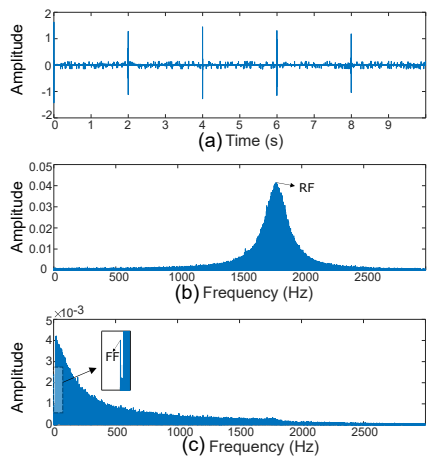

Figure 8. Harmonic interference suppression evaluation with BATCN method (a) time domain, (b) frequency domain, (c) Hilbert envelope spectrum.

Physical Experiments

This section uses three physical cases to test the proposed BATCN method. Data in case 1 is from a public dataset MFPT*. Data in case 2 is collected in a lab environment from wind turbine pitch bearings that have been serviced for 15 years in real wind farms. Data in case 3 is obtained

*[Online]. Available: https://github.com/mathworks/RollingElementBearingFaultDiagnosis-Data

from a field wind turbine in a real-world wind farm but is known to work abnormally.

Case 1

This case here is related to an open-source dataset, MFPT, that is collected in general bearing with sampling frequency 48828 Hz. MFPT dataset contains two fault types, inner fault (case 1.1) and outer fault (case 1.2). The corresponding fault frequencies (FF) of the inner fault and outer fault are named BPFI and BPFO, respectively. We first test the proposed BATCN method in MFPT general bearing (case 1), and then method validation for real wind turbine pitch bearing will be subsequently executed in case 2 and case 3. Fig. 9 is the analysis results for normal bearing. From Fig. 9 (d) or (e), it can observe that we can not identify any fault frequency (FF). Fig. 10 and Fig. 11 represents the damaged conditions, with inner fault and outer fault, respectively.

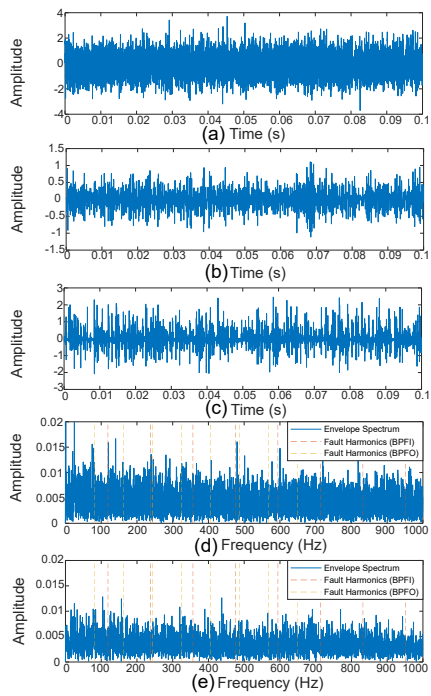

Figure 9. MFPT normal bearing (a) raw signals, (b) SK denoised signals, (c) BATCN denoised signals, (d) Hilbert envelope spectrum based on SK denoised signals, (e) Hilbert envelope spectrum based on BATCN denoised signals.

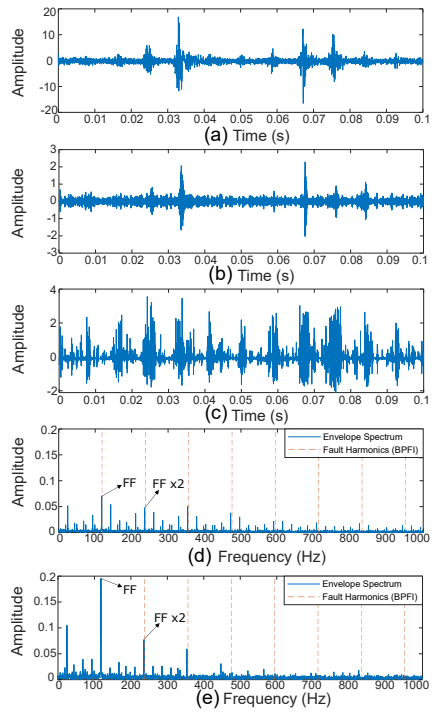

Figure 10. MFPT inner fault bearing, case 1.1 (a) raw signals, (b) SK denoised signals, (c) BATCN denoised signals, (d) Hilbert envelope spectrum based on SK denoised signals, (e) Hilbert envelope spectrum based on BATCN denoised signals.

As can be seen in Fig. 10, when using SK method to filter the raw signal with inner fault, the Hilbert envelope spectrum Fig. 10 (d) may still exist some interference frequency. However, the BATCN filtered signal can clearly identify FF in its Hilbert envelope spectrum Fig. 10 (e). In addition, using the BATCN method, higher fault harmonic components in the envelope spectrum seem to decay quickly.

By observing Fig. 11, we can draw similar conclusions to those of Fig. 10. One notable point in these conclusions is that the higher fault harmonic components present in the envelope spectrum also exhibit a rapid reduction. This phenomenon is more significant for slow-speed bearings, such as wind turbine pitch bearings, because the higher fault harmonics in the slow-speed condition may overlap with other types of FF and make it challenging to identify the correct FF.

To quantitatively compare the performance between SK and BATCN, we can define amplitude of corresponding frequency in Hilbert envelope

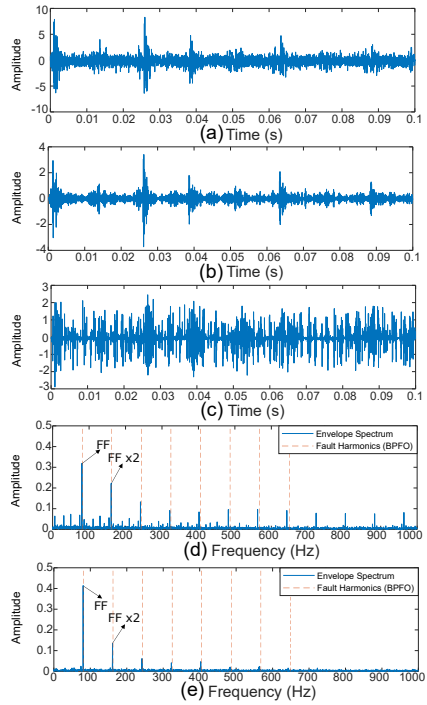

Figure 11. MFPT outer fault bearing, case 1.2 (a) raw signals, (b) SK denoised signals, (c) BATCN denoised signals, (d) Hilbert envelope spectrum based on SK denoised signals, (e) Hilbert envelope spectrum based on BATCN denoised signals.

Table 1. Case 1: quantitative results of signals processed by SK and BATCN.

	FF*1 AMP	FF*2 AMP	FF*2 / FF*1 AMP	FF*5 / FF*1 AMP
Case 1.1 (SK)	0.0704	0.0475	0.6747	0.3049
Case 1.1 (BATCN)	0.1957	0.0769	0.3933	0.0603
Case 1.2 (SK)	0.3186	0.2221	0.6971	0.2646
Case 1.2 (BATCN)	0.4140	0.1342	0.3242	0.1123

spectrum as AMP. FF*1 is fault frequency, FF*2 and FF*5 represent fault frequency times 2 and 5, respectively. As shown in Tab. 1, BATCN has the larger amplitude at FF*1 frequency than SK. Regarding the large FF*2 / FF*1 and FF*5 / FF*1 for BATCN, indicating that the fault harmonics

decay more rapidly in the Hilbert envelope spectrum of the BATCN method, which is advantageous for fault diagnosis.

Case 2

The data, in this case, is collected from the industrial scale wind turbine bearing lab at the University of Manchester. The bearings have been operated in a real-world wind farm for over 15 years and have naturally incurred defects. The bearings have a diameter of 1m and and a mass of 261 kg. The sampling frequency is initially chosen as 100 kHz. This high sampling frequency produces significant computational demands on the proposed method. Therefore, the collected data is downsampled to 1 kHz. In this case, this proposed method only needs low computations while maintaining good filtering and fault diagnosis abilities. The case 2.1 is executed without loadings. The case 2.2 is operated with loadings, which may result in a little bit rotating speed fluctuation. The left and front view of the test rig are shown in Fig. 12 (a) and Fig. 12 (b), respectively. Tab. 2 describes the three types of fault frequencies (FF) of the wind turbine pitch bearing, and the FF in this case is the inner fault frequency, namely BPFI.

On top of the SK method, the temporal convolutional network (TCN) method is also added to compare with our proposed BATCN method. Fig. 13 and 14 compare different methods in the time and frequency domains for case 2.1, while Fig. 15 and 16 do the same for case 2.2. Observing the magnified time-domain plot in Fig. 13 and 15, it can be observed that, compared to other methods, BATCN demonstrates a better ability to extract fault signals. Observing the frequency-domain results in Fig. 14 and Fig. 16, it can be found that compared to other methods, BATCN can better extract the characteristic frequencies of the FF and double FF (FF *2), and the FF is more obvious compared to other orders of FF.

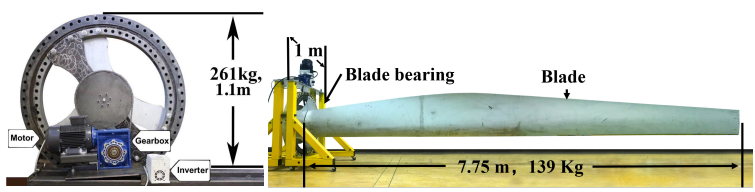

Figure 12. View of test-rig.

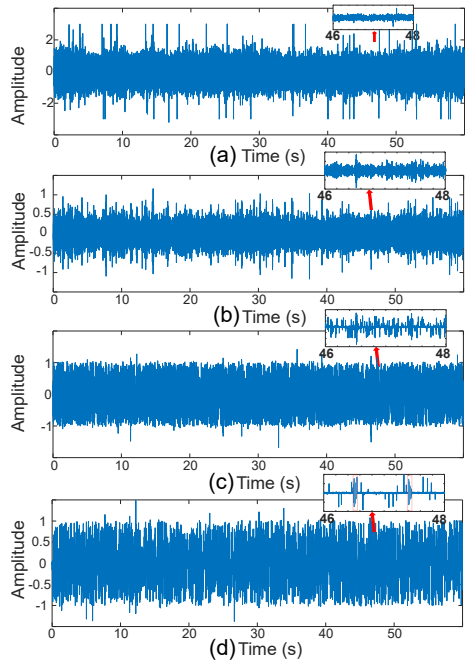

Figure 13. Laboratorial wind turbine pitch bearing fault, case 2.1 (a) raw signals, (b) SK denoised signals, (c) TCN denoised signals, (d) BATCN denoised signals.

To quantitatively compare the performance between SK and BATCN, we can define amplitude of corresponding frequency in Hilbert envelope spectrum as AMP. FF*1 is fault frequency, FF*2 and FF*5 are fault frequency times 2 and 5, respectively. As shown in Tab. 3, the BATCN can obtain larger AMP of FF*1 and FF*2 than SK method. In addition, the AMP of FF*2/FF*1 and FF*5/FF*1 are both smaller than 1 in BATCN. These findings denote that BATCN is more capable of obtaining a spectrum that can capture the fault frequencies.

Table 2. Theoretical FF of the wind turbine pitch bearing.

Case	Bearing Speed (r/min)	BPFI (Hz)	BPFO (Hz)	BSP (Hz)
Case 2.1	2.145	1.109	1.035	0.331
Case 2.2	1.658	0.857	0.800	0.256

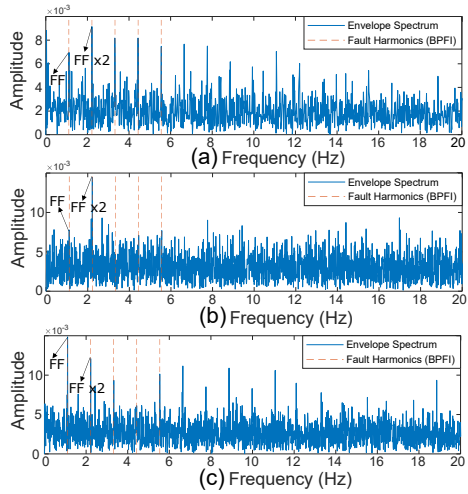

Figure 14. Laboratorial wind turbine pitch bearing fault, case 2.1 (a) Hilbert envelope spectrum based on SK denoised signals, (b) Hilbert envelope spectrum based on TCN denoised signals, (c) Hilbert envelope spectrum based on BATCN denoised signals.

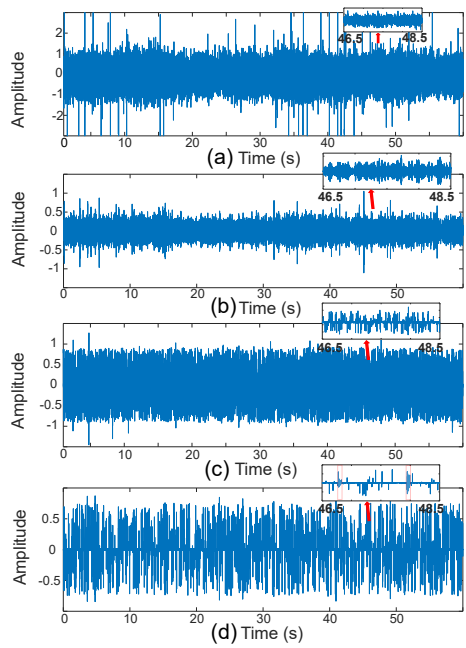

Figure 15. Laboratorial wind turbine pitch bearing fault, case 2.2 (a) raw signals, (b) SK denoised signals, (c) TCN denoised signals, (d) BATCN denoised signals.

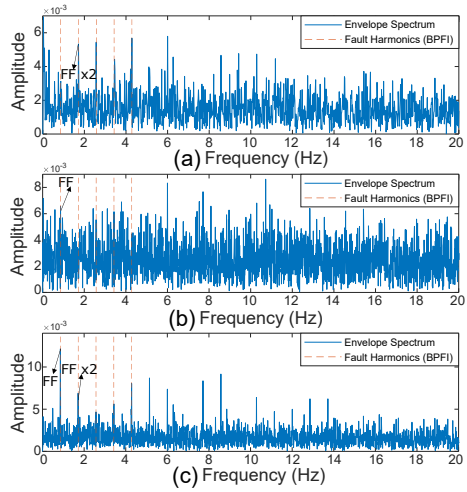

Figure 16. Laboratorial wind turbine pitch bearing fault, case 2.2 (a) Hilbert envelope spectrum based on SK denoised signals, (b) Hilbert envelope spectrum based on TCN denoised signals, (c) Hilbert envelope spectrum based on BATCN denoised signals.

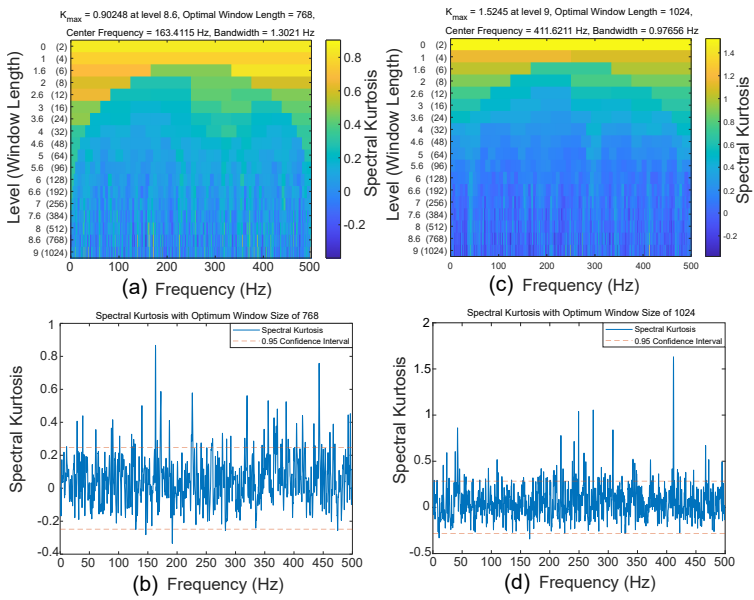

Figure 17. Visible intermediate processing results (a) fast kurtogram results, (b) spectral kurtosis with best window size.

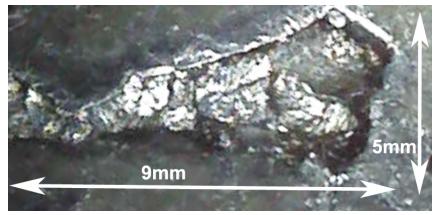

Figure 18. Inner race defect through endoscope.

Table 3. Case 2: quantitative results of signals processed by SK and BATCN.

	FF*1 AMP	FF*2 AMP	FF*2 / FF*1 AMP	FF*5 / FF*1 AMP
Case 2.1 (SK)	0.0069	0.0091	1.3188	1.0724
Case 2.1 (BATCN)	0.0156	0.0135	0.8653	0.8205
Case 2.2 (SK)	0.0031	0.0053	1.7096	1.8064
Case 2.2 (BATCN)	0.0121	0.0068	0.5619	0.6611

In addition, Fig. 17 presents the visible intermediate processing results of SK method, which could explain why SK method perform less satisfactory in this case. As can be seen in Fig. 17 (a) and Fig. 17 (c) the best window size 768 and 1024 are determined by fast kurtogram algorithm, respectively. Subsequently, using this window size to calculate corresponding spectral kurtosis Fig. 17 (b) and Fig. 17 (d). We can observe that spectral kurtosis is uniformly distributed, and values around the best center frequency are not all higher than the other frequencies. This denotes that SK method may be difficult to distinguish the FF from other frequencies.

To inspect the inner conditions of pitch bearings of the industrial-scale wind turbine, an electronic endoscope is applied in this case. Fig.18 from endoscope indicates that the defect truly exists at the bearing inner race, measured length 9 mm and width 5 mm. Furthermore, no obvious damage is found in the bearing balls or outer race. This inspection result denotes that the aforementioned diagnostic results are convincing. As a result, the proposed BATCN method in this study might be beneficial for fault diagnosis of wind turbine pitch bearings, demonstrating that this method has a wide range of applications in a natural industrial occasion.

Case 3

The data in case 3 is also collected with a sampling frequency of 100k Hz and then downsampled to 1k Hz from a real wind turbine pitch bearing, which is working in a real world wind farm but working abnormally. The speed of wind turbine pitch bearing is estimated to range from 1.009 r/min to 1.075 r/min (average speed: 1.042 r/min). Then, substituting into inherent parameters and average speed, the theoretical fault frequencies can be calculated: BPFI 0.535 Hz, BPFO 0.503 Hz and BSP 0.161 Hz.

Fig.19 shows the actual wind turbine pitch bearing in field test. The characteristic of this case is that the data was collected during the operation of the wind turbine, which means that the collected raw data may have a certain degree of disorderliness, as depicted in Fig. 20 (a). This disorderliness is usually caused by the reciprocating motion of the bearing[41]. To eliminate the influence of reciprocating motion, the collected data is spliced together to obtain an ordered signal, as depicted in Fig. 20 (b).

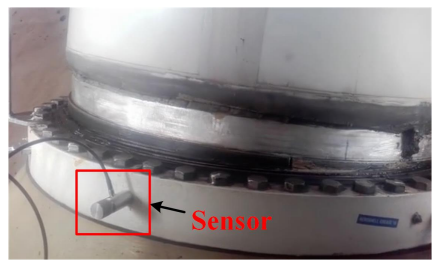

Figure 19. The actual wind turbine pitch bearing in field test.

Table 4. Case 3: quantitative results of signals processed by SK and BATCN.

	FF*1 AMP	FF*2 AMP	FF*2 / FF*1 AMP	FF*5 / FF*1 AMP
Case 3 (SK)	0.0010	0.0010	1	1.6
Case 3 (BATCN)	0.072	0.0041	0.5694	0.5555

As shown in Fig. 21, the proposed BATCN method can identify fault frequency (FF) more clearly than the SK method. It is noted that the FF marked in Fig. 21 (c) is 0.530 Hz. By comparing the theoretical fault

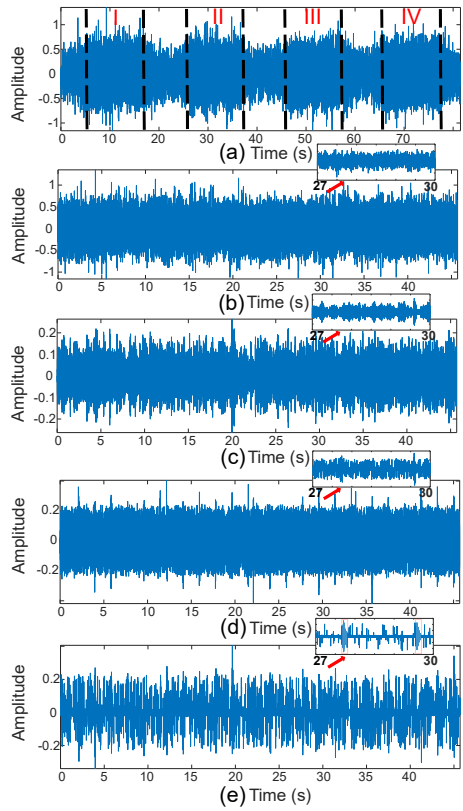

Figure 20. Field wind turbine pitch bearing, case 3 (a) raw signals under the reciprocating operation condition, (b) recombined signals with the four parts in (a), (c) SK denoised signals, (d) TCN denoised signals, (e) BATCN denoised signals.

frequencies, we can find that the fault frequency match BPFI, so we can infer that this bearing may exist the inner fault. Similar to the Tab. 1 and Tab. 3, the Tab. 4 also demonstrates that the BATCN method is more effective in capturing the fault frequencies in the spectrum.

Meta-analysis

This section would show a comprehensive analysis on our proposed method. To automatically reflect the damage information from spectra, an indicator I[41] is defined as follows:

$$I = (|F_d - F_f|/F_f) \times 100\% \tag{22}$$

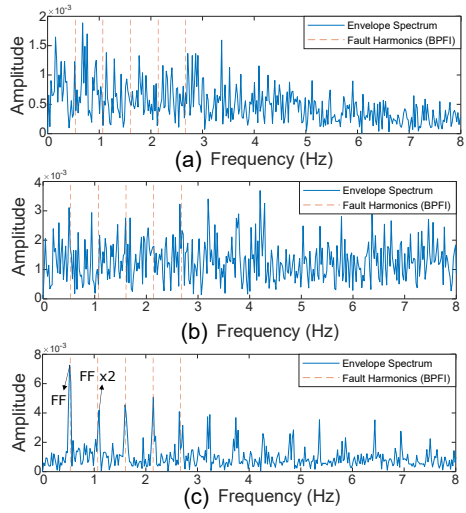

Figure 21. Field wind turbine pitch bearing, case 3 (a) Hilbert envelope spectrum based on SK denoised signals, (b) Hilbert envelope spectrum based on TCN denoised signals, (c) Hilbert envelope spectrum based on BATCN denoised signals.

where F_d denotes the dominant frequencies for each test and F_f can be selected as BPFI, BPFO and BSP. I represents the error between the identified dominant frequency and the theoretical defect frequency. As shown in Tab. 5 , the indicator evaluation for cases 1,2 and 3 are executed. It is found that the indicator I in all cases could detect the corresponding fault. For example, the bearing most likely has an inner race fault in case 1.1, because it has the smallest I. Additionally, in the case of real fault types, BATCN consistently yielded lower values for the indicator I compared to SK, indicating improved accuracy in identifying the dominant frequency associated with the theoretical defect frequency.

Table 5. Indicator evaluation for cases 1, 2, and 3.

Case	Inner race fault	Outer race fault	Ball fault
Case 1.1 (BATCN/SK)	**0.06**%/0.73%	46.44%/45.45%	-
Case 1.2 (BATCN/SK)	32.42%/31.77%	0.97%/**0.03**%	-
Case 2.1 (BATCN/SK)	**0.36**%/0.81%	6.76%/6.28%	233.8%/232.3%
Case 2.2 (BATCN/SK)	**0.23**%/12.25%	6.875%/9.37%	233.9%/183.2%
Case 3 (BATCN/SK)	**0.18**%/0.37%	6.875%/6.16%	231.6%/231.0%

To further evaluate our proposed method, we conducted extensive experiments under laboratory conditions to simulate different operating speeds. As shown in Tab. 6, the bearing defect frequencies were successfully identified using the proposed method and all of them matched the theoretical inner race defect frequencies. In addition, the BATCN method is more capable of detecting the correct fault frequency than the SK method, as evidenced by the indicator I. However, it is noteworthy that in cases where the bearing inner race is severely damaged, the condition of the balls and outer race may also be affected. In the present study, a comprehensive inspection of the bearing was conducted using an endoscope, which revealed no significant damage to the balls and outer race except for a few minor dents. Compared to the visible damage on the inner race, the damage on the balls and outer race is insignificant.

In our experiments, the SK method could also detect the fault frequency in some cases, such as results in Fig. 14. Unsatisfactory results may be caused by insufficient data length. One characteristic of the SK method is that, as a frequency method, the detection performance improves as the number of repetitions of the fault signal increases. It means that the data length may need to be ensured to obtain multiple repetition of fault signals. This paper could demonstrate that the BATCN is capable of working effectively with limited data length, while the classical SK method is not.

Table 6. Extensive experiments.

Rotated speed (rpm)	1.05	1.35	2.05	2.25	3.05	3.15
Identified dominant frequency (Hz) BATCN/SK	0.545/0.582	0.698/0.719	1.064/1.099	1.1635/1.1845	1.5816/1.59	1.629/1.677
Inner race I BATCN/SK	**0.45**%/7.37%	**0.10**%/3.02%	**0.43**%/3.74%	**0.02**%/1.82%	**0.3**%/1.08%	**0.08**%/3.02%
Outer race I BATCN/SK	7.63%/15.05%	7.26%/10.39%	7.61%/11.15%	7.17%9.11%	7.47%/8.30%	7.24%/10.37%
Ball I BATCN/SK	236.5%/259.7%	235.3%/245.1%	236.4%/247.5%	235.1%/241.6%	236.0%/238.6%	235.3%/245.1%

Conclusion

A signal denoising method, Bayesian augmented temporal convolutional networks (BATCN), is proposed for wind turbine pitch bearing defect detection with slow speed. The BATCN method can learn the intrinsic features of fault signals to guarantee the extraction of non-stationary fault signals and realize noise suppression and harmonic interference suppression during the process of signal denoising, with deep learning techniques. In addition, this new method is able to spontaneously find the

best patch length that is a significant performance influence factor under the Bayesian framework. The effectiveness of the new method has been extensively validated on simulation examples and three real cases (open-source data, lab data, and field data). The comprehensive results show the BATCN method is effective in detecting faults for slow speed wind turbine pitch bearings due to its superior filtering capacity and it also outperforms the popular SK method.

References

1. Dhanda KK and Hartman LP. The ethics of carbon neutrality: A critical examination of voluntary carbon offset providers. *Journal of Business Ethics* 2011; 100(1): 119–149.
2. Karin O. Gwec global wind report 2019. Technical report, Council, Global Wind Energy, 2017.
3. Zhang C, Liu Z and Zhang L. Wind turbine blade bearing fault detection with bayesian and adaptive kalman augmented lagrangian algorithm. *Renewable Energy* 2022; 199: 1016–1023.
4. Yan R, Gao RX and Chen X. Wavelets for fault diagnosis of rotary machines: A review with applications. *Signal processing* 2014; 96: 1–15.
5. Chen L, Zhang Y and Xia X. Contact stress and deformation of blade bearing in wind turbine. In *2010 International Conference on Measuring Technology and Mechatronics Automation*, volume 1. IEEE, pp. 833–836.
6. Liu R, Yang B and Hauptmann AG. Simultaneous bearing fault recognition and remaining useful life prediction using joint-loss convolutional neural network. *IEEE Transactions on Industrial Informatics* 2019; 16(1): 87–96.
7. Poore R. 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. Technical report, National Renewable Energy Lab.(NREL), Golden, CO (United States), 2008.
8. García Márquez FP, Bernalte Sánchez PJ and Segovia Ramírez I. Acoustic inspection system with unmanned aerial vehicles for wind turbines structure health monitoring. *Structural Health Monitoring* 2022; 21(2): 485–500.
9. Sun C, Zhang Z, Luo X et al. Support vector machine-based grassmann manifold distance for health monitoring of viscoelastic sandwich structure with material ageing. *Journal of Sound and Vibration* 2016; 368: 249–263.
10. Liu Z, Zhang L and Carrasco J. Vibration analysis for large-scale wind turbine blade bearing fault detection with an empirical wavelet thresholding method. *Renewable Energy* 2020; 146: 99–110.
11. Elasha F, Greaves M and Mba D. Planetary bearing defect detection in a commercial helicopter main gearbox with vibration and acoustic emission. *Structural Health Monitoring* 2018; 17(5): 1192–1212.
12. Randall RB and Antoni J. Rolling element bearing diagnostics—a tutorial. *Mechanical systems and signal processing* 2011; 25(2): 485–520.
13. Zhao Z, Qiao B, Wang S et al. A weighted multi-scale dictionary learning model and its applications on bearing fault diagnosis. *Journal of Sound and Vibration* 2019; 446: 429–452.

14. Pan Y, Hong R, Chen J et al. Performance degradation assessment of wind turbine gearbox based on maximum mean discrepancy and multi-sensor transfer learning. *Structural Health Monitoring* 2021; 20(1): 118–138.
15. Antoni J and Randall R. Unsupervised noise cancellation for vibration signals: part i—evaluation of adaptive algorithms. *Mechanical Systems and Signal Processing* 2004; 18(1): 89–101.
16. Antoni J and Randall R. Unsupervised noise cancellation for vibration signals: part ii—a novel frequency-domain algorithm. *Mechanical Systems and Signal Processing* 2004; 18(1): 103–117.
17. Antoni J. The spectral kurtosis: a useful tool for characterising non-stationary signals. *Mechanical systems and signal processing* 2006; 20(2): 282–307.
18. McDonald GL and Zhao Q. Multipoint optimal minimum entropy deconvolution and convolution fix: Application to vibration fault detection. *Mechanical Systems and Signal Processing* 2017; 82: 461–477.
19. Liu Z, Tang X, Wang X et al. Wind turbine blade bearing fault diagnosis under fluctuating speed operations via bayesian augmented lagrangian analysis. *IEEE Transactions on Industrial Informatics* 2020; 17(7): 4613–4623.
20. He G, Ding K and Lin H. Fault feature extraction of rolling element bearings using sparse representation. *Journal of Sound and Vibration* 2016; 366: 514–527.
21. Aye SA, Heyns PS and Thiart CJ. Fault detection of slow speed bearings using an integrated approach. *IFAC-PapersOnLine* 2015; 48(3): 1779–1784.
22. Guo Y, Zhao Z, Sun R et al. Data-driven multiscale sparse representation for bearing fault diagnosis in wind turbine. *Wind Energy* 2019; 22(4): 587–604.
23. Li X, Shao H, Lu S et al. Highly efficient fault diagnosis of rotating machinery under time-varying speeds using lsismm and small infrared thermal images. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 2022; 52(12): 7328–7340.
24. Han H, Wang H, Liu Z et al. Intelligent vibration signal denoising method based on non-local fully convolutional neural network for rolling bearings. *ISA transactions* 2022; 122: 13–23.
25. Zhang D, Chen Y, Guo F et al. A new interpretable learning method for fault diagnosis of rolling bearings. *IEEE Transactions on Instrumentation and Measurement* 2020; 70: 1–10.
26. Chen Z, Gryllias K and Li W. Intelligent fault diagnosis for rotary machinery using transferable convolutional neural network. *IEEE Transactions on Industrial Informatics* 2019; 16(1): 339–349.
27. Li T, Zhao Z, Sun C et al. Waveletkernelnet: An interpretable deep neural network for industrial intelligent diagnosis. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 2021; 52(4): 2302–2312.
28. Kumar A, Vashishtha G, Gandhi C et al. Novel convolutional neural network (ncnn) for the diagnosis of bearing defects in rotary machinery. *IEEE Transactions on Instrumentation and Measurement* 2021; 70: 1–10.
29. Wang H, Liu Z, Peng D et al. Attention-guided joint learning cnn with noise robustness for bearing fault diagnosis and vibration signal denoising. *ISA transactions* 2022; 128: 470–484.
30. Li W, Shang Z, Qian S et al. A novel intelligent fault diagnosis method of rotating machinery based on signal-to-image mapping and deep gabor convolutional adaptive pooling network. *Expert Systems with Applications* 2022; 205: 117716.

31. Pei X, Zheng X and Wu J. Rotating machinery fault diagnosis through a transformer convolution network subjected to transfer learning. *IEEE Transactions on Instrumentation and Measurement* 2021; 70: 1–11.
32. Chen Z, Guo L, Gao H et al. A fault pulse extraction and feature enhancement method for bearing fault diagnosis. *Measurement* 2021; 182: 109718.
33. Long J, Shelhamer E and Darrell T. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 3431–3440.
34. Bai S, Kolter JZ and Koltun V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling, 2018. arXiv:1803.01271.
35. He K, Zhang X, Ren S et al. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 770–778.
36. Martinez-Cantin R, De Freitas N, Brochu E et al. A bayesian exploration-exploitation approach for optimal online sensing and planning with a visually guided mobile robot. *Autonomous Robots* 2009; 27(2): 93–103.
37. Brochu E, Brochu T and De Freitas N. A bayesian interactive optimization approach to procedural animation design. In *Proceedings of the 2010 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. pp. 103–112.
38. Press WH, Teukolsky SA, Vetterling WT et al. *Numerical Recipes with Source Code CD-ROM 3rd Edition: The Art of Scientific Computing*. Cambridge University Press, 2007.
39. Huang W, Gao G, Li N et al. Time-frequency squeezing and generalized demodulation combined for variable speed bearing fault diagnosis. *IEEE Transactions on Instrumentation and Measurement* 2018; 68(8): 2819–2829.
40. Wang J, Qiao W and Qu L. Wind turbine bearing fault diagnosis based on sparse representation of condition monitoring signals. *IEEE Transactions on Industry Applications* 2018; 55(2): 1844–1852.
41. Liu Z and Zhang L. Naturally damaged wind turbine blade bearing fault detection using novel iterative nonlinear filter and morphological analysis. *IEEE Transactions on Industrial Electronics* 2019; 67(10): 8713–8722.