Using wavelet transform analysis and the support vector machine to detect angular misalignment of a rubber coupling

Juhamatti Saari∗
Johan Odelius ** Jan Lundberg ** Matti Rantatalo**

∗ SKF-LTU University Technology Centre, Luleå University of Technology, SE 97187 Luleå, Sweden (e-mail: juhamatti.saari@ltu.se)
** Division of Operation, Maintenance and Acoustics, Luleå University of Technology, SE 97187 Luleå, Sweden

Abstract: Shaft misalignment is a common problem for many types of rotating systems. It can cause machine breakdowns due to the premature failure of bearings or other components. Common diagnostic approaches rely on detecting increasing vibration response spectra at multiples of the shaft speed. However, in many time-variant systems, such as wind turbines, the speed and load vary considerably, which can make spectrum analysis insufficient. In this paper, a method for detecting shaft misalignment by using wavelet analysis is proposed. The method was experimentally evaluated in a laboratory test rig for four different operating conditions by varying the rotational speed and load. An angular misalignment was introduced between a hydraulic pump (load) and a medium-sized industrial gearbox connected with a rubber coupling. Vibration data were collected by using two accelerometers mounted in an axial and a radial direction directly on the gearbox casing. The features extracted from wavelet representation were classified by using a support vector machine algorithm. The detection of misalignment and the sensitivity of the proposed method are presented using validation data and confusion matrices. The final results of the confusion matrices clearly indicate that this method can detect misalignment even when the speed and load vary. The proposed method can be used for systems which are connected with shafts and there are many similar systems (comprising an electric motor, a gearbox and a centrifugal pump) working under the same circumstances.

Keywords: Shaft misalignment, wavelet transform, support vector machine, machine learning

1. INTRODUCTION

A properly aligned rotating machine can reduce the maintenance cost by increasing the lifespan of the system and its components. It can reduce bearing, coupling, seal and shaft failures and improve the manufacturing quality by reducing the overall vibration levels. Moreover, an aligned system will enjoy reduced energy consumption. It has been estimated that the energy consumption can be reduced by 4-5% on average Luedeking (2012). Furthermore, this is just the tip of the iceberg, which is relatively easy to see. The biggest savings are to be gained through the increased lifetime of the system and the reduced unscheduled downtime and repair expenses. Therefore, the need for automatic detection of misalignment definitely exists, even though this need is not recognized by many, since misalignment may remain hidden and only secondary failures of other components may be detected. In order to increase the lifetime of the system or accurately estimate its remaining technical life, knowing the goodness of the shaft alignment plays an important role. In this paper, a continuous wavelet transform and a support vector classifier have been tested to detect the angular misalignment of a rubber coupling.

Nowadays, when the computational power of computers has increased and more methods are available for the use of this computer power to solve condition monitoring problems, there is a demand for automatic fault diagnosis methods. Methods for fault diagnosis usually involve experts performing the final analysis and/or scheduled routines for manual inspections Kothamasu et al. (2006). Well-known methods are the extraction of sensitive features from the time domain and spectrum analysis (performed before or after enveloping). Based on the spectrum, known hit repetition frequencies can be located which can indicate certain faults. To diagnose the state of a machine, experts often use methods in both the time and the frequency domain, as well as many more (non-vibration based) methods, to draw the final conclusions about the state of the machine. In order to develop automatic fault diagnosis methods, it should be possible to use similar approaches. However, there are some problems that need to be solved before this will be possible. First of all, experts usually have a deep knowledge of similar machines and they can use all of their senses to gather information from the surroundings of the machine. For example, for a human it is easy to determine from the vibration signals if the noise level is too high or the speed variation is too big to use spectrum analysis. However,
for unsupervised algorithms these steps are much harder. Condition monitoring (CM) experts do not necessarily need any charts or thresholds in order to realize when to use more advanced condition monitoring techniques, such as filtering or signal averaging techniques, to process the signal before analysis; but again, that is not the case for machines. Even the smartest algorithms usually rely on people who first set up these thresholds before the algorithm can work. Fuzzy logic algorithms are a good example of this type of algorithm Yao (1998). However, a large-scale automated diagnosis method should work even without this step, and such methods are currently one of the main topics in the field of CM.

An interesting candidate for selection as an automated diagnosis method is a combination of existing CM methods with machine learning techniques developed in computer science for data analysis. The idea of using machine learning techniques is to ease the load on trained experts by drawing simple conclusions automatically, or to transform the information to a level which is more understandable for the operator.

One approach worth considering for automated fault diagnosis is not to use traditional time domain or frequency analysis separately for feature extraction, but instead to use both at the same time. This type of analysis is referred to as time-frequency analysis. It is more suitable than frequency analysis when the speed varies, e.g. for moving vehicles and wind turbines. It can also reveal small impacts in the continuous signal, which may be impossible for pure frequency analysis. Some of the known time-frequency analysis methods used for condition monitoring are the Wigner-Ville distribution (WVD), the short-time Fourier transform (STFT) and the wavelet transform (WT).

The WVD is the oldest known time-frequency transform method. Wigner (1932) applied it to quantum mechanics in the beginning of the 1930s, and in the 1940s Ville (1948) applied the transform to signal processing, which explains the origin of its name. Since then it has been used to diagnose many types of machine faults. For example, Staszewski et al. (1997) used it to detect gearbox faults. Although the WVD leads to a superior frequency-domain resolution, it can produce high-energy coefficients in the transform plane, even though no such coefficients actually exist. This inference term can be filtered, but then some of the excellent frequency resolution will be lost. When filtering of the plane is included, the WVD can be referred to as a pseudo-Wigner-Ville distribution. A comparative study of the use of this method for CM is to be found in an article written by Baydar and Ball (2001).

The STFT, or windowed Fourier transform, uses a windowing function to separate a small section from the signal (short time) and produces a snapshot of the signal. Overlapping each analysed segment and summing them will lead to an image (named a spectrogram) which can represent how the signal will vary in time. The windowing function can have different shapes, as in standard frequency analysis. By choosing the optimum windowing function, the detection performance can be improved. However, the benefits can be quite minimal, and mostly a Gaussian

windowing function is used, since it has been proven to work in many cases Wang and McFadden (1993). The STFT has also been widely used for detecting faults in rotating machines Cohen (1989); Bartelmus and Zimroz (2009). Unfortunately, because of the nature of the calculation of the spectrogram, it is not possible to know the exact time-frequency representation; i.e. by knowing the frequency component precisely, the exact time instance is unknown. This is also known as a manifestation of the famous Heisenberg Uncertainty Principle, Allen and Mills (1993).

This fixed resolution pitfall of the STFT is one of the reasons why the wavelet transform was developed Peng and Chu (2004). Wavelets are functions whose translations and dilations can be used for expansions of square-integrable functions. Instead of having a fixed window shape, as is the case in the STFT, the clever idea was devised of using the same basic filter shape (mother wavelet) and shrinking its time domain extension. This leads to a time-scale representation that can have a good time resolution for high-frequency events and a good frequency resolution for low-frequency events. For this reason the wavelet transform is a promising tool which can be used for many types of machine faults Peng and Chu (2004). It can detect transient signals whose origin is, for example, a broken gear tooth Fan and Zuo (2006); Bafroui and Ohadi (2014) or signals which are longer in duration and are caused, for example, by worn gears Bafroui and Ohadi (2014) or worn bearings Li and Ma (1997).

Even though wavelets have been studied with a view to optimizing and automating feature extraction for problems such as gear and bearing faults Rafiee et al. (2010), less effort has been made to solve problems such as misalignment. The most common method for diagnosing misalignment still relies on detecting increasing vibration response spectra at the shaft speed or its harmonics. However, this can be problematic when the speed varies or low frequency noise of the machine is masking the signal. Moreover, it has been shown by many researchers working on the misalignment problem that, for example, for some systems the 2nd harmonic in frequency domain is the most dominant, while for others the 4th in the time domain can be the most dominant Lahdelma and Laurila (2012). To make the problem even more complicated, misalignment can be highly non-linear, in such a way that a bigger misalignment does not necessarily mean that the amplitude will increase.

Although wavelet analysis can be a very effective method for diagnosing many types of faults, it requires trained eyes to detect different types of faults from the image (WT scalogram), just as it requires a professional to interpret the spectrum. However, a substantial amount of research has been conducted to analyse images by using machine learning techniques. Recently these techniques have become more popular and many people have used them to diagnose machine failures. However, in order to apply these methods, the following two crucial steps are needed:

(1) finding sensitive features that represent well all the possible changes which can be seen in the scalogram,
Suitable machine learning techniques for classifying features extracted from the scalogram images are a topic of great current interest and are dealt with in Agarwal et al. (2004). Deeper knowledge of how the classifying algorithm chosen for the present study (the SVM algorithm) works and how it is used for condition monitoring can be found in Widodo and Yang (2007).

SVM algorithms are efficient learning algorithms for non-linear functions. They can separate non-linear regions by using kernel functions to measure similarity, based on the dot products of the vector space. By separating the data into training, testing and validation sets, it is possible to find a model which will maximize the decision boundary, and new measurements performed online can be tested without defining any threshold limits manually. The main advantage of using the SVM is that the value of each feature does not necessarily have to increase in order to separate the healthy system from the unhealthy one. In the case of a misalignment this is important, since sometimes it might happen that the vibration level of a certain frequency component is actually decreasing when the axles are poorly aligned. This might be due to an increase in the static load, which may not be detected by using a vibration sensor, which is more sensitive to dynamic loads. One disadvantage of the SVM is that there is no standard method for choosing the kernel function. Moreover, as in the case of other machine learning techniques, over-fitting or under-fitting the data can happen when the technique is not applied correctly in all its parts or when features are chosen poorly. Therefore, it is important to separate the data properly into the training, testing and validation sets. A limitation of the SVM which explains why it is not commonly used for CM diagnosis is that both healthy and unhealthy data are needed for training the algorithms to detect faults in the future (Widodo and Yang 2007). However, this is a common problem for all supervised methods and not only for the SVM. To solve this problem, it has been suggested that synthetic data should be used to simulate all or some of the failure modes Leturiondo et al. (2015). Another way of obtaining both healthy and unhealthy data is to apply a good management protocol in which, each time a failure has happened, the data are labelled correctly with a good description from the system. Later on these data can be used to train the classifying algorithm to diagnose the failure the next time it occurs.

2. METHOD

In this study, complex Morlet wavelets have been used to denoise the raw vibration signal by extracting features with varying scaling and bandwidth parameters. From the resultant coefficients, statistical parameters were chosen to obtain as much information as possible for the classification algorithm. The classifier used was a support vector machine (SVM), which is a supervised machine learning model. The model was validated using a new set of data and a confusion matrix was used to present the classification results. In Figure 1 the steps needed to build up the model can be seen.

Misalignment was introduced between a hydraulic pump (load) and a medium-sized industrial gearbox connected with a rubber coupling. The coupling type was a so-called doughnut type, which is one of the best types of rubber couplings for enduring a large degree of misalignment. This type of coupling is mainly used in cases where sudden impacts are likely to happen, such as highly loaded gearboxes.

The test rig (Figure 2) used in this study was specifically designed for testing several different failure modes in an environment where natural noise is coming from other components (e.g. hydraulic systems, and bigger electric motors), since the scale of the test rig is much closer to the size of normal machine components. In Table 1 some of the main component of the test rig are listed.

<table>
<thead>
<tr>
<th>Specification</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motor</td>
<td>75 kW three phase AC electric motor</td>
</tr>
<tr>
<td>Controller</td>
<td>ABB inverter</td>
</tr>
<tr>
<td>Max RPM</td>
<td>1600 RPM</td>
</tr>
<tr>
<td>Gearbox</td>
<td>Mekanex 602A Ratio:3.61</td>
</tr>
<tr>
<td>Coupling</td>
<td>Rubber doughnut coupling</td>
</tr>
<tr>
<td>Load</td>
<td>Hydraulic pump max. torque 500 Nm</td>
</tr>
<tr>
<td>Vibration sensors</td>
<td>IMI PCB 10mV/g</td>
</tr>
<tr>
<td>Torque sensor</td>
<td>Linear range 0-1000 Nm</td>
</tr>
</tbody>
</table>

2.1 Wavelet transform and feature extraction

The selection of the mother wavelet for wavelet analysis is still an open question, although many researchers have tried to find the best solution or a method for choosing the mother wavelet by comparing the signal to a best suitable one Raiee et al. (2010). Some have also used the well-known trial-and-error method to define it before the analysis. However, previous studies have ascertained that, even though the Morlet wavelet may not be the best candidate, it is usually one of the best candidates for many types of faults Raiee et al. (2010). The reason for this is that impacts usually create a transient signal whose shape is very close to the shape of the Morlet wavelet after damping effects. Therefore, it is reasonable to choose the Morlet wavelet as the mother wavelet when trying to implement an automated fault diagnosis method based on wavelet analysis.

The most common feature of interest in traditional CM methods is the root mean square (rms) value of the vibration acceleration. This value is a measure of the energy content of the signal and is a good indicator of the overall health of the system. However, it can still be a challenge to diagnose the fault just by using the rms value, since the root cause is very difficult to isolate and attribute to a certain component and its problem.

To overcome the difficulty of identifying the reason for a vibration level increase, the common procedure is to use a band alarm in such a way that the raw vibration signal is pre-processed by filtering over a chosen frequency band, and then a proper threshold is selected which will trigger an alarm if the amplitude of the frequency component reaches this level. These bands can be chosen.
by using physical models for which, for example, the ball passing frequency for the outer race of the bearing or the mesh frequency of two meshing gears is known. The limitations of this method are that usually the speed needs to be known precisely and the variation of the speed should be minimal. This is especially the case when the rotation speed of the shaft is low and other passing frequencies are occurring at close frequencies. Moreover, the limitation of just detecting the peak of a certain frequency component is that the increased level of the sidebands may go undetected, or the band must be wide enough and other statistical parameters (e.g. the rms) need to be used for the band-passed signal. However, the risk of detecting other passing frequencies or pure noise is increasing at the same time. Using band alarms for detecting misalignment can be challenging, since there are no specific passing frequencies which are to be located and isolated. However, misalignment is usually detected through the observation of an increased level of vibration in the shafts rotation speed or in its harmonics bands using these frequencies. The difficulty of using this technique is that misalignment can be highly non-linear, in such a way that the vibration level might actually decrease when a bigger misalignment occurs. In addition, for some cases the 2nd harmonic in the frequency domain is the best indicator, while the 3rd or 4th harmonic is the best for others.

Using wavelet transforms for feature extraction has several advantages. First of all, the WT method is a combination of time domain and frequency analysis and can detect transient signals and give an indication of the condition of the system based on the overall behaviour and what is usually seen in the time domain analysis. The continuous wavelet transform (CWT) can be expressed in the following equation:

$$ M_\Psi f = S(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} f(t) \Psi(\frac{t-b}{a}) \, dt, $$

where \( a \) is the scaling factor and \( b \) is the translation value. \( \Psi(t) \) is a continuous function usually referred to as the mother wavelet. In this study a complex Morlet wavelet (CMW) has been used as a mother wavelet. A CMW can be defined as,

$$ \Psi(t,f_c,f_b) = \frac{1}{\sqrt{\pi f_b}} \exp(-t^2/f_b) \exp(j2\pi f_c t), $$

where \( f_c \) is the centre frequency and the \( f_b \) is the bandwidth parameter. Figure 3 illustrates how changing the parameter \( f_c \) will affect the time-frequency resolution of the wavelet analysis. Note that when the bandwidth parameter value is big, Morlet wavelet is very close to a sinusoidal wave.

After a WT has been performed, some statistical parameters are still needed in order to reveal the change in the attributes of each decomposed WT parameter. Usually for spectrum analysis, the amplitude of each frequency component is used to determine the behaviour of each feature. However, in wavelet analysis each scale coefficient
can be treated as the time domain signal and more values than just the peak value can be used. In this article, five different statistical parameters (see Table 2) have been used to extract as much information as possible from each scale coefficient.

Table 2. Statistical parameters extracted from the wavelet coefficients.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Equation</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peak</td>
<td>max(</td>
<td>x_k</td>
</tr>
<tr>
<td>RMS</td>
<td>\sqrt{\frac{1}{N}\sum_{k=0}^{N}x_k^2}</td>
<td>Measure of the energy content of the signal</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>\frac{E((x-\mu)^3)}{\sigma^3}</td>
<td>Measure of the peakedness of the probability distribution of a real-valued random variable</td>
</tr>
<tr>
<td>Skewness</td>
<td>\frac{E((x-\mu)^3)}{\sigma^3}</td>
<td>Measure of the asymmetry of the data around the sample mean</td>
</tr>
<tr>
<td>Range</td>
<td>max(x_k) - min(x_k)</td>
<td>Difference of the maximum and minimum value of the signal</td>
</tr>
</tbody>
</table>

*E(t) is the expected value of the quantity t

2.2 Principal component analysis

Principal component analysis (PCA) is a method for identifying patterns in data or expressing data in such a way that the similarities and differences can be highlighted. Another main advantage of PCA is its capability to reduce the number of data dimensions without losing too much information. In the present study, PCA was first used to reduce the dimensions to two in order to gain a better understanding from the SVM results. However, this step is not compulsory and sometimes may weaken the sensitivity of the model. Accordingly, to reduce the dimensions without losing too much information, PCA was used in such a way that a 95% variance was retained each time before training the SVM.

2.3 Support vector machine

To achieve non-linear classification, a Gaussian kernel was used for the SVM. To avoid problems such as over-fitting and converging to a local minimum, 5-fold cross-validation was implemented and 20 random initial values were used. From these sets, the Matlab fminsearch function was used to choose the best parameters. The data were always separated using 60-20-20% sections (training, testing and validation). The actual number of data instances is to be found in Table 3.

2.4 Test setup

The time length of each data instance was 2.5 seconds and for each speed and load, data instances were collected by using an accelerometer which was stud-mounted directly on the gearbox in an axial direction, as can be seen in Figure 2. Data were also collected using another accelerometer mounted in a radial direction. However, since angular misalignment is usually easier to detect using an axial direction, only data from the accelerometer mounted in an axial direction were used to train the final model. Nevertheless, it would be possible to repeat the same steps for a radial direction or even combine features extracted from both locations, before training the SVM. All the data were collected using a National Instruments PXI platform. The sampling frequency used was 102.4 kHz. However, the linearity of the accelerometers used was within an error or 15% up to 10 kHz, and therefore, before any wavelet analysis, all the signals were low-pass filtered and downsampled to 10.24 kHz using the decimate function of Matlab. All of the signal processing and analysis was carried out using Matlab software. Data were collected using two different speeds and loads. The rotating speed was obtained using a tachometer and the load by using a torque sensor located between the electric motor and the studied gearbox. Both the speed and the load were kept constant, but a slight natural variation of the load occurred. Accurate mean load and speed values are to be found in Table 3. The load and the speed were measured from the input shaft side.

Table 3. Number of data instances collected for each case.

<table>
<thead>
<tr>
<th>Speed [RPM]</th>
<th>Load [Nm]</th>
<th>0deg</th>
<th>1deg</th>
<th>2deg</th>
<th>3deg</th>
</tr>
</thead>
<tbody>
<tr>
<td>1500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>66</td>
<td>102</td>
<td>51</td>
<td>52</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>101</td>
<td>51</td>
<td>52</td>
<td>52</td>
</tr>
<tr>
<td>1000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>65</td>
<td>101</td>
<td>51</td>
<td>51</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>37</td>
<td>101</td>
<td>51</td>
<td>51</td>
<td>52</td>
</tr>
</tbody>
</table>

3. RESULTS AND DISCUSSION

In this section the preliminary results are first shown using acceleration (g) signal and its Fourier transform.
(Figures 4 and 5). Then the spectra for each data instance are displayed in waterfall plots (Figures 6, 7, 8 and 9) in order to visualize how each misalignment case affects the vibration levels of the frequency components.

Each step of the flow chart (Figure 1) was performed and the results are to be found in Figures 10 – 16 and Table 4. First the SVM was trained using only individual scaling factor values to determine whether it was possible to use only one set of features to detect misalignment when the speed and load varied. Later the SVM was trained using feature sets gotten using the first four scaling factors.

**Visual inspection of signals time and frequency domains**

Typical signals of one data instance and their amplitude spectra are shown in Figure 4 for the case where the axes were aligned, and in Figure 5 for the case where the misalignment was three degrees. Comparing these extreme cases, it seems that the overall vibration level does not change dramatically. Figures 6, 7, 8 and 9 show that the rubber coupling works well and that the frequency components of the rotation frequencies and their harmonics remain almost the same for the healthy signal and for each misalignment case; this is especially the case for low frequencies. However, the change in the behaviour of the system was different for each case, which was easy to detect by listening to the running sound, at least when the misalignment reached its highest point. Moreover, small bits of rubber were detached from the rubber coupling, so that, clearly, the misalignment would have been harmful to the system in the long run. However, with traditional techniques, the misalignment might have been hard to detect. In addition, since the maximum speed for the test rig is 1500 RPM and the ratio is 3.61, the calculated maximum rotation frequency of the output shaft is 6.93 Hz and, when the speed is 1000 RPM, the rotation frequency is 4.62 Hz. Therefore, the functionality of detecting the rotation frequency is on the verge of its limitation. Even though the sensitivity of the accelerometers is 10 mv/g, we are operating under the standard factory calibration limit of 10 Hz. Therefore, the use of the first rotation frequency might be obscured.

**Wavelet analysis and SVM classifier**

According to preliminary observations, it seems that changing both the scaling factor (a) and the centre frequency \( f_c \) of the Morlet wavelet has little or no effect, since doubling the value a and halving the value \( f_c \) lead to almost the same scalogram image. This is in contrast to the findings of Rafiee et al. (2010) who found some minor differences by using different centre frequencies. In the present study, only the bandwidth parameter \( f_b \) and the value a had chosen parameters which were altered. For the \( f_b \) values, 0.1, 1 and 10 were chosen, as shown in Figure 3. For all the cases, the scaling factor values were chosen by taking the mean speed so that the first scaling factor’s pseudo-frequency was centred around the rotation speed. In total ten different scaling factors were used and all of them were centred like the first one, such that the harmonics of the rotation speeds were chosen to be up to ten times the rotation speed.

Before choosing the value a according to the rotation frequency, other a values were also tested to see how well the misalignment could be detected. It was found, surprisingly, that by using some a values not related to the rotation frequencies the misalignment could be detected better than by using the rotation frequencies to calculate the a values. However, these other a values were excluded from the final models and only the first 4 a values were used, which all were related to the rotation speed. The reason for this was that there are no good way of knowing whether other faults would also have an affect and disturb the sensitivity of the final model. Therefore, the functionality of the method was tested only using a values centred around the rotation frequencies which should improve the robustness of the final model.

Since misalignment can be very sensitive to changes in the load and speed, the SVM was first trained by choosing feature sets which consist only wavelet coefficient using
Fig. 6. Acceleration (g) waterfall plot with a speed of 1500 RPM and a torque of 35 Nm.

Fig. 7. Acceleration [g] waterfall plot with a speed of 1500 RPM and a torque of 66 Nm.

Fig. 8. Acceleration [g] waterfall plot with a speed of 1000 RPM and a torque of 37 Nm.

Fig. 9. Acceleration [g] waterfall plot with a speed of 1000 RPM and a torque of 65 Nm.

scaling factors from 1 – 10 individually. All the individual cases are shown in Figures 10 – 16 in the same graph. Mostly a 95 % variance was obtained with 4 features out of total 5. When comparing these figures, it can be seen that some of the models are able to detect the misalignment with good confidence. However, the most sensitive feature set do not remain the same when the load or speed changes and the first four feature sets are rarely 100% accurate. It seems that it is reasonable and necessary to use more than one scaling factor values to make the model more robust.

Furthermore, to test how well the feature set which is chosen using only one a value would work when the load and speed varied, all the data cases were combined into three different sections. Figures 14 and 15 show the results for two cases where the speed remained the same and the load varied, while Figure 16 shows the results for all the cases combined. Again it seems that some of the feature sets work quite well, but they do not show the same performance when the speed changes. In addition, combining all four cases seems to give the worst results, which clearly indicates that the feature set which is based only to one scaling factor value cannot be used individually to detect the misalignment with good confidence. The tests for all the cases shown in Figures 10-16 were performed when \( f_b = 10 \).

To achieve a better and more robust model, the first four feature sets (four different scaling factor values) and all four cases were combined and used together to train the SVM. Moreover, this procedure was carried out by using three different \( f_b \) parameters to determine whether this would have any effect on the final model. Results for these three different bandwidth parameters are to be found in Table 4. Surprisingly, this method was almost 100% accurate, even though individually feature set led to quite poor results. Mostly a 95% variance (see Section 2.2) was obtained by using six dimensions out of the chosen 20 (4 a values and 5 features).

Confusion matrices were used to determine the evaluation metrics of the misalignment classifier. For each class, the following four classification outcome states were employed.
Fig. 10. Accuracy of the SVM models when using only individual scaling factors (a) from 1 to 10. Rotation speed is 1500 RPM and torque is 35 Nm.

Fig. 11. Accuracy of the SVM models when using only individual scaling factors (a) from 1 to 10. Rotation speed is 1500 RPM and torque is 66 Nm.

Fig. 12. Accuracy of the SVM models when using only individual scaling factors (a) from 1 to 10. Rotation speed is 1000 RPM and torque is 37 Nm.

Fig. 13. Accuracy of the SVM models when using only individual scaling factors (a) from 1 to 10. Rotation speed is 1000 RPM and torque is 65 Nm.
Fig. 14. Accuracy of the SVM models when using only individual scaling factors (a) from 1 to 10. Speed is 1500 RPM and torque is 35 or 66 Nm.

Fig. 15. Accuracy of the SVM models when using only individual scaling factors (a) from 1 to 10. Speed is 1000 RPM and torque is 37 or 65 Nm.

Fig. 16. Accuracy of the SVM models when using only individual scaling factors (a) from 1 to 10. All cases (see Table 3) combined.
(1) True positive (TP): The system is healthy and was predicted to be healthy.
(2) False positive (FP): The system is faulty, but was predicted to be healthy.
(3) False negative (FN): The system is healthy, but was predicted to be faulty.
(4) True negative (TN): The system is faulty and was predicted to be faulty.

The accuracy of the SVM model, which is presented in Table 4, could be calculated using the following equation: Accuracy = (TP + TN) / (P + N), where P is the number of positive data instances and N is the number of negative data instances. In all the cases, P and N were randomly selected using the data instances listed in Table 3.

From Table 4 it can be seen that changing the \( f_b \) parameter of the mother wavelet does not affect the results that much. The reason for this might be that the speed was constant without any minor fluctuation around the mean speed. However, it would be interesting to determine whether this would also be the case when there is some fluctuation around the mean speed. Our estimation is that there is a high correlation between the speed variance and the bandwidth parameter.

### Table 4. Confusion matrices when SVM is trained using all data instances with 3 different bandwidth parameters.

<table>
<thead>
<tr>
<th>fc=2, fb=0.1 and scales (a) 1-4</th>
<th>Predicted class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Misalignment=</td>
<td>Healthy</td>
</tr>
<tr>
<td>Healthy Actual</td>
<td>37</td>
</tr>
<tr>
<td>Faulty Actual</td>
<td>0</td>
</tr>
<tr>
<td>Accuracy [%]</td>
<td>100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>fc=2, fb=1 and scales (a) 1-4</th>
<th>Predicted class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Misalignment=</td>
<td>Healthy</td>
</tr>
<tr>
<td>Healthy Actual</td>
<td>38</td>
</tr>
<tr>
<td>Faulty Actual</td>
<td>0</td>
</tr>
<tr>
<td>Accuracy [%]</td>
<td>100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>fc=2, fb=10 and scales (a) 1-4</th>
<th>Predicted class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Misalignment=</td>
<td>Healthy</td>
</tr>
<tr>
<td>Healthy Actual</td>
<td>38</td>
</tr>
<tr>
<td>Faulty Actual</td>
<td>0</td>
</tr>
<tr>
<td>Accuracy [%]</td>
<td>96.7</td>
</tr>
</tbody>
</table>

4. CONCLUSIONS

The features which were extracted by using Morlet wavelet transforms with scaling factor (a) centred around the harmonics of the rotation frequencies and which were classified using the SVM can be used to detect and classify misalignment with great confidence, as seen in Table 4. Varying the bandwidth parameter of the Morlet wavelet seems to have little or no effect when the speed does not fluctuate around its mean value.

Feature sets using only individual scaling factors are very sensitive to changes in the rotation speed and torsional load and cannot be used to detect the misalignment according to this test setup.

5. FUTURE WORK

This research study has only dealt with one failure type, namely shaft misalignment. Future research should investigate whether the model is still valid when other types of failure, e.g. bearing or gear damage, are introduced in the same system. It is only after such research has been performed that it will be possible to classify and validate the sensitivity of the chosen features for individual failure modes with great confidence.

Furthermore, the process parameters, e.g. the load and speed, should vary even more and the speed should fluctuate around the mean value during each measured time segment.
To improve the sensitivity of the method, it could be beneficial to integrate the acceleration signal to obtain the velocity. We can also calculate an infinite number of signals by means of fractional order derivation. However, in the test setup of the present study, it was sufficient to use acceleration.

This type of model can also be used as an input for estimating the remaining technical life of a system. Since sometimes it is impossible to stop the machine under investigation (owing to distance or location, or because it is part of a bigger system) and fix a misalignment, the model can support decision making by giving some estimations based on the nature of the misalignment, even if such estimations may not be 100% accurate.

**ACKNOWLEDGEMENTS**

The authors would like to thank SKF AB, Vinnova and National Instruments for their contributions and support.

**REFERENCES**


