Measurement 137 (2019) 287–301

Contents lists available at ScienceDirect

Measurement

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

Detection and identification of windmill bearing faults using a one-class support vector machine (SVM)

Juhamatti Saari [a,b,*], Daniel Strömbergsson [a,c], Jan Lundberg [b], Allan Thomson [d]

[a] SKF-LTU University Technology Centre, Luleå University of Technology, SE-97187 Luleå, Sweden
[b] Division of Operation, Maintenance and Acoustics, Luleå University of Technology, SE-97187 Luleå, Sweden
[c] Division of Machine Elements, Luleå University of Technology, SE-97187 Luleå, Sweden
[d] SKF (U.K), Industrial Digitalisation & Solutions, Livingston EH54 7DP, Scotland, United Kingdom

ARTICLE INFO

Article history:
Received 20 December 2017
Received in revised form 13 November 2018
Accepted 11 January 2019
Available online 14 January 2019

Keywords:
Novelty detection
Wind turbine
Bearing fault diagnostics

ABSTRACT

The maintenance cost of wind turbines needs to be minimized in order to keep their competitiveness and, therefore, effective maintenance strategies are important. The remote location of wind farms has led to an opportunistic maintenance strategy where maintenance actions are postponed until they can be handled simultaneously, once the optimal opportunity has arrived. For this reason, early fault detection and identification are important, but should not lead to a situation where false alarms occur on a regular basis. The goal of the study presented in this paper was to detect and identify wind turbine bearing faults by using fault-specific features extracted from vibration signals. Automatic identification was achieved by training models by using these features as an input for a one-class support vector machine. Detection models with different sensitivity were trained in parallel by changing the model tuning parameters. Efforts were also made to find a procedure for selecting the model tuning parameters by first defining the criticality of the system and using it when estimating how accurate the detection model should be. Method was able to detect the fault earlier than using traditional methods without any false alarms. Optimal combination of features and model tuning parameters was not achieved, which could identify the fault location without using any additional techniques.

1. Introduction

Wind energy is an important source of clean and renewable energy. In order to keep its competitiveness, the maintenance cost needs to be minimized and, therefore, effective maintenance strategies need to be adopted. The remote location of wind farms has led to an opportunistic maintenance strategy where component degradations are monitored over a long period and any non-critical maintenance actions are postponed until they can be handled simultaneously, once the optimal opportunity has arrived [6]. For this reason, early fault detection and identification are important, especially for the rotational structures of wind energy converters (WECs) [5].

Two approaches commonly used for WEC fault detection are to look for inconsistencies in the process data which indicate the presence of faults [19,12] and to measure vibrations in the converters [17,1]. Vibration analysis is especially applicable to monitor the gears and bearings in the gearbox, the bearings in the generator and the main bearing [5]. Vibration analysis is carried out using spectral analysis and involves a process where, based on knowledge of mechanics, defect frequencies are isolated which can give an early indication of certain faults. Furthermore, envelope analysis (a.k.a the high-frequency resonance technique) is found to be the foremost technique [11], since it is able to detect defect frequencies even if they are masked by vibration generated by structural elements and other machine elements. When applying this technique, each time a defect in a rolling element bearing enters the rolling element/raceway contact under load, it will cause an impact which usually excites a resonance in the system at a high frequency (above 10 kHz), and thus the defect in question is easier to detect using the developed demodulation technique [9].

After relevant features are extracted (for instance the peak of the defect frequencies), one must define an appropriate threshold for the alarm in order to know when the incipient fault has reached the limit when actions are necessary for retaining or restoring the original function of the component in question (i.e. trend analysis). A common method for setting the threshold is to monitor the values under nominal conditions, to define the threshold manually by

* Corresponding author at: SKF-LTU University Technology Centre, Luleå University of Technology, SE-97187 Luleå, Sweden.
E-mail address: juhamatti.saari@ltu.se (J. Saari).

https://doi.org/10.1016/j.measurement.2019.01.020
0263-2241/

assuming the data points to have a Gaussian distribution, and to use the standard deviation together with the "68–95–99.7" rule. This is a fast and rough way to find the anomalies caused by faults in the system. However, this technique has some limitations. Firstly, it can only be applied for one feature at a time and cannot spot the small changes which may together be considered as the fault anomaly, since the features are not compared with each other, and since several features are merely combined and summed together. This may lead to a situation where different features counterbalance each other. Secondly, it is difficult to define which sigma range should be used for selecting the confidence level and the number of outliers which should be allowed; this requires a manual effort.

In order to avoid these problems, machine learning techniques have been investigated where multiple key indicators can be used as an input. Using supervised methods, a fault detection threshold is set by using existing historical data, which are divided into two groups labelled nominal condition data and faulty condition data. These data are used later for optimizing the boundary between these two classes. However, the drawback of machine learning techniques is the need to have failure data covering all the possible failure modes. For windmill WECs, this problem is sometimes avoided by training fault detection algorithms using data collected during failures observed in other similar turbines. However, applying this approach, the working environment and data collection are considered to be similar, which may not be a totally realistic assumption.

According to Hameed et al. [5], an Aegis PR pattern recognition tool has been developed which is capable of detecting wind turbine blade faults using unsupervised techniques. However, in the review presented in Hameed et al. [5], no mention is made of using similar techniques for WEC gearbox faults. In unsupervised techniques, the threshold is set using nominal data collected before any faults have been observed. There is no requisite for failure data in such techniques and they can work with multiple input features. One unsupervised technique tested for real machine fault monitoring problems is the kernel-based one-class support vector machine (SVM) [14,15,4]. Shin et al. [15] used a one-class SVM for detecting several machine faults and concluded it to be superior to the compared artificial neural network methods. However, the one-class SVM was shown to be sensitive to the selected parameters and, therefore, follow-up studies are needed which investigate how to select appropriate parameters for one-class SVMs [15]. Fernández-Francos et al. [4] demonstrated how the one-class SVM was able to detect bearing defects by using energy-related features extracted by dividing the power spectrum of the raw vibration signals into several sub-bands. After an anomaly detection, an envelope analysis was performed where the classification of faults was carried out. In their experiment, accurate fault detection was achieved each time using the same parameter values, which may reduce the sensitivity, as stated by Shin et al. [15].

The aim of the present study was to detect and identify WEC bearing faults by extracting relevant features using envelope analysis, and then use these features as an input for the unsupervised one-class SVM. In order to diagnose the presence of a specific fault as early as possible, several detection models were trained in parallel by changing the input features together with the model tuning parameters. This was expected to lead to a situation where faults could be detected and identified by comparing the accuracy of the different models. Moreover, efforts were made to find a procedure for selecting the model tuning parameters by first defining the criticality of the system by estimating how accurate the detection model should be. The analyses were compared with analyses performed using traditional methods where the failure thresholds were set manually.

2. Measurement campaign and feature extraction

An SKF CMSS WIND-100-10 (100 mV/g) accelerometer was mounted in the axial direction on the gearbox housing close to the high-speed shaft (HSS) generator side (GS) bearing. Fig. 1 shows the schematics of the WEC, where the gearbox stages are seen. The annulus rings of both planetary stages were fixed and the output shaft is on the generator side. The accelerometer was mounted near the HSS shaft where bearing 1 and 2 are located (see Fig. 1). The HSS bearings, which were damaged during the measurement campaign, has been shown in statistical analyses to most often develop a fault of the gearbox bearings [7]. The failed bearing was a tapered roller bearing and the diameter was over 30 cm.

In total 518 measurements (one per day) were collected. The sample rate was 12.8 kHz and the measurement time per measurement was 1.28 s. This led to a frequency resolution of 0.7813 Hz for each bin. During the measurement campaign, the speed of the high-speed shaft varied from 702 to 1174 RPM. Moreover, a process parameter which indicates how much power the converter is producing was recorded (referred to as the feature P). The difference when dividing the highest value of P by the smallest one was 2.77. In order to trend the wind turbine degradation, key condition indicators (CI) were extracted (see Table 1) which are sensitive to certain faults. The defect frequency multipliers for bearing 1 (B1)

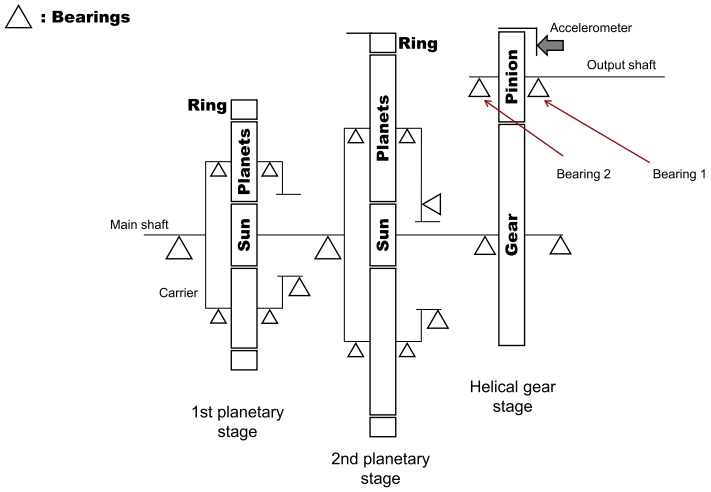


Fig. 1. Schematic figure of the WEC where the location of the accelerometer is shown.

Table 1
Explanation of uncommon features.

Feature	Explanation	Unit
BPFI	Sum of the BPFI and the absolute values of the± 3 shaft speed SBs	
BPFI 1th H	Sum of the 1th harmonic of the BPFI and the absolute values of the ±3 shaft speed SBs	g
BPFI 2th H	Sum of the 2th harmonic of the BPFI and the absolute values of the ±3 shaft speed SBs	g
BPFIrss	Sum of the absolute values of all the BPFI-related defect frequency components	g
*BPFOrss	Sum of the absolute value of all the BPFO-related defect frequency components	g
P	Process (produced power)	Watt

*The BPFO-related features are calculated similarly to how the BPFI-related features are calculated, without including any of the peak values of the shaft speed SBs.

Abbreviation	Meaning
BPFI	Ball pass frequency inner
BPFO	Ball pass frequency outer
SB	Sideband

were 11.37 (BPFI) and 8.63 (BPFO). For bearing 2 (B2), these multipliers were 9.25 (BPFI) and 6.75 (BPFO).

3. One-class support vector machine

The aim of novelty detection is to identify new or unknown data which a machine learning system is not aware of during training [8]. A novelty detection algorithm uses two main approaches for estimating the probability density function (PDF), namely the parametric approach (for known distributions) and the non-parametric approach (for unknown distributions) [3]. In this study we used a method called the one-class support vector machine (OCSVM, one-class SVM or nu-SVM), which was originally developed by Schölkopf et al. [14]. It is a non-parametric method where the PDF is estimated using the kernel method.

Since only data from one-class are available (nominal data in this case), the one-class SVM algorithm cannot maximize the margin between two classes similarly to what is done when using the regular SVM [14]. Instead, the goal is to develop an algorithm which returns a function that takes value 1 in a small region (a nominal region in this case) and -1 elsewhere (a faulty region). The strategy is to map the data into feature space F (using a known kernel function) and to separate them from the origin with the maximum margin, which is also known as the hyperplane. The objective function of a one-class SVM (Eq. 1) resembles that of the two-class SVM with some small differences. Instead of the cost function, in the one-class SVM it is the parameter $v \in [0,1]$ that characterizes the solution by solving a quadratic solution, where Schölkopf et al. [14]

$$\min_{\omega,\xi,\rho} \frac{\|\omega\|^2}{2} + \frac{1}{vn}\sum_{i=1}^{n}\xi_i - \rho \tag{1}$$

Subject to:

$$(\omega \cdot \Phi(x_i)) \geqslant \rho - \xi_i \quad \forall i \in \mathbb{N}$$

$$\xi_i \geqslant 0 \quad \forall i \in \mathbb{N},$$

where n is the number of instances, ρ is the offset parameter and ξ is the slack variable. ω and ρ are hyperplane parameters represented in the equation, $\omega^T x + \rho = 0$.

This quadratic optimization problem is solved using Lagrange multipliers, and the decision function rule for a datapoint x then becomes [14]

$$f(x) = sgn\left(\sum_{i=1}^{n}\alpha_i y_i K(x, x_i) + b\right), \tag{2}$$

where the coefficient $\alpha_i > 0$, b is a constant and K is the kernel function.

To ensure an excellent performance, the kernel must be chosen with great care, as it has a crucial effect on the performance [10]. Determining which kernel to use depends on the data and the number of features. In this study the Gaussian radial base function (RBF) was used:

$$K(x, x') = exp(\gamma\|x - x'\|^2), \tag{3}$$

where γ is the kernel parameter and $\|x - x'\|$ is the dissimilarity measure.

4. Training the one-class SVM

Selecting the Gaussian (RBF) kernel and using the default tolerance of the termination criterion, there are two parameters (γ and v), which needs to be optimized when using one-class SVM toolbox [2]. γ defines the extent of the influence of a single training example. Increasing the γ value, lowers the influence. v sets an upper bound on the fraction of outliers (training examples regarded out-of-class) and it is a lower bound on the number of training examples used as Support Vector [13].

5. Estimating the criticality of the system

The trade-off between sensitivity (how well the model is able to predict when the system is in a faulty state) and specificity (how well the model is able to predict when the system is in a healthy state) is often plotted as an ROC (receiver operating characteristic) space, as seen in Fig. 2. The challenge is to find a reasonable, rational and desirable balance between sensitivity and specificity [16].

In the ROC space diagram, there is one area called the conservative area and one called the liberal area. In general, conservative models are able to detect when the system is healthy with fewer false alarms related to a false positive, but may miss faults by giving more false negatives. Liberal models should be able to trigger an alarm signalling when the system is no longer healthy, but may give more false alarms by increasing the false positive value. The remote location of wind farms makes them hard to maintain. Therefore, the cost of false alarms is high and should be avoided. On the other hand, critical faults must be avoided since the initial cost of building a new wind turbine is extremely high. For these reasons, the wind turbine can be considered as a slightly liberal system. Considering these attributes, the closest estimated area for detecting WEC faults should be near the A2 area seen in Fig. 1. Models under group A2 should give a rather good detection sensitivity when a fault is present, but may sometimes give false alarms. The other areas tested in this study are those marked as A1 and A3. Theoretically, area A1 should give fewer false alarms, but should give a later indication of when a fault is present than models near area A2 or A3. Models in group A3 should give more

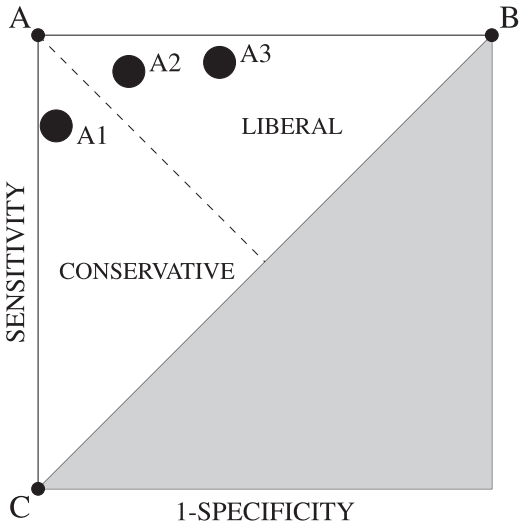


Fig. 2. Illustration of an ROC curve. Point A represents the ideal classifier, which is able to classify all the data points correctly with zero false positives and zero false negatives. Point B represents the classifier which is able to classify correctly the times when the machine is faulty with zero false negatives. Point C represents the classifier which is able to classify correctly every time when the machine is healthy with zero false positives. Areas A1 (specificity 1–0.95), A2 (specificity 0.8–0.75) and A3 (0.6–0.55) are areas which can be targeted when setting the initial trade-off accuracy between sensitivity and specificity.

false alarms than those in the other groups, but should give earlier indications of when a fault has occurred. In this study these three areas are estimated using the following procedure when training the one-class SVM detection algorithm.

1. Nominal data were collected before any faults were present (covering 120 days).
2. The nominal data were divided into two classes, i.e. a training set (covering 100 days) and a testing set (covering 20 days).
3. The initial testing set (covering 20 days) was used to calculate the accuracy (later referred to as the baseline specificity). Three baseline specificity values were targeted, 0.95, 0.85 and 0.7, and they were based on the selected areas A1, A2 and A3.
4. A grid-search method was used where 15 ν and γ parameters were and the maximum ν and γ values were chosen up to a value which would give a baseline accuracy of 50 % or higher.
5. Four models from each A area were selected. If more than four models from the selected area were found, the models were selected where ν and γ are equal or almost equal.
6. Once the parameter values were selected, the models were re-trained using all the nominal data (covering 120 days).

6. Results

The presentation of the results is divided into two parts. First, we present an analysis of the condition of the wind turbine performed using traditional condition monitoring methods (i.e. key indicators extracted from the time domain and enveloped signals), and then we present a corresponding analysis performed using the one-class SVM.

6.1. Traditional analysis

Before any hindsight failure analysis, two events were detected using available traditional vibration analysis tools. These events (Event 1 and 2) are marked into Fig. 3 with arrows. Event 1 indicates a point (on the 240th day) when a small fault was first noticed, but could not be identified or verified. Event 2 (on the 347th day) indicates a time when the fault was identified to be in the inner race of bearing 1. Detailed spectra for both of these events and two measurement during the spall growth are shown in the Fig. 4.

Spectra for the event 1 is shown in the Fig. 4a, where BPFI defect frequency component, its two harmonics and sidebands are marked with narrow lines. This event was estimated by condition monitoring technicians to most likely be a surface anomaly, which

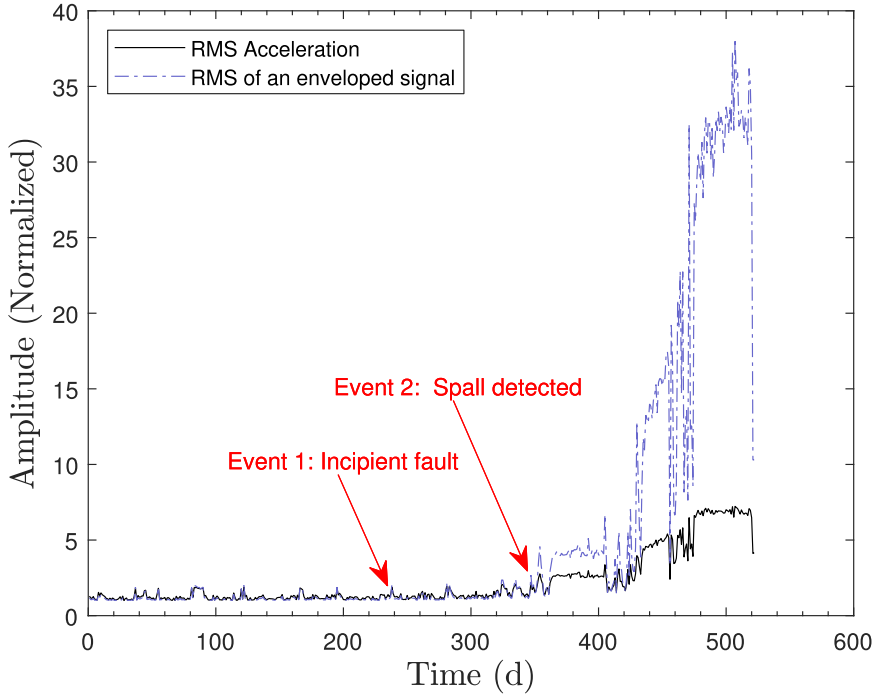
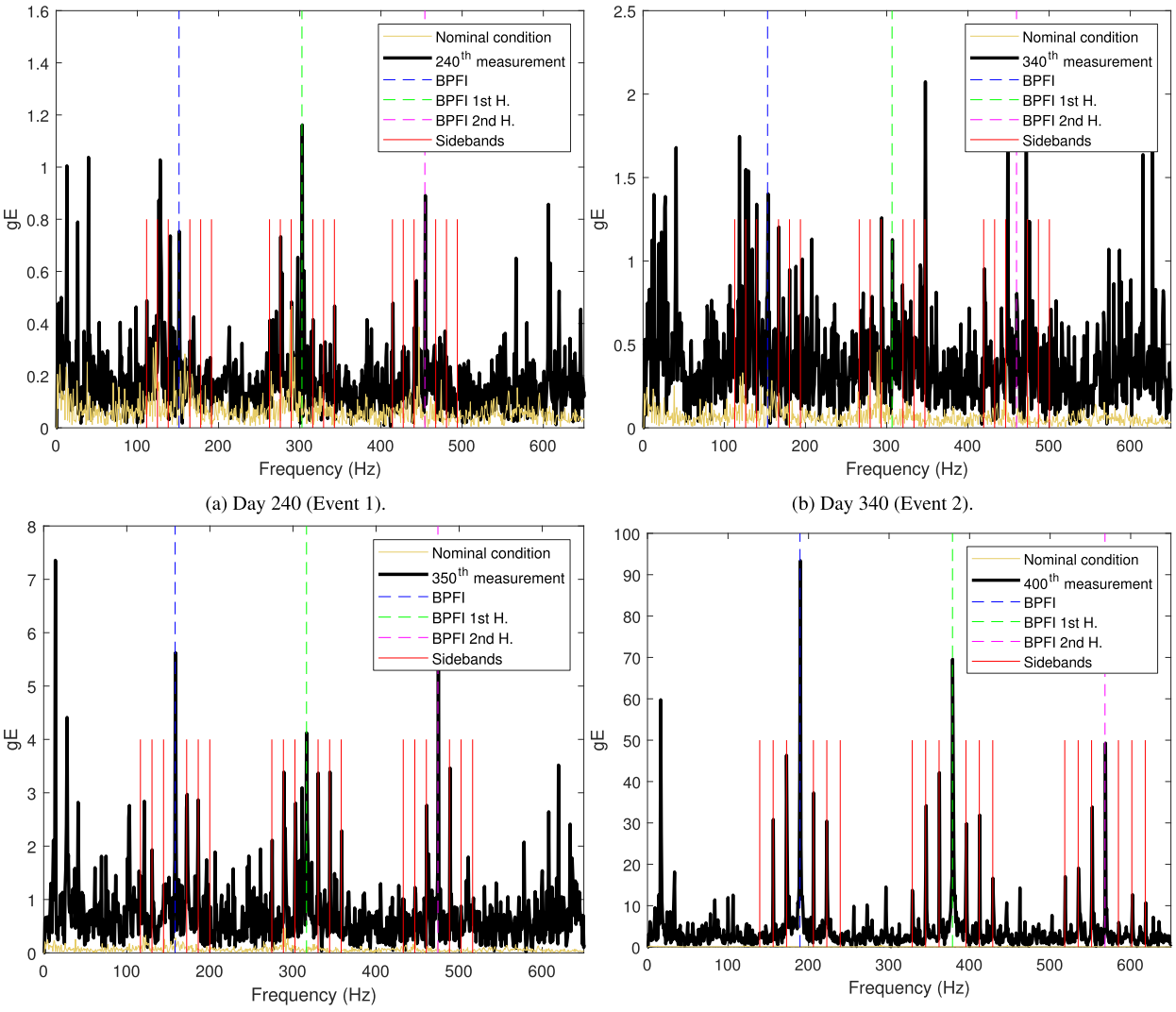


Fig. 3. RMS level of the measured acceleration signals from the wind turbine.



(a) Day 240 (Event 1).

(b) Day 340 (Event 2).

(c) Day 350.

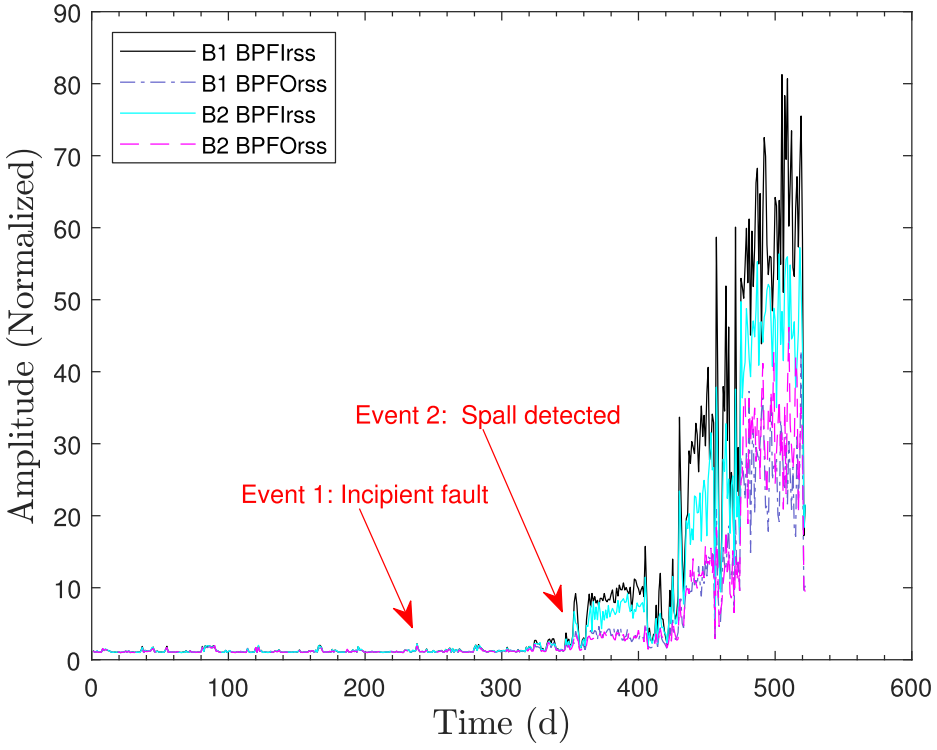
(d) Day 400.

Fig. 4. Enveloped spectra of the acceleration signals. Yellow curve shows the spectra calculated during the first measurement day. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

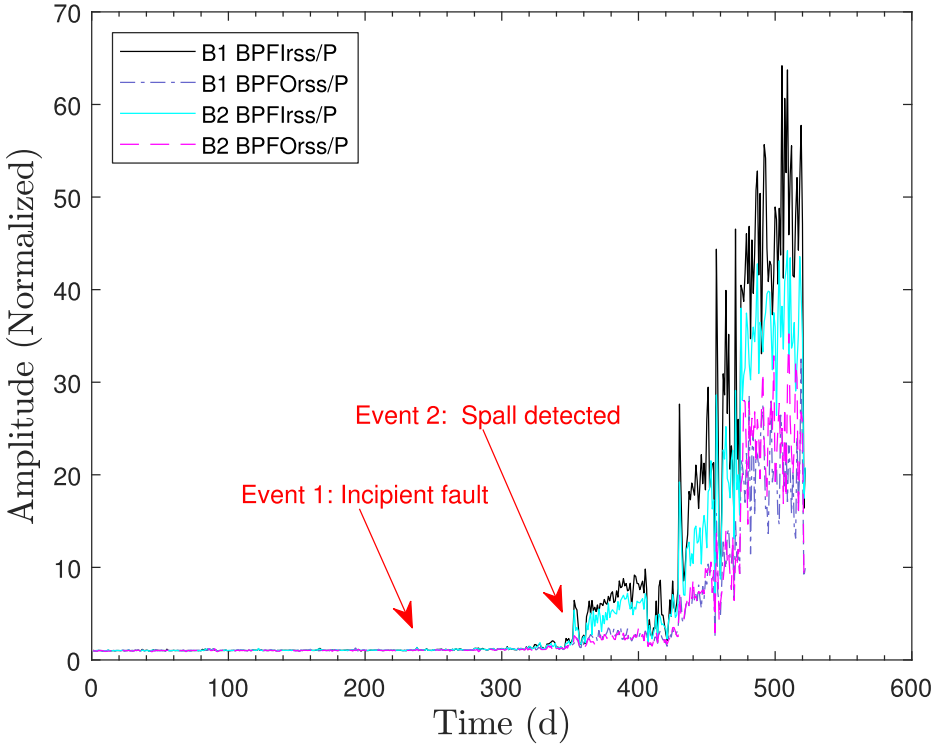
was not a spall yet. As can be seen from the Fig. 4a, overall level of vibration has increased when comparing to the nominal condition. However clear indications cannot be seen at the close proximity of the marked defect frequencies. Spectra for the event 2 seen in the Fig. 4b shows even more elevated overall vibration level of multiple frequency components. Nevertheless, BPFI defect frequency component nor its harmonic are significantly higher than any other frequency components. Some of the sidebands are elevated higher than any other frequency components, which may indicate the start of a spall. Five days later, BPFI defect frequency component and its harmonics can clearly be seen by looking at the Fig. 4c. At this point, most of the vibration frequency components are ten times higher than what they were in the beginning. At this point, most of the good vibration analysis tools should be able to detect the presence of a fault. Fig. 4d, shows the spectra when spall growth has continued for at least 50 days. At this point BPFI defect frequencies were clearly separated from the other frequency components.

Fig. 3 shows the RMS level of the accelerometer signal measured near the high-speed shaft of the WEC as a function of time. The RMS and enveloped RMS values in event 1 are only slightly higher than those seen at days between 81–89 and 167. Therefore, the threshold should have been placed exactly on the correct level in order to detect this incipient fault without any false alarms. The identification of the inner race fault (event 2) might have taken place earlier, since similar shapes (to those appearing during event 2) were also seen twice (see the peaks on day 324 and day 335) before the confirmed spalling. This would indicate that the spalling started at least as early as day 324, but was unnoticed. However, when comparing the peaks seen in event 2 to the one seen at day 81–89, an identification was made rather early, since the spalling could easily have been undetected with the threshold not set exactly to the correct position. Furthermore, most likely the latest detection time using RMS or enveloped RMS indicators would have been around the 354th measurement, since the levels are twice as high as those seen on day 8 (an RMS over 5 and an enveloped RMS over 8). Note that the RMS levels would only be able to send an alarm and would be unable to identify the location of the fault.

Fig. 5a shows the trend of four key indicators used for identifying bearing faults. These features were calculated from



(a) Features without scaling with the process parameter.



(b) Features scaled with the process parameter.

Fig. 5. Trend of wind turbine features calculated from the enveloped accelerometer signals.

the enveloped accelerometer signals and each feature is explained in Table 1. Moreover, event 1 and 2 are marked similarly to how they were marked in Fig. 3.

At event 1, the time of the incipient fault, no difference between inner and outer race features can be observed. Without any further processing, the four features do not provide data to determine fault location at this early stage of the degradation. When comparing the levels of each feature presented in Fig. 5a near event 2, a slight increase is seen in the level of BPFIrss in bearing 1. These results indicate that the fault is located in the inner race in bearing 1. At

Table 2
Separating the wind turbine degradation data into several segments according to the condition of the bearing using traditional methods.

Segments (d)	WEC condition	Estimated stage	Symptoms
1–100	Nominal	Training	Low level of overall vibration
101–120	Nominal	Baseline specificity testing/Training	
121–237	Nominal	False positive	
238–340	Incipient fault	Early alarm	Sudden increase of CI levels
341–350	Faulty system	Late alarm/Early fault identification	Sidebands around BPFI defect frequencies
351–430	Faulty system	Late fault identification	Increase of BPFI related defect frequencies
431-end	Faulty system	Fast fault propagation	High level of overall vibration

day 324 increased level of BPFIrss is also seen and this is most likely the first point when some indications of the fault location could have been identified using features based on the envelope analysis.

Fig. 5b shows the same analysis as presented in Fig. 5a, but with the difference that each feature has been divided with the process parameter P. When the features are scaled with P, the peaks previously seen at days 81 to 89 and day 167 are smeared out. Because the process parameter P is correlated to the load, the peaks seen at days 81–89 and day 167 were probably caused by high load in the WEC system. In fact, during the day 167 the P feature value was eight largest value seen during the whole measurement campaign. It could therefore be advantageous to scale vibration levels with load based features when engineering input features for one-class SVM.

6.2. Fault detection using a one-class SVM

Based on the vibration analysis performed using the traditional methods, the WEC degradation data set was divided into several segments (see Table 2). These segments represent estimated stages of fault progression. For instance, if the trained one-class SVM model sends an alarm between days 121 and 237, it is considered to be a false alarm. Note, that these segments are estimated using manual vibration analysis and cannot be validated with high certainty, especially during the early period of the measurement campaign. Therefore, in the Table 2 description of the symptoms is given how the segments are divided based on the vibration levels during the measurement campaign.

One-class SVM models for fault detection were selected using the steps explained in Section 5. The nominal (healthy) data were divided into two sets, the first of which was used for initial training while the second was used for calculating the baseline specificity. The input features for the results presented in Table 3 were BPFI, BPFI 1th H and BPFI 2th H for bearing 1. The testing set consisted of twenty points and, therefore, the accuracy is increasing by five points. As can be seen in Table 3, the accuracy is decreasing almost consistently when the ν value is increasing. γ value was affecting less to the baseline accuracy.

Based on these results presented in the Table 3, 12 model parameters were selected by choosing four models from area 1 (highlighted with a light grey background), where the targeted baseline specificity was 95 %. In this area the ν values were between 0.045–0.15 and γ values were between 0.045–0.5. Four models were selected from area 2 (highlighted with a medium grey background), where the targeted baseline specificity was 85 %. In this area the range of the ν value was between 0.19–0.26 and γ value was between 0.19–0.26. Four models were selected from area 3 (highlighted with a dark grey background), where the targeted baseline specificity was 70%. In this area the range of the ν value was always 0.29 and γ value was between 0.01–0.5. If more than four models had the same initial specificity within the selected range, the models were chosen to be near the diagonal vector where ν and γ had equal or almost equal values. After this step, models with the chosen model tuning parameters were re-trained using all 120 nominal points in order to increase the training set size.

Later the accuracy of each chosen model was tested using the remaining data set (from measurement point 121 to the end) by taking a sliding window where the five latest measurement points were evaluated. When five (0/5 healthy points in this case) measurements indicated the presence of a fault, an alarm was raised and the further accuracy was set to zero.

6.2.1. Fault detection using specific fault frequency features

Fig. 6 shows the times of the alarms for each of the 12 models when the input features were specified for bearing 1, and Fig. 7 shows the corresponding results when the input features were

Table 3
One-class SVM baseline specificity (%) by varying the ν and γ parameters. Three estimated areas (A1–A3) are highlighted using different shades of grey background. Three BPFI-related features were used as an input.

ν \ γ	**0.01**	**0.045**	**0.08**	**0.12**	**0.15**	**0.19**	**0.22**	**0.26**	**0.29**	**0.33**	**0.36**	**0.40**	**0.43**	**0.47**	**0.5**
0.01	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
0.045	95	95	95	95	95	95	95	95	95	95	95	95	95	95	95
0.08	95	95	95	95	95	95	95	95	95	95	95	95	95	95	95
0.12	95	95	95	95	95	95	95	95	95	95	95	95	95	95	95
0.15	95	95	95	95	95	95	95	95	95	95	95	95	95	95	95
0.19	85	85	85	85	85	85	85	90	85	90	85	85	85	85	85
0.22	85	85	85	85	85	85	85	85	85	85	85	85	85	85	85
0.26	80	80	80	80	80	80	80	80	80	80	85	85	85	85	85
0.29	70	70	70	70	70	70	70	70	70	70	70	70	70	70	70
0.33	60	60	60	60	60	60	60	60	60	60	60	60	60	55	55
0.36	60	60	60	60	60	60	55	55	55	55	55	55	55	55	55
0.40	50	50	50	50	50	50	50	50	50	50	55	55	55	55	55
0.43	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50
0.47	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50
0.5	45	45	45	45	50	50	50	50	50	50	50	50	50	50	50

A1 A2 A3

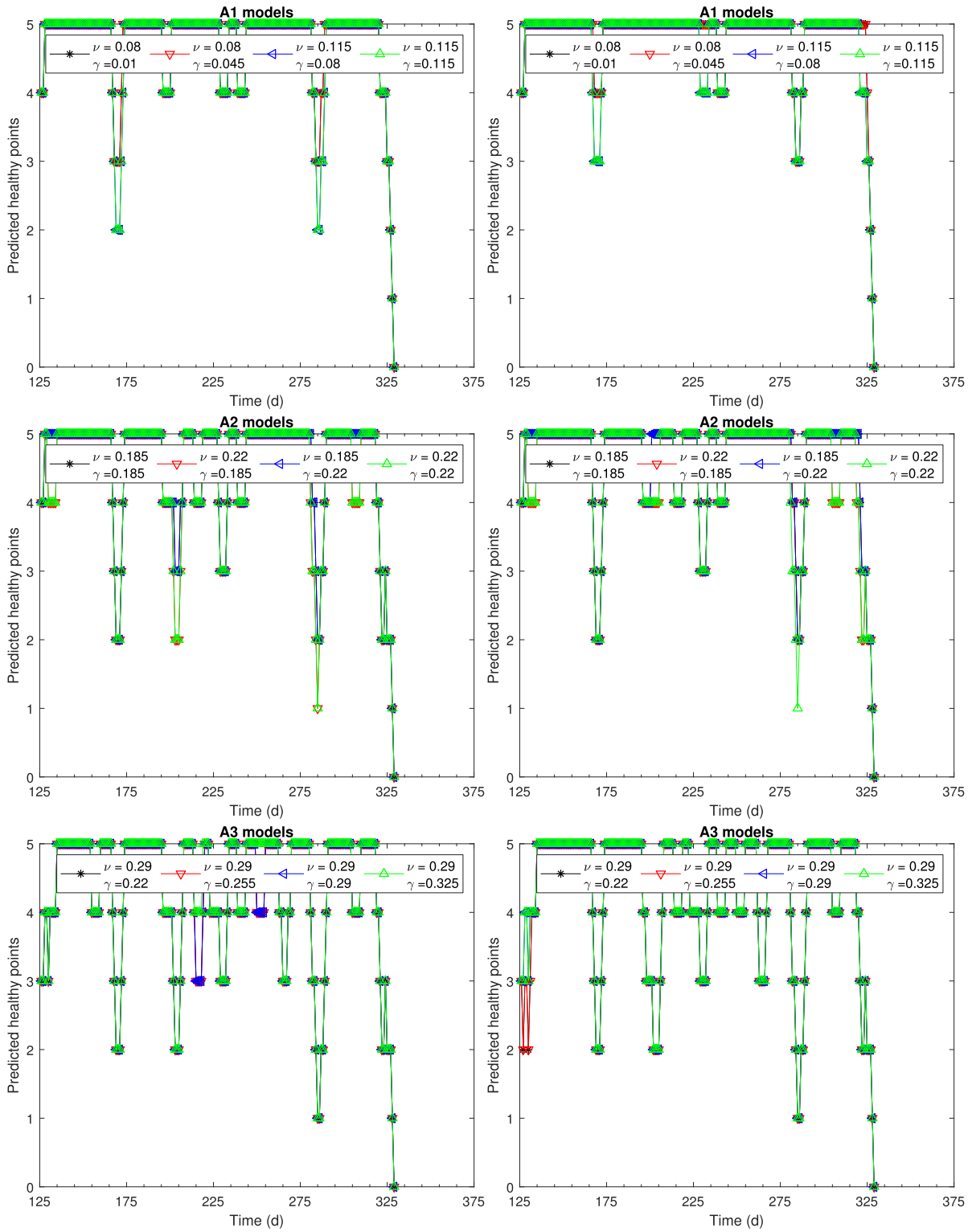


Fig. 6. Detecting bearing 1 faults using the one-class SVM algorithm. The left-hand column shows the number of predicted healthy points when the models have been trained using features related to inner race faults (BPFI, BPFI 1th H and BPFI 2th H). The right-hand column shows the corresponding results when the models have been trained using features related to outer race faults (BPFO, BPFO 1th H and BPFO 2th H).

specified for bearing 2. In these figures, the first row shows four models whose model parameter values were selected using a baseline accuracy of 0.95. These models are considered to be in area 1.

When comparing these first-row models to each other, the results are very similar. All model specified for inner race fault B1 sent an alarm on day 328. This can be considered as a good result for detection. However, when comparing these results

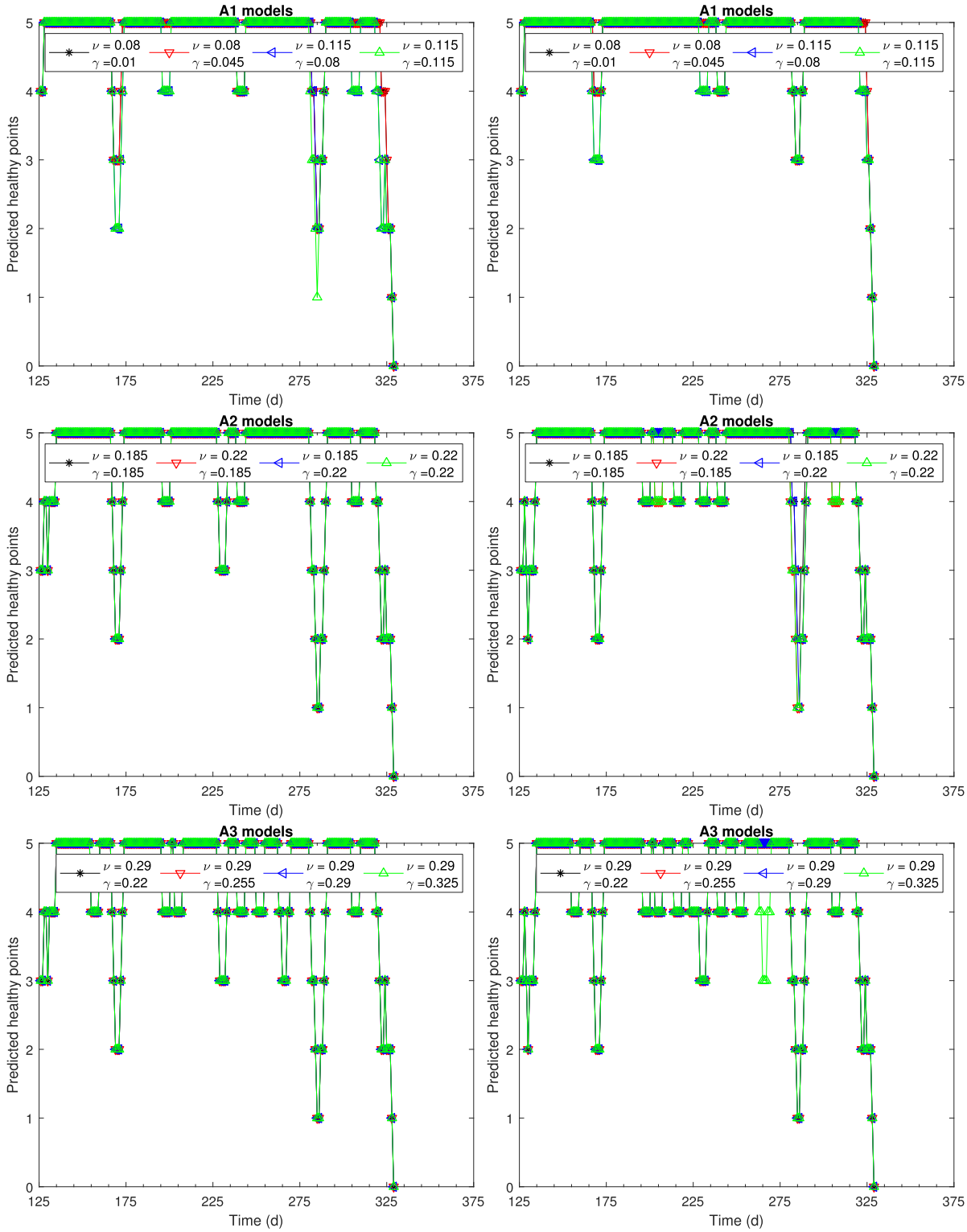


Fig. 7. Detecting bearing 2 faults using the one-class SVM algorithm. The left-hand column shows the number of predicted healthy points when the models have been trained using features related to inner race faults (BPFI, BPFI 1th H and BPFI 2th H). The right-hand column shows the corresponding results when the models have been trained using features related to outer race faults (BPFO, BPFO 1th H and BPFO 2th H).

against models specified for detecting B1 outer race faults, the fault detection is occurring at the same time. Therefore, it is not possible to accurately say if the fault is at the inner raceway or outer raceway.

By comparing these models against models trained using feature set sensitive to detect bearing faults of a bearing, which remained healthy (upper row, Fig. 7), similar results were seen. Therefore, it can be concluded that these A1 models could have been used for detecting abnormal behaviour when 5 consecutive points are sent as anomalous event, but it was not possible to identify the fault location. Surprisingly A1 model ($\nu = 0,0115, \gamma = 0,115$) trained to detect inner race fault of a

bearing 2 would have seen abnormal behaviour with 4/5 consecutive anomalous points at day 285. Even though this can be considered as early alarm time, it would have been difficult differentiate it from a false alarm, especially since this bearing was not the one that was damaged.

The second rows of Figs. 6 and 7 show four models where the model parameter values were selected using an initial accuracy of 0.85. These models are considered to be in area 2 and should give an earlier alarm than A1 models, but may produce more false alarms. Also, these models would have raised the alarm at the same time (day 328), if all measurement points in the tested window should be seen as anomalous. However, there are two models, which would have raised the alarm at day 285 with 4/5 consecutive anomalous points. However, the sensitivity was the same as it was when the model was trained using feature set specified to detect bearing 2 inner race faults.

The third rows of Figs. 6 and 7 show four models whose model parameter values were selected using an initial accuracy of 0.6. In

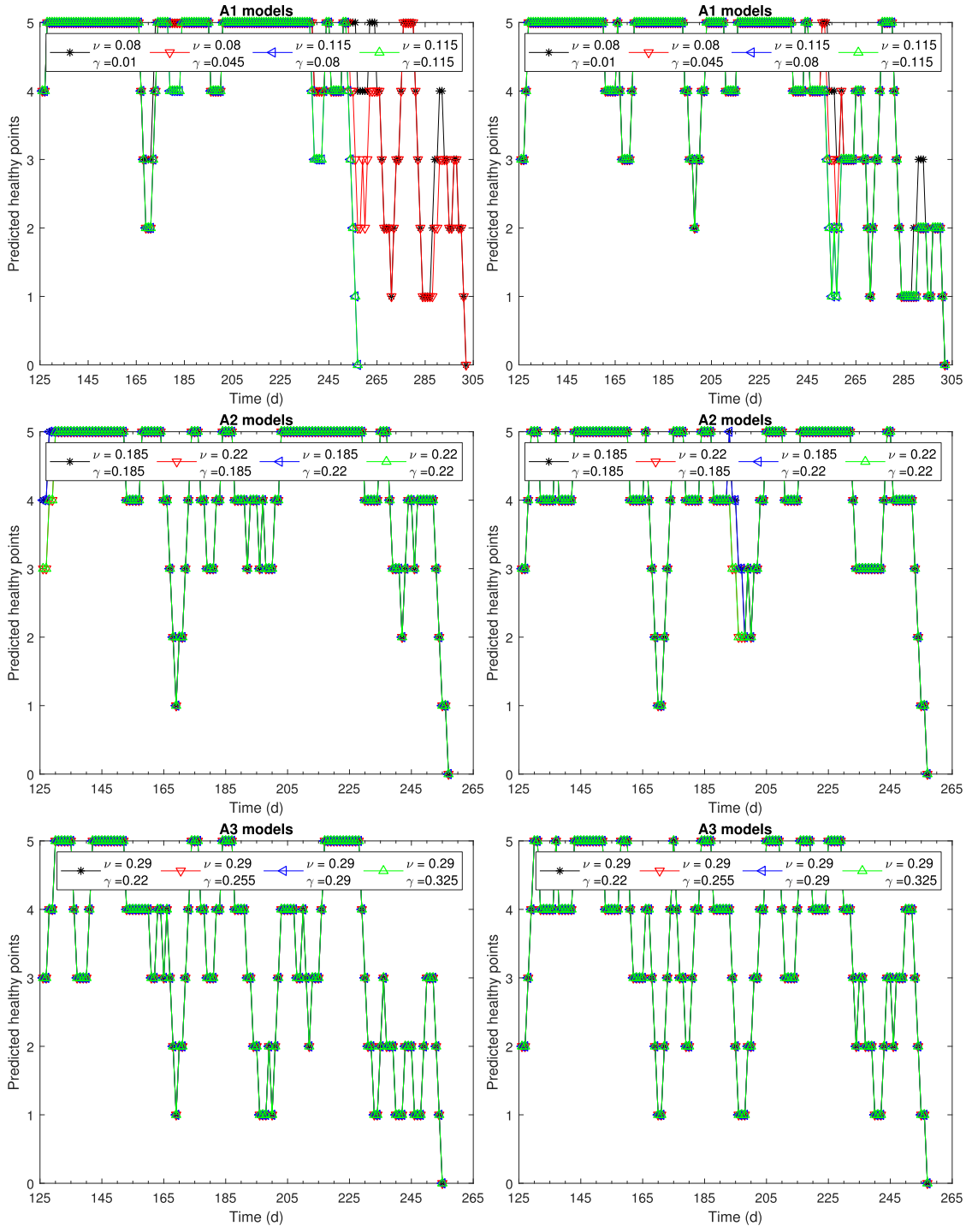


Fig. 8. Detecting bearing 1 faults using the one-class SVM algorithm. The left-hand column shows the number of predicted healthy points when the models have been trained using features related to inner race faults (BPFI, BPFI 1th H and BPFI 2th H), which are divided point-wise by feature vector P. The right-hand column shows the corresponding results when the models have been trained using features related to outer race faults (BPFO, BPFO 1th H and BPFO 2th H), which are divided point-wise by feature vector P.

these models, the results were rather similar as was seen with A2 models. Only exception was that all models would have caused the alarm on a day 285 with 4/5 consecutive anomalous points.

In general, these results indicate that using collective anomalies where 5 points are causing the alarm as a threshold, the selection of model tuning parameter is less important. However, when using 3/5 anomalous consecutive point or less as the threshold, big difference within models were seen. For instance using 3/5 rule as the threshold for the A1 models and comparing the models presented in the first rows of Figs. 6 and 7 with those presented in the second and third rows, there is only one period of time when false alarms are seen, but using the same threshold value for A2 and A3 models, there are two periods (first one starting at a day 170 and second one starting at a day 203) where multiple false alarms are seen.

Note that the parameter selection based on the calculated baseline specificity was only performed using features related to the inner race fault in order to be able to compare the models, which were trained using the same tuning parameters. In reality, parameter values based on the baseline specificity should be

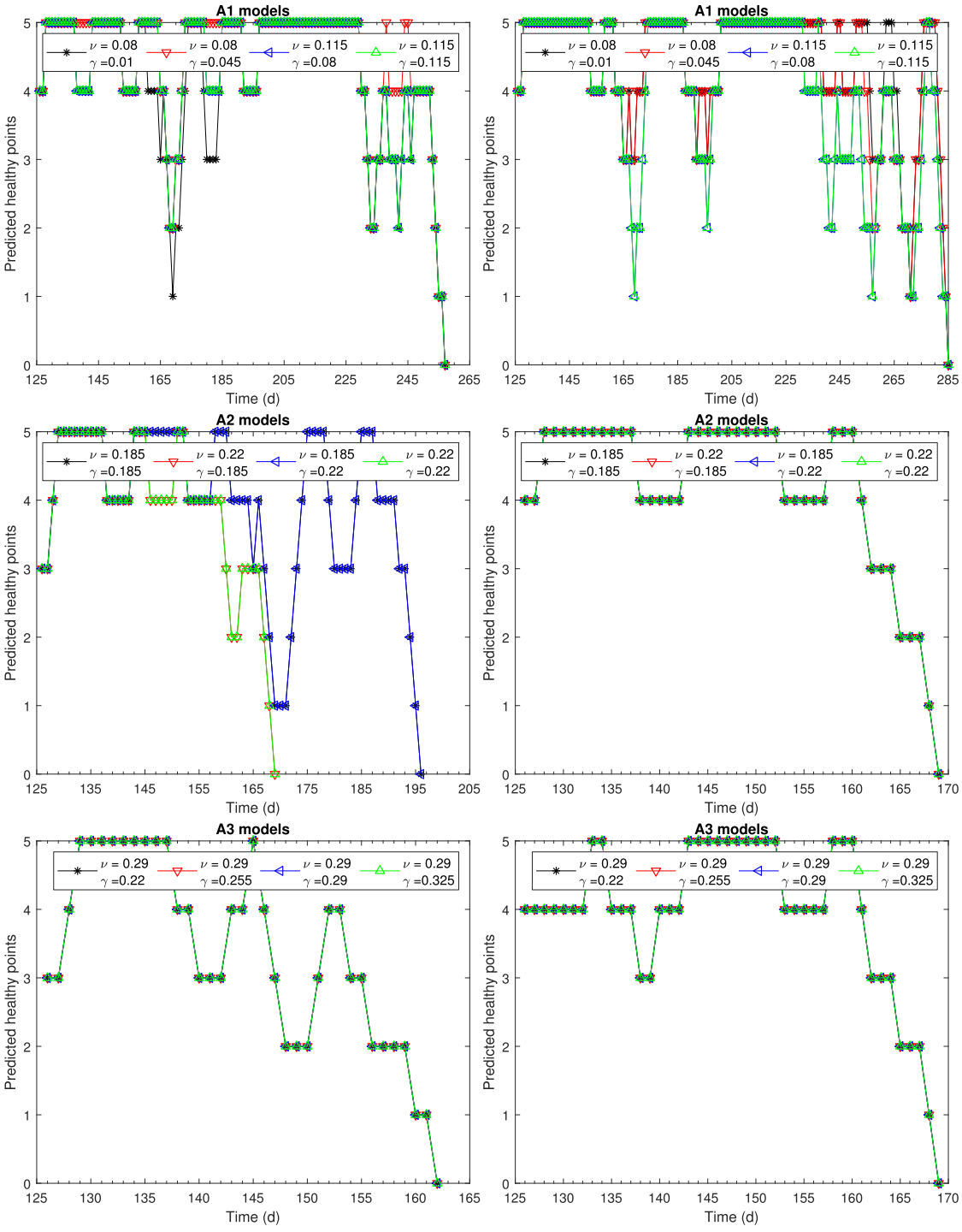


Fig. 9. Detecting bearing 2 faults using the one-class SVM algorithm. The left-hand column shows the number of predicted healthy points when the models have been trained using features related to inner race faults (BPFI, BPFI 1th H and BPFI 2th H), which are divided point-wise by feature vector P. The right-hand column shows the corresponding results when the models have been trained using features related to outer race faults (BPFO, BPFO 1th H and BPFO 2th H), which are divided point-wise by feature vector P.

re-calculated when using different input features in order to obtain the correct model accuracy for other faults too.

6.2.2. Fault detection using specific fault frequency features which are divided using the process parameter

The results using normalized features are presented in Figs. 8 and 9. The approach is similar to the previously presented result, expect now each feature is divided with the measured and normalized process feature (values are normalized between 1–2). By comparing first row models in Fig. 8 it can be seen that fault was detected at the day 257 using the model with $\nu = 0.115$ and $\gamma = 0.08$ and the model with $\nu = 0.115$ and $\gamma = 0.115$, respectively. The other two models were able to detect the fault at day 301. The fault detection on day 257 and 301, can both be considered to be early alarms. As a comparison, models trained to detect inner race faults of bearing 2 were raising the alarm at day 257. This indicates that fault detection using features divided by the process parameter are not as promising as it first seems. For the second row models (A2 area), an alarm was triggered for bearing 1 again at day 257 as shown in Fig. 8. Furthermore, all A2 models would have caused false alarms before day 196 and third row models were even worse than the A2 models.

In overall, based on the results presented in Figs. 8 and 9, the detection and identification of the fault using one-class SVM performed worse when compared to traditional methods. Features divided with process parameter caused more false positive point anomalies and good results were not achieved despite that some of the peaks were smeared, which were caused most likely due to increase stresses in the WEC. Therefore, using the process parameter for reducing the effects of operational changes, was unsuccessful.

6.3. Fault identification by combining faults-specific features

Because the identification of fault location was unsuccessful using models with inner ring and outer ring features separately, one-class SVM models including both these features were tested. In total, six features were considered: BPFI, BPFO, BPFI 1th, BPFO 1th, BPFI 2th and BPFO 2th. Models with combined features, which are able to detect both inner and outer rings faults, would also avoid the problem of redundant detection models in systems with several bearings.

Before training one-class SVM models, baseline accuracy was tested using the grid search with the 15 ν and γ parameters. Results are shown in Table 4. From the table four models were selected from each area (A1–A3) for further studies.

The results when defect frequencies were targeted to detect faults in bearing 1 are presented in Fig. 10. As can be seen in the figure, all A1 models detected the fault at day 355. This can be considered as a late fault detection time. Before day 325 there were no anomalous points, expect immediately after day 125. Using 3 out of 5 consecutive anomalies as the threshold of detection, these models would have detected anomalous event around day 325, which can be considered to be an early time of detection. A2 models detected the anomalous event at day 329 by using 5/5 as the threshold of alarm. By comparing these models against models seen in the group A1, there are more occasions where one or two measurements are seen as anomalies. This is expected and is aligned with the initial hypothesis of using the ROC space selection of models. A3 models also detected the anomalous event at day 329. Because A3 models are slightly more over-fitted than A2 models, more anomalies were seen when compared to the A2 group models.

By selecting a threshold where four faulty cases would raise an alarm (0 or 1 out of 5 predicted as healthy), all 12 models would have detected the anomalous event at the early stage of the fault.

In reality it is not useful that detection models are only targeted for one component in a system. Therefore one-class SVM was trained using the same technique, but selecting features that are more sensitive of finding bearing 2 faults. Since the bearing 2 remained healthy during the measurement campaign, the fault in bearing 1 should be detected and localized later or not at all.

The result for the one-class SVM trained for bearing 2 is shown in Fig. 11. The fault in bearing 1 was detected for all A1 models at day 357 when features were sensitive to find faults in bearing 2 using 5 out 5 fault points as threshold. This is only two days later than models trained to detect bearing 1 faults. Using 4/5 as threshold, the anomalous event was detected 10 days earlier when compared to using features sensitive of detecting bearing 1 faults. These results indicate that one-class SVM can detect the abnormal

Table 4
One-class SVM baseline specificity (%) by varying the ν and γ parameters. Three estimated areas (A1–A3) are highlighted using different shades of grey background. Three BPFI-related features were used as an input.

ν \ γ	**0.01**	**0.045**	**0.08**	**0.12**	**0.15**	**0.19**	**0.22**	**0.26**	**0.29**	**0.33**	**0.36**	**0.40**	**0.43**	**0.47**	**0.5**
0.01	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
0.045	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
0.08	100	100	100	100	100	100	100	95	95	95	95	95	95	95	95
0.12	95	95	95	95	95	95	95	90	90	90	90	90	90	90	90
0.15	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90
0.19	85	85	85	85	85	85	85	85	85	85	85	85	85	85	85
0.22	75	75	75	75	75	75	75	75	75	75	75	75	85	85	80
0.26	70	70	70	70	70	70	75	75	75	75	70	70	70	70	70
0.29	70	70	70	65	65	65	65	65	65	65	65	65	65	65	65
0.33	55	55	55	55	55	55	55	55	55	55	55	55	55	55	65
0.36	55	55	55	55	55	55	55	55	55	55	55	55	55	55	55
0.40	55	55	55	55	55	55	55	55	55	55	55	55	55	55	55
0.43	50	50	50	50	50	50	50	50	50	50	50	50	50	55	55
0.47	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50
0.5	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50

A1 A2 A3

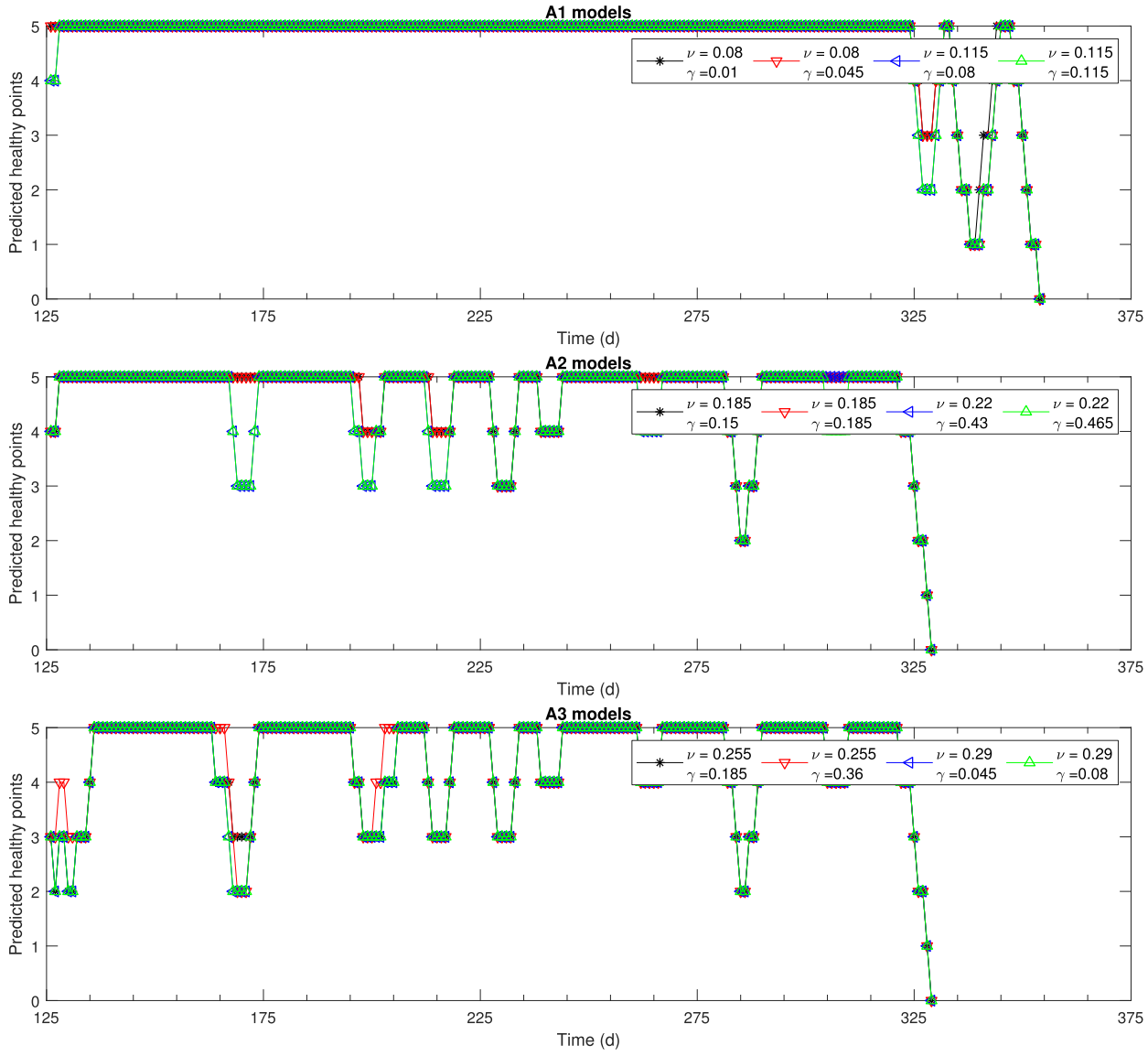


Fig. 10. Detecting bearing 1 faults using the one-class SVM algorithm. Models have been trained using features related to inner race faults (BPFI, BPFI 1th H and BPFI 2th H) and to outer race faults (BPFO, BPFO 1th H and BPFO 2th H).

events using the A1 models, even though the fault is not occurring at the bearing it was trained for. The reason can be an overall increased vibration level as well as leakage of frequency components as explained by Wu and Zhao [18]. To improve fault localization, each feature could be divided with the overall vibration level of the signal envelope.

The A2 group models detected the anomalous event at the same time or some 27 days later (with the threshold 0/5) when compared to models trained for the fault in bearing 1. For the A3 group, detection occurred at the same time as when using features sensitive to detect bearing 1 faults.

Comparing the A1, A2 and A3 groups, the best results were achieved for one-class SVM models in the A2 group. For this group, the anomalous event was detected either earlier or at the same time as using features insensitive for the occurred inner race fault of a bearing 1. Furthermore, these models were able to send the alarm earlier than it was detected using traditional methods.

Because the one-class SVM models using both inner and outer ring features were not able to identify the location of the fault, further development is needed. One approach is to calculate the median value of each specific defect frequency components in the beginning of the test and compare it against the calculated mean value of the measurement points (points inside the detection window), once the model has detected the anomalous event.

If the investigated method is used for detecting other type of faults or in other wind turbines some aspects ought to be considered.

- Training set should be long enough to cover even some of the rarely seen operating conditions. Otherwise too many false alarms are seen and the selection of good tuning parameters is difficult to achieve.
- Since best results were achieved by using 4/5 or 5/5 consecutive anomaly points as the alarming threshold, collective anomaly should work better than point anomaly detection.
- In most of the cases, features specified to detect a certain fault were as accurate as features specified to detect another type of faults. Therefore, the one-class SVM model cannot identify the location and further studies are needed.
- Since the identification was not possible, it is advised to use feature set where numerous defect frequency related features are combined.

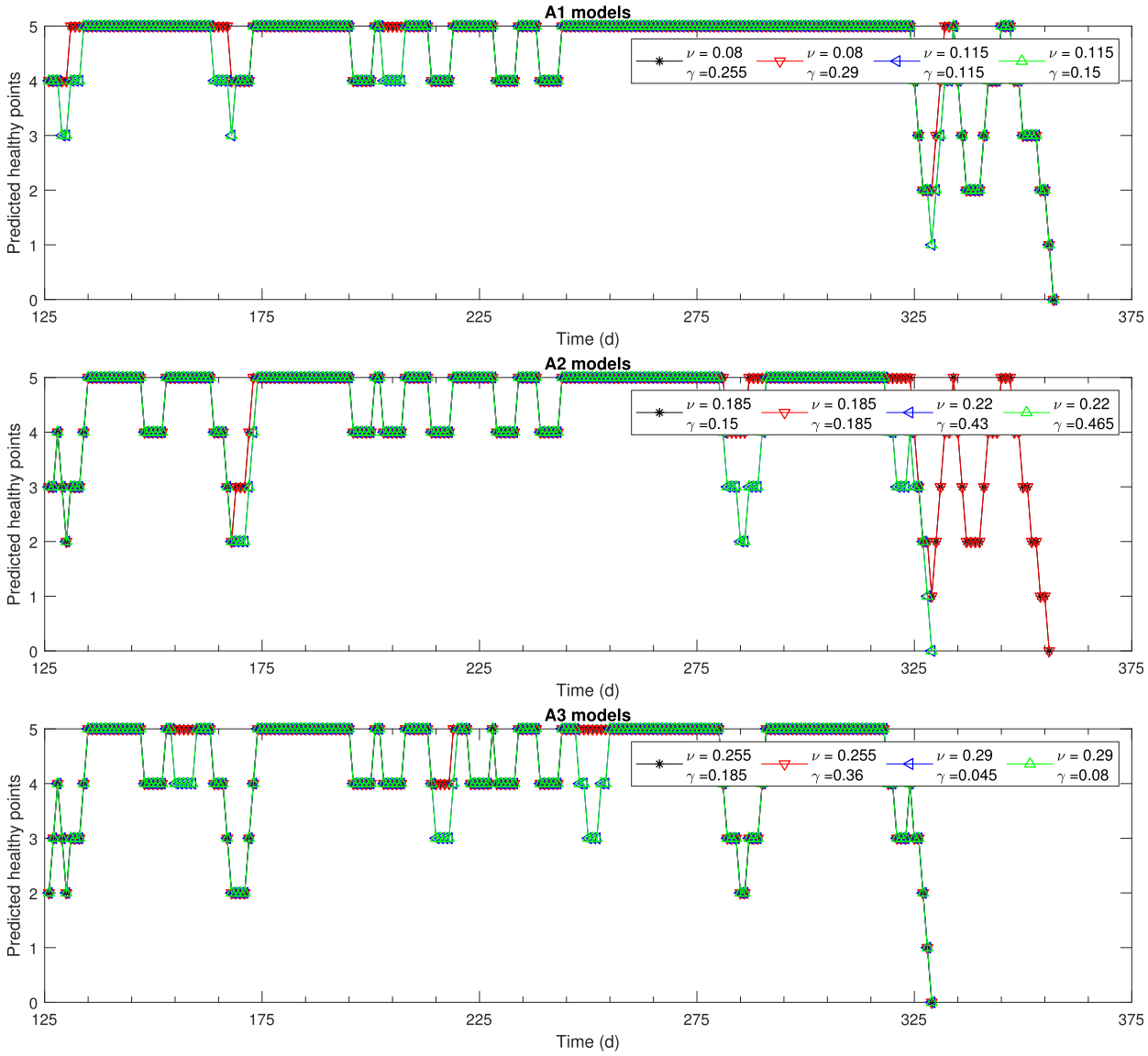


Fig. 11. Detecting bearing 2 faults using the one-class SVM algorithm. The left-hand column shows the number of predicted healthy points when the models have been trained using features related to inner race faults (BPFI, BPFI 1th H and BPFI 2th H) to outer race faults (BPFO, BPFO 1th H and BPFO 2th H).

7. Conclusions

The present study arrived at the following findings concerning the detection of WEC bearing faults by using a one-class SVM with the given data set, where an inner race spalling was later confirmed.

- Accurate fault detection was achieved when BPFI-related and BPFO-related features were combined as one feature set and the model parameters were selected using pre-defined baseline specificity of 0.85.
- Fault detection capability was worse when each measured feature was scaled using the process parameter.
- A combination of model tuning parameters and feature set was not found that were able to identify the bearing fault location without using any additional post-process techniques.

Acknowledgements

The authors would like to thank SKF AB and Vinnova for their financial support.

References

[1] T. Barszcz, R.B. Randall, Application of spectral kurtosis for detection of a tooth crack in the planetary gear of a wind turbine, Mech. Syst. Signal Processing 23 (4) (2009) 1352–1365.
[2] C.-C. Chang, C.-J. Lin, Libsvm: a library for support vector machines, ACM Trans. Intelligent Syst. Technol. (TIST) 2 (3) (2011) 27.
[3] M. Desforges, P. Jacob, J. Cooper, Applications of probability density estimation to the detection of abnormal conditions in engineering, Proc. Inst. Mech. Eng., Part C: J. Mech. Eng. Sci. 212 (8) (1998) 687–703.
[4] D. Fernández-Francos, D. MartíNez-Rego, O. Fontenla-Romero, A. Alonso-Betanzos, Automatic bearing fault diagnosis based on one-class ν-svm, Comput. Ind. Eng. 64 (1) (2013) 357–365.
[5] Z. Hameed, Y. Hong, Y. Cho, S. Ahn, C. Song, Condition monitoring and fault detection of wind turbines and related algorithms: A review, Renewable Sustainable Energy Rev. 13 (1) (2009) 1–39.
[6] M.C.O. Keizer, S.D.P. Flapper, R.H. Teunter, Condition-based maintenance policies for systems with multiple dependent components: A review, Eur. J. Operational Res. (2017).
[7] J. Keller, M. McDade, W. LaCava, Y. Guo, S. Sheng, Gearbox reliability collaborative update. US National Renewable Energy Laboratory (NREL) report PR-5000-54558, 2012.
[8] M. Markou, S. Singh, Novelty detection: a review-part 1: statistical approaches, Signal Processing 83 (12) (2003) 2481–2497.
[9] P. McFadden, J. Smith, Vibration monitoring of rolling element bearings by the high-frequency resonance technique-a review, Tribology Int. 17 (1) (1984) 3–10.

[10] K.-R. Muller, S. Mika, G. Ratsch, K. Tsuda, B. Scholkopf, An introduction to kernel-based learning algorithms, IEEE Trans. Neural Networks 12 (2) (2001) 181–201.
[11] R.B. Randall, J. Antoni, Rolling element bearing diagnostics-a tutorial, Mech. Syst. Signal Processing 25 (2) (2011) 485–520.
[12] M. Schlechtingen, I.F. Santos, Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection, Mech. Syst. Signal Processing 25 (5) (2011) 1849–1875.
[13] B. Schölkopf, A.J. Smola, R.C. Williamson, P.L. Bartlett, New support vector algorithms, Neural Comput. 12 (5) (2000) 1207–1245.
[14] B. Schölkopf, R.C. Williamson, A.J. Smola, J. Shawe-Taylor, J.C. Platt, et al., Support vector method for novelty detection, NIPS, Vol. 12, Citeseer, 1999, pp. 582–588.
[15] H.J. Shin, D.-H. Eom, S.-S. Kim, One-class support vector machines-an application in machine fault detection and classification, Comput. Ind. Eng. 48 (2) (2005) 395–408.
[16] J.A. Swets, R.M. Dawes, J. Monahan, Psychological science can improve diagnostic decisions, Psychological Sci. Public Interest 1 (1) (2000) 1–26.
[17] L.F. Villa, A. Reñones, J.R. Perán, L.J. De Miguel, Angular resampling for vibration analysis in wind turbines under non-linear speed fluctuation, Mech. Syst. Signal Processing 25 (6) (2011) 2157–2168.
[18] J. Wu, W. Zhao, A simple interpolation algorithm for measuring multi-frequency signal based on dft, Measurement 42 (2) (2009) 322–327, URL: http://www.sciencedirect.com/science/article/pii/S026322410800105X.
[19] A. Zaher, S. McArthur, D. Infield, Y. Patel, Online wind turbine fault detection through automated scada data analysis, Wind Energy 12 (6) (2009) 574–593.