

Model Based Detection and Diagnosis of Bearing Defects

Henk Mol, Antonio Gabelli

SKF Research and Technology Development, Nieuwegein, 3439MT, the Netherlands

henk.mol@skf.com, antonio.gabelli@skf.com

ABSTRACT

In this paper we make use of the Amplitude Spectral Density of Spectrogram (ASDS) (Mol and van Nijen 1997) and describe how its output is post processed for determining the size of defects in bearings using acceleration signals. Bearing modeling and experiments with defective railway wheel bearings are then used to define the lead time between detection and stop of the operation in order to replace the bearing.

Initially, small defects in the order of a 0.1 to 1 millimeter may arise in the running surface of the bearing. This is the moment when vibrations start to be excited by the rolling contact forces and can be observed by accelerometers on the structure of the machine. At this stage most machines are still operating well. The rolling bearing modeling work shows that characteristic bandwidth of the force impulse for defects smaller than the Hertzian length approximates the speed rolling divided by the sum of the length of the defect and length of the Hertzian contact. This simple result has been proven experimentally by running a bearing with a small defect and varying the running speed, and by varying the size of the Hertzian contact by changing the load.

When the small defect grows, the amplitude of the acceleration is growing with defect size but its bandwidth is hardly changing until the Hertzian contact is completely disrupted. At that point the force impulse's bandwidth starts going down as the defect length becomes the dominant factor. This property is used to calculate the actual size of the defect. Eventually when the defect is about 20 percent of the roller spacing in the bearing, the bandwidth starts to become similar to the repetition frequency and the ASDS transform will no longer be able to estimate it. Experimental work with artificial defects proves the size quantification.

The diagnosis with defect size quantification makes future prognosis possible. For small defects, we have proof that the defect's rate of growth is slow (Morales-Espejel and Gabelli 2015), (Rycerz, Olver and Kadiric 2017). When the defect passes a certain size, the erosion process will exponentially

Henk Mol et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

increase the area damaged in the bearing. Using rolling contact fatigue modeling, under the condition of defects in a rolling bearings, it is possible to predict the likely defect growth rate according to the work of Morales-Espejel and Gabelli (2015) and Gabelli and Morales-Espejel (2017). By fitting the estimated defect size to the exponential growth rate curve, the decision to replace the defective bearing can be deferred to the right moment without much risk of immediate failure.

1. INTRODUCTION

1.1. Background

Like in many rotating machines, railway wheel bearings are not monitored yet in order to track their wear. They are inspected for any damage and refurbished with new lubricating grease and if necessary new seals if found free of damage and put back on the rolling stock. The bearing inspection typically is done when the wheels are replaced. Inspection intervals can be optimized see e.g. the work by Saleh and Marais (2006) and ten Wolde and Ghobbar (2013). However, monitoring the bearings is now becoming increasingly attractive due to several developments.

First, there is an increase of wheel life by re-profiling the wheel in a different way than before. Instead of taking 6 mm off and doing this only when form and roundness of the wheels requires it, there is a tendency to go to a fixed but short interval with 1 mm removal typically every 70000 km. Wheel life can be potentially extended from 0.8 to about 1.6 million km (Müller *et.al.*, 2013), which is a doubling of the replacement interval length.

Secondly there are enabling factors. The sinking costs for the monitoring electronics and software, the possibility to have ubiquitous connectivity for mobile monitoring devices, and the advancement of methods to detect and quantify the level of damage from vibrations.

Modern rolling element railway wheel bearings show a nominal life L_{10} in the order of 4 million km. The bearing failure rate, assuming a Weibull distribution with coefficient β of 1.3, see SKF General Catalogue (2018) and the work of Ferreira, Balthazar and Araujo (2003), is then typically ~2 % of the population of bearings within the current inspection

interval of 0.8 million km. When enlarging inspection intervals to 1.6 million kilometers this will grow to potentially ~5%.

If not observed for 1.6 million km, a small damage may have propagated all through the bearing, like the photograph shown in Figure 1.

Obviously it is not straightforward then to double the inspection interval for bearings. Monitoring the wheel bearings may mitigate the problem, under the assumption that the transition between initial damage and final failure is well defined and detectable and offers enough lead time to plan a maintenance action.

SKF is now rolling out autonomous monitoring of railway wheelsets with embedded temperature and vibration measurement. It collects signals, processes them to data sets and connects via the GSM network. In the back office with automated post processing the analysis is done to detect, diagnose problems and report a prognosis. This forms the basis for recommended maintenance.



Figure 1. Inner ring of a failed spherical roller bearing after continuing the machine's operation without condition monitoring applied.

1.2. Objectives for Vibration Monitoring

The detection of bearing damage with vibrations is a well-established field. Experienced maintenance personnel will acknowledge that (depending on the application conditions) bearings are parts that can exhibit a rather slow progression of their damage in use, and often an initial failure is not the only criterion to stop a running machine. The effect of stopping it is taken into account before decisions are taken. Therefore the question is mainly how to quantify the size of a defect in a damaged bearing when running in the application.

Although there are root causes of bearing failure which are separate mechanisms, they lead to mostly spalling of the

running surface. The damage is visible as surface defects in the contact surfaces of rings and rolling elements. Inside the bearing, pieces of material have been removed by fatigue cracking due to the cyclic stress around the roller to raceway contacts. As the bearings are integrated into the machinery it is in principle not possible to directly observe the wear and failure of a bearing, so monitoring the bearing's condition must be done by observing symptoms. The most common method is to observe the mechanical system's response to the changes in the rolling bearing contact forces due to the wear of the bearing by means of vibration, as this is least intrusive on the equipment.

1.3. Bearing Defect Growth During Continued Running

Once a bearing has a beginning defect, the unit can be officially considered as "failed". However, the time between a minor irregularity and an extended damage on the raceways is found to be long enough for many applications to continue operations until a favorable moment for maintenance.

Laboratory testing of taper roller bearings, under a test load giving rise to a maximum contact pressure of 2 GPa, have shown that small defects will grow to a significant size for as much as 20–25 million revolutions during which the defect will double its size every 5 million revolutions quite similar to the ball bearing case observed in the work of Morales-Espejel and Gabelli (2015).

In railway wheel bearing applications, the maximum contact pressure is generally smaller than 1.5 GPa making it likely to have significant lead time after detection of the defect before having to stop operating the vehicle.

2. THEORY OF THE VIBRATION BASED METHOD

2.1. Vibration Baseline

Obviously, when bearings are transferring the load from a rotating component to another component, there will be variations in the load equilibrium within the machine. This causes a vibration as a base line response:

1. Any rotating machine has already a certain amount of harmonic vibrations caused by mass imbalance and geometric tolerances in the components.
2. Perfect rolling bearings will cause some harmonic vibrations: forces are carried by rolling contact areas at a finite number of locations and they move around (Wensing, 1998).
3. In real life, rolling bearings are never perfect. They have been produced with a finite three dimensional geometrical accuracy of the rings and rolling elements, and this is causing additional dynamic forces in the system.

The baseline therefore is not a zero level vibration. The complexity is that dynamic response of the machine is often

not entirely repeatable nor reproducible between identical copies.

Stationarity of a mechanical system in the long run is usually not proven. On the short term it will depend on (amongst others) the load forces themselves, while on the long term ageing of the loaded structure will show up as a drift in some of the dominant modes (Eigen frequencies) and their damping coefficients. Not unusual is to find drift of internal dimensions and fittings leading to changes of internal (pre) loading of bearings.

With regard to reproducibility, the excitation as well as dynamic response is dependent on manufacturing tolerances in the machines and their components. Two copies of a machine design may show significant differences in their vibration response to the same forces, which causes difficulties when monitoring groups of identical copies of machines, e.g. wind turbines in a park. In the rail wheelset area however the reproducibility is found to be remarkably good (based on vibrations recorded by sensors on many similar pilot cases with trains).

2.2. Force and Response Due to Bearing Defects

The time dependent contact forces in a bearing are the main property with regard to vibrations in the presence of defects on the running surfaces. This contact force is normally harmonic in nature, but when the contact zone between roller and ring passes a defect, the pressure distribution will be different from normal, and the force carried through the rolling element changes. If we assume that the forces in the bearing are in equilibrium, then the passing of a contact area over a defect will cause a temporary change of the force equilibrium. The force $F_r(t)$ carried by the rolling element will then become a sum of the stationary force of the bearing without defect and a force impulse $\Delta F_r(t)$ due to the defect. For geometrically rectangular shaped defects it has been found that the load on the rolling element is changing during a time $T_{h,max}$ equal to:

$$T_{h,max} = \frac{2b + L}{v} . \quad (1)$$

In this formula L is the length of the defect in the direction of rolling (m), v is the speed with which the contact area moves through the defect (m/s), and $2b$ is the semi-minor axis of the projected contact ellipse (m).

Assuming a sum of step functions that create this pulse, the force pulse' spectrum $S(\omega)$ will be:

$$S(\omega) = -\Delta F_r T_h \frac{\sin\left(\omega \frac{T_{h,max}}{2}\right)}{\omega \frac{T_{h,max}}{2}} . \quad (2)$$

The characteristic time $T_{h,max}$ of the force impulse is reflected in the frequency domain by a regular series of "notches" on

the characteristic frequencies $f_{h,min} = n/T_{h,max}$. where n is a positive integer. At these particular frequencies, no vibrations will be observed.

The simple model above is stating the maximum time for T_h . In the bearing the force will not change with a step, but more gradually when the Hertzian contact moves into the defect.

For defects much smaller than $2b$ the time is obviously dominated by this Hertzian contact length itself. The validity of the simple approximation of the force $F_r(t)$ can be proven experimentally using a scale model ball bearing using a small defect. Changing speed v will show whether the notch is proportionally following the speed, and changing the load will change the Hertzian contact length $2b$ where an increase of load will decrease the first notch frequency f_h .

When the defect grows from well under a length of $2b$, the rolling element which is at the defect has a decreasing contact force with increasing defect length, until the rolling element is no longer keeping even partial contact with the ring surfaces. An example of the quasi-static equilibrium roller loads has been calculated numerically for a railway wheel bearing system using two SKF 229750 double-row spherical roller bearings in an Y32-R18 axlebox and is shown in **Figure 2**. Each bearing in this arrangement is carrying 26900 N. The roller force is decreasing from nominal to zero with increasing defect size until $L = 4b$. In other words, the amplitude of the force pulse will no longer grow after the roller has lost contact. In the small defect size regime therefore, the amplitude is a potential measure of the size of the defect, in the large defect regime the duration of the force pulse is the measure of the size.

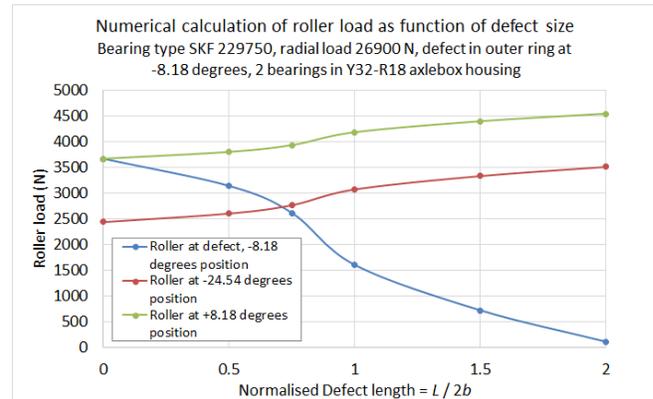


Figure 2. Numerical modeling result of roller load F_r versus relative defect length $L/2b$ for bearing type SKF 229750.

The bearing modeling also shows that the rollers in the vicinity (as well as the second roller row in the bearing, not shown in this graph), take over the load in the quasi static equilibrium. The outer ring moves closer to the inner ring when the defect reduces the contact force on the roller at the defect and thus keeping the roller in partial contact. For defects larger than $4b$ the force function can be approximated

by a block impulse negative load at the position of the defect in the bearing equal to the nominal roller load. This approach is also found in the work from Sawalhi and Randall, (2011), and Moazen-Ahmadi and Howard (2016). A simplified model for the purpose of estimating pulse time T_h is proposed for the roller load. It assumes linear change of net contact load when the Hertzian contact is moving into the defect, as well as out of the defect. The interaction starts when the centre of the Hertzian contact ellipse is at position $x = -b$ and ends at $x = L+b$. Taking into account that the rings are moving closer to each other, which keeps the roller partially in contact with the edge of the defect, the force can be summarized as:

$$\begin{aligned}
 F_r(x) &= F_{r, equilibrium} \quad \text{for } x < -b \text{ and for } x > L+b \\
 F_r(x) &= F_{r, equilibrium} \left(1 - \frac{x+b}{3b}\right) \quad \text{for } -b < x < 2b \\
 F_r(x) &= 0 \quad \text{for } 2b < x < L-2b \\
 F_r(x) &= F_{r, equilibrium} \frac{L+2b-x}{3b} \quad \text{for } L-2b < x < L+b
 \end{aligned} \tag{3}$$

The modeled relative contact force is shown in **Figure 3**.

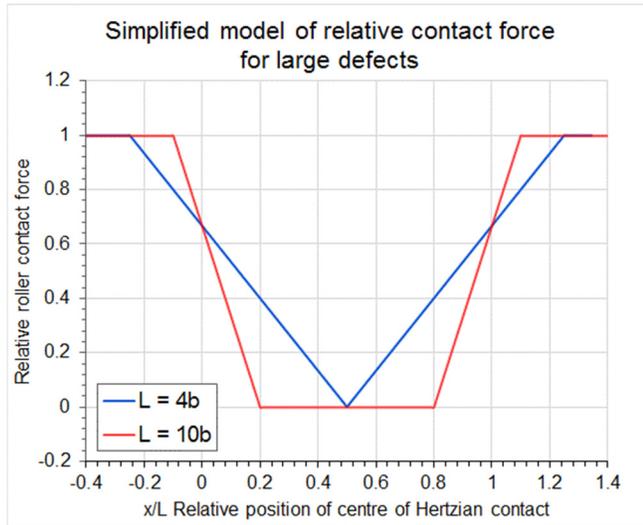


Figure 3. Simplified model of the relative roller load used for explaining net pulse time T_h

The inertia of the system (bearing and axlebox housing with suspension spring and damper) will dominate the forces of the system at high speeds, but as only few cases were numerically analyzed the detailed results will not be discussed in this paper. Getting out of the defect will lead to a short duration force which is several times the static roller load for this particular bearing situation depending on mass and stiffness. This dynamic result is not taken into account for the simplified model, which intends to explain the net pulse time $T_h < T_{h,max}$, and its corresponding notch frequency f_h .

2.3. Description of the implemented Amplitude Spectral Density of Spectrogram

2.3.1. Spectrogram Setting

To reconstruct the notch frequencies (the local minima in the spectrum) from a repetitively occurring (i.e. cyclo-stationary) pulse force in a noisy mechanical structure with several force sources, the spectrogram and its amplitude spectral density is used. The method has been first implemented inside SKF for this purpose in 1991 (van Nijen, 1990) and has been published as a patent in 1997, see the work of Mol and van Nijen (1997). The method is also known from later publications, see amongst others the work of Antoni (2007) and Antoni, Abboud and Xin (2016).

First step is to set up the STFT (Short Time Fourier Transform) based spectrogram part, in order to calculate the amplitude as function of frequency and of time. The choice of window and the overlap of subsequent intervals of the time signal will be explained.

Assume that the vibration signal $x(t)$ is sampled and therefore time discrete. It has a sampling time ΔT , and the sampling frequency ω_s is equal to $2\pi/\Delta T$. The short time average amplitude $A(t)$ of the vibration signal $x(t)$ can be calculated as function of time t over a finite interval $T = N\Delta T$. In order to be meaningful, this interval T is smaller than the repetition time T_b of the repetitive excitation by the defect in the bearing or vice versa its repetition frequency ω is less than $2\pi/T$. The amplitude $A(t)$ is the root of the power $P(t)$

The STFT calculates the functions $C_n(t)$ which are the complex Fourier coefficients determined by the discrete Fourier transformation at time t and determined over the short time interval T :

$$C_n(t) = \sum_{m=0}^{N-1} (x(t + m\Delta T) w(m\Delta T) e^{-j\omega_1 m n \Delta T}). \tag{4}$$

The angular frequency ω_1 is the frequency resolution of the discrete Fourier spectrum thus constructed:

$$\omega_1 = \frac{2\pi}{T}. \tag{5}$$

The function $w(t)$ represents the window function. The window function $w(t)$ is zero for $t < 0$ or $t > T$.

Subsequent calculation of the spectrum $|C_n(t)|$ is not done for every time step ΔT . This is depending on the window's ability to reduce aliasing and is explained in the next paragraph.

2.3.2. Spectrogram Window Function Choice

The window function is to be optimized for the next step, which is to calculate the amplitude spectrum of the spectrogram. A different choice has been made in this implementation based on the work of Mol and van Nijen (1997) with respect to the choices made in Antoni (2007).

The amplitude $A(t, \omega) = |C_n(t)|$ of the coefficients represents the amplitude as function of time and as function of the discrete frequency $\omega = n \omega$. The amplitude $A(t, \omega)$ does not have to be recalculated for each discrete time point $t = i\Delta T$, first of all because it is only interesting to observe the low frequency variation of the amplitude A and secondly because the window function w , which is in front of the Fourier transform, is acting as a low pass filter for the amplitude estimation. The window function $w(t)$ will lead to aliasing however if the coefficients would only be calculated with a periodic time equal to $T = N\Delta T$ because the characteristic impulse time of the window is shorter than T . The optimum time interval between subsequent Fourier transformations depends on the impulse response of the window used in the spectrogram calculation. The impulse response H of any amplitude $A(t, \omega)$ equals:

$$H(t) = w(t) \sqrt{\frac{1}{N}}. \quad (6)$$

The frequency spectrum of this impulse response has the same contents as the amplitude density spectrum of the window function $w(t)$. The subsequently calculated spectra will therefore need to have an overlap in the input data. How much overlap is needed, will be a result of a compromise between aliasing error and calculation effort. Generally the spectrum will be calculated at times $t = d \Delta T$, and the output sampling frequency of the coefficients $|C_n(t)|$ is equal to the original sampling frequency divided by d . The value d is therefore called the down-sampling rate. The parameter d is an integer between 1 and N , and the part of the data segments which is overlapping the previous one will contain $N-d$ samples.

For the Hanning window, it can be shown that the down-sampling rate d causes an aliasing error less than 1 percent when $d \leq 5N/32$. But, a 1% error is inadequate for the reconstruction.

A better choice is a Modified Bessel window $w_b(t)$ as it trades off between frequency resolution and aliasing suppression in the spectrogram. The zero order modified Bessel function $I_0(z)$ with the following definition has less than 0.02% aliasing error and therefore this window is implemented in the spectrogram part:

$$w_b(t) = I_0(z). \quad (7)$$

The parameter z in Eq. (7) equals:

$$z = B \sqrt{(1 - (2\frac{t}{T} - 1)^2)} \quad \text{for } 0 \leq t \leq \frac{T}{2}, \text{ and} \quad (8)$$

$$z = B \sqrt{(1 - (2\frac{T-t}{T} - 1)^2)} \quad \text{for } \frac{T}{2} < t < T.$$

The shape parameter $B=9.6$ has been chosen to provide a good compromise between frequency resolution and the

achievable down-sampling rate d in relation to the total error due to aliasing.

As the window with modified Bessel function had to be implemented in a table for a microcontroller of an embedded condition monitoring device the function was approximated with a standard series expansion:

$$I_0(z) = 1 + \sum_{n=1}^{\infty} \frac{(\frac{z^2}{4})^n}{(n!)^2}. \quad (9)$$

2.3.3. Amplitude Spectral Density of Spectrogram

The spectrogram, which is the collection of regularly spaced time-frequency signals $A(t, \omega)$, represent the amplitude of the vibration signals as function of time and frequency with R spectra, each with $N/2$ lines. It can now be transformed to the discrete frequency-frequency domain for the identification of the bearing by its repetition frequencies ω_r over the finite time interval $R d \Delta T$:

$$A(\omega_r, \omega_n) = \left| \sum_{m=0}^{m=R-1} |C_n(m d \Delta T)| w_r(m d \Delta T) e^{-j m d \Delta T \omega_r} \right|. \quad (10)$$

In this, the repetition frequency $\omega_r = \frac{r \omega_s}{d R}$ and r and R are integers such that

$$0 \leq r < \frac{R}{2}. \quad (11)$$

where ω_s is the sampling frequency of the time signal $x(t)$, and d is the down-sampling rate between original sampling frequency and the sampling frequency for the spectrogram output.

The window function w_r for the repetition frequency domain can now be any suitable standard window. In this case the Hanning window is used for w_r :

$$w_r(m) = \frac{1 - \cos(-\pi + \frac{2\pi m}{R})}{2}. \quad (12)$$

Note: The amplitude spectral density of the spectrogram is abbreviated to ASDS, in order to keep the text compact. It is similar but not entirely equal to the cyclo-stationary analysis presented in Antoni, Abboud and Xin (2016).

The result of the calculation is thus an amplitude with repetition frequency and signal frequency as variables. Graphically, it is convenient to plot this in a height map with colors. With a repeating impulse as time signal, the correctly constructed map will show vertical lines where the signal energy is concentrated.

The important feature is that any cross section on a specific repetition frequency shows an amplitude spectral density of the repetitive phenomenon at that specific repetition frequency. It does not lead to any other sum or difference frequencies when there are more vibration excitation sources with different repetition frequencies.

2.3.4. Post Processing of the ASDS

There are three meaningful cross sections which are used to detect defects and estimate their size as graphically shown in **Figure 4**:

- The cross section at the repetition frequency, in order to show the spectral density of the phenomena belonging to a bearing defect. Taking the cross section of $A(\omega, \omega_n)$ at the particular repetition frequency $f_{r, \text{bearing defect}}$ is showing the excitation response amplitude spectrum for $N/2$ frequencies.
- The RMS summation of the random vibration signals between two repetition frequencies. It is a spectral density with the same frequencies as the cross section at the defect frequency, but the amplitude is dominated by random vibrations. This cross section is used to calculate the signal to noise ratio. However as explained in section it is also a hypothesis for blind estimation of the frequency transfer function.
- The cross section over one or the RMS summation over a certain number of frequencies ω_n is representing the amplitude spectral density of the signal for all possible repetition frequencies within a certain band. In fact, this is representing the “envelope spectrum” as obtained by more classic filtering-rectifying bearing condition monitoring methods. However now the band selection is just a post process part, and not a time domain filter operation.

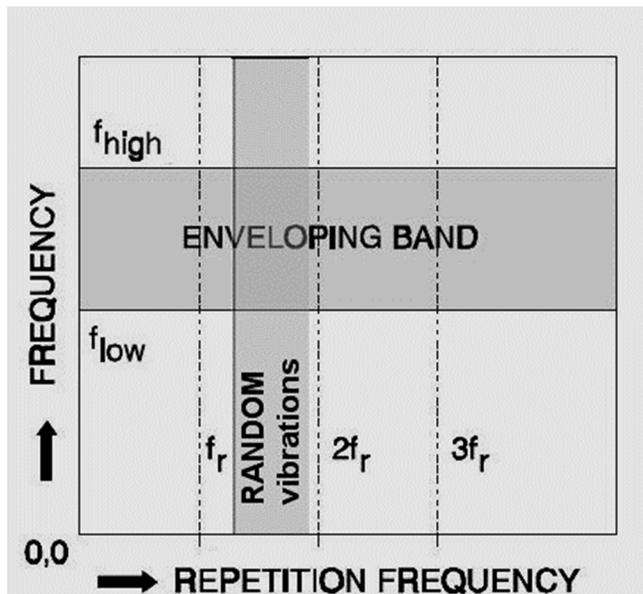


Figure 4. Three cross sections of the ASDS output as post processing steps: vibrations at the repetition frequency f_r due to the defect in the bearing, a band selected for a random vibration spectrum, and the band selection for envelope spectrum.

2.3.5. Post Processing to Signal-To-Noise Ratio

There are two practical reasons use the ratio of two spectra, that is the ratio of RMS amplitude at the defect frequency, and a spectrum at a random set of frequencies.

First of all, it is practical to identify the range of frequencies for a classic envelope often realized in embedded form in monitoring real-time hardware. Enveloping only makes sense if the vibration for the bearing defect is not under the mechanical noise floor.

Second, if the impulse response $H(\omega)$ could be estimated from the random vibrations as an amplitude spectrum $|H_{noise}(\omega)|$, this might be used to divide the response by this transfer function. The risk is however two fold. First, the transfer function $H(\omega)$ may have be not smooth at all, and in particular dividing by small random values will give problems. Second, the source of random vibrations is not necessarily the bearing and consequently not the same vibrational modes may be observed. These questions will be addressed by experimental means.

2.3.6. Post Processing to an Envelope Spectrum

A very useful cross section of the ASDS is the Envelope Spectrum, and it is used to detect and identify the frequencies where the cross section for the defect frequency is located, after which the spectrum of the response to the repetitive force pulse can be extracted. Enveloping is used frequently as it is relatively simple, achievable with low effort, and is conservative in its energy use in embedded autonomous sensing implementations. Both hardware and software implementations are commonly found.

Originally the enveloping method is a time signal processing which consists of three stages:

- Stage 1 is a bandpass filter, which has to select only those frequencies where there is a significant signal from the defect above the background vibrations. Let this filter act from frequency f_1 to f_2 . A common industrial choice is a band between 500 Hz and 10000 Hz.
- Stage 2 is the power calculator: this is in the ideal case made by using taking the square root of the sum of the square of the signal and the square of its Hilbert transformation (see Oppenheim and Schaffer, 1975 and Rice, Venkatachalam and Wegmann, 1988). In practice, it can be approximated using an absolute value (full wave rectifier) or root-mean-square function. The output frequencies of this stage will contain frequencies from zero to at least 2 times f_2 and the wanted demodulated signal components will be at frequencies near zero.
- Stage 3 is the low pass filter, to pass only the frequencies which represent the demodulated signal. After this low pass filter, a (re)sampling can be done with a frequency smaller than the original bandwidth f_2 , but larger than

twice the highest repetition frequency of the vibrations caused by the defective bearing.

Generally the output of stage 3 is processed into an amplitude spectral density, called the Enveloped Spectrum. It shows with which frequencies the phenomena are repeating and thus can be attributed to specific components based on their characteristic kinematic frequencies.

In fact, the ASDS method and the enveloping in the classical sense can be compared directly by means of the Enveloped Spectrum $|A_{env}(\omega_r)|$ which is approximated by taking the root of the summed powers from the N_1 -th frequency line (at N_1 times ω_s) to the N_2 -th frequency in the ASDS:

$$|A_{env}(\omega_r)| \sim \sqrt{\sum_{n=N_1}^{n=N_2} (A(\omega_r, \omega_n))^2} \quad (13)$$

3. EXPERIMENTAL VERIFICATION OF NOTCH FREQUENCY

3.1. Experiments with Scale Model Bearings

Initially, the ASDS method to estimate defect sizes by use of the notch frequency is verified by tests with a line shaped artificial defect of 0.052 mm wide in the outer raceway of a type SKF 6207 deep groove ball bearing. The defect has been made using spark erosion. The machine used to test the bearings is a standard life testing equipment type R2 (Harris, 2001). The vibrations were measured using a B&K type 4374 accelerometer on the housing of the bearing in the direction of the load. The bearing is lubricated with oil (Shell TT40).

The first experiment changes the speed v . Three speeds are used: 495, 3060, and 6100 RPM. The bearing is loaded with 4 kN in the radial direction. Calculation of the Hertzian contact size with (Harris, 2001) resulted in $2b = 0.479$ mm. At 495 RPM (i.e. 8.250 Hz inner ring rotation frequency), the rolling surface speed v is 0.633 m/s. The pulse time T_h becomes 0.798 ms and the first notch frequencies should appear at 1.25 kHz intervals or 152 orders of the shaft rotation.

It is most clearly visible when the modeled spectrum of the force impulse is directly compared to the observation for 495 RPM as shown in Figure 5.

The pulse spectra are collected for the three speeds in Figure 6. Local minima are roughly lining up in a pattern with an average spacing at about 167 orders, although the spectrum looks very capricious for 3070 and 6100 RPM. This is about 9% larger than the predicted 152 orders.

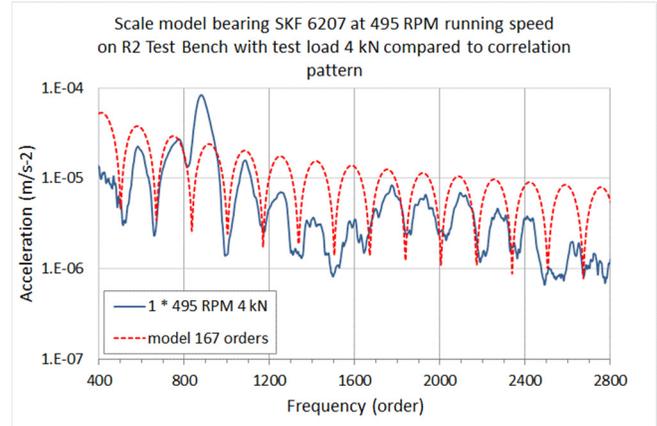


Figure 5. Actual observed response spectrum (blue) compared with spectrum of the model of the force spectrum (red, dashed) based on Eq. (3).

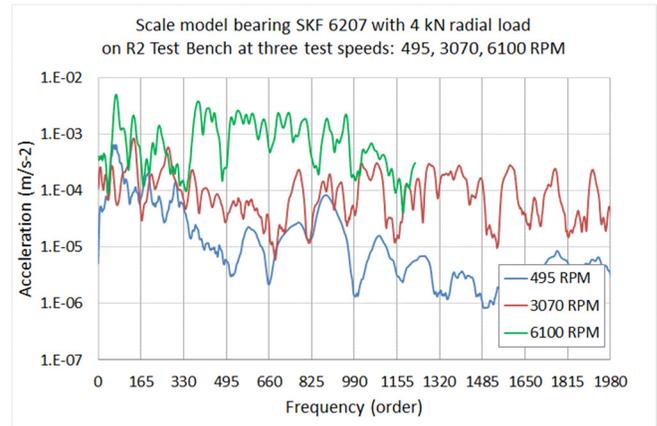


Figure 6: response spectra observed from a test with a very small defect in a DGBB type 6207 for three speeds (all spectra have the same vertical scale).

Three different loads yield three different Hertzian contact sizes, and shall lead to corresponding predicted notch frequencies as in Table 1.

Table 1. Notch frequency as function of load at 495 RPM.

Bearing Load (kN)	$2b$ (mm)	Notch frequency (orders)	
		Prediction	Experiment
1	0.302	228	285
4	0.479	152	166
9	0.628	119	125

The result of the experiment at 495 RPM is shown in **Figure 7**. Note that the signals are scaled such that the graphs are clearly visible. A correlation fit with an absolute sine function has been done to estimate the experimental values for the notch frequency which are shown in Table 1. The result at 4 and 9 kN are deviating less than 10% from the prediction.

This can still be explained by assuming less radial bearing clearance in mounted condition than presumed with the analytical calculation. However the 1 kN result is 25% larger than predicted by Eq. (1) and (2). The ball load at the defect seems to be significantly below the expected ball load and consequently the value for $2b$ seems to be overestimated.

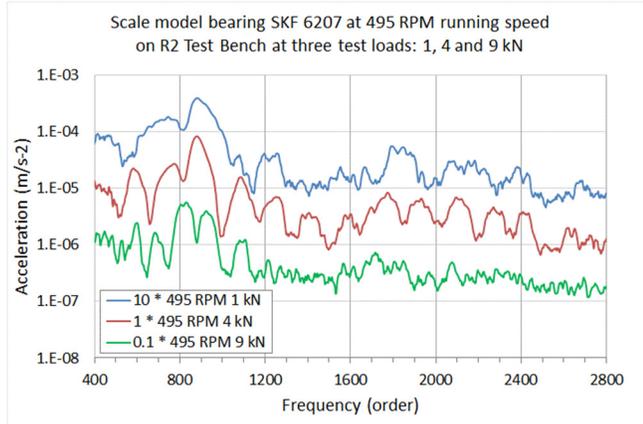


Figure 7. Response spectra for 495 RPM and three different loads. Note that the signals are scaled such that the individual graphs are clearly visible, and the noise dominated frequencies (below 400 orders) are left out.

The tests do show that the ASDS is useful for the estimation of the duration of the force pulse, and thus can be used to estimate the size of a defect.

3.2. Experimental Verification of Defect Size Estimation with Railway Bearings

3.2.1. Set-Up for Experiments with Rail Bearings

The determination of the defect size is to be verified by experiments. In this case, a simple bearing system is chosen with a practical application, being a railway axlebox, on a test machine for railway bearings that has programmable load and speed capability. This set-up is shown in Figure 8.



Figure 8. Test setup for evaluating railway wheel bearing defect size estimation: R3 endurance test machine with Y32-R18 axlebox housings, primary springs, and SKF 229750 railway bearings inside the axlebox.

It consists of a stiff shaft which is belt driven by an electrical motor, and which is supported by two spherical roller bearings type 23024. The test bearings are mounted on each side of the machine at the end of the shaft. The test bearings are fitted inside the axlebox housing which in its turn is mounted inside the suspension parts of the train. The wheel suspension chosen is a variant of a bogie construction, type Y32-R18. It has a simple arm which pivots around a rubber bushing annex bearing that is mounted on the bogie, and has a set of primary springs and a damper. The test arrangement on the machine has left out the damper, because the movement on the train is many centimeters (this is the main function of this so-called primary suspension system) while on the test machine there is hardly any movement (less than a mm). The spring and its auxiliary components (rubber slabs, steel plates) however are important as it does add stiffness as well as a realistic mass and damping for the vibrations that are caused by the bearing's forces when running through the defect.

The load is applied by a hydraulic system pushing at the top of the suspension springs. In this case, the standard load for the chosen test bearings is set to 55 kN i.e. each bearing in the axlebox is subjected to 27.5 kN radial load. The test conditions are typically a simulation of a driving train, with speeds from 30 km/h (quasi static behavior for the bearing system) to 160 km/h (behavior will be depending on the inertia), and loads are kept constant with controlled hydraulic actuators.

The bearing used is a common railway wheel bearing of SKF, type 229750 (outer diameter 220 mm, inner diameter 130 mm, width 72 mm). This spherical roller bearing has two rows. In each axlebox there are two such bearings.

The main geometric parameters are summarized in Table 2 and the calculated frequencies and speeds are summarized in Table 3.

Table 2. Key geometric properties of the test bearing.

SKF 229750 Bearing property	Value
Size (ID x OD x Width)	130 x 220 x 73 mm
Number of rolling elements per row	22
Roller diameter	20.5 mm
Roller set pitch diameter	178.8 mm
Contact angle	11.6 degrees
Distance between subsequent Hertzian contacts of rollers on outer ring running surface	28.5 mm

Table 3. Key frequencies and speeds for the test bearing.

SKF 229750 frequencies at 1 Hz shaft rotation frequency	Value
Inner ring defect frequency	12.236 Hz
Outer ring defect frequency	9.764 Hz
Roller defect frequency	8.611 Hz
Bandwidth for $L+2b = 1$ mm	277 Hz
Roller surface traveling speed v	0.2773 m/s
Equivalent train speed assuming 0.920 m wheel diameter	10.405 km/h

3.2.2. Defect size experiments with railway bearings

The experiments have been done with bearings in which the outer ring has been foreseen with an artificial defect. The bearing defect causes the roller to lose its load when moving through the defect. Numerical quasi-static calculation of the bearing's elastic deformation assuming normal radial clearance and pure radial load shows that the contact force on the most loaded roller comes down from 15 kN without any defect to zero when the defect's length L reaches about $4b$. With a $2b$ of 0.55 – 0.60 mm, the defect length L must be approximately $4b$ i.e. 1.2 mm before the Hertzian contact is entirely lost

Measurements using a bearing with one roller missing and running at very low speeds (10 – 20 km/h equivalent) confirm a housing rigid body movement with regard to the shaft as predicted by the quasi-static calculations. An observed periodic movement of about 7 micrometer could be proven. Dynamic tests showed the first dominant resonance frequency of the system at about 255 Hz and operating as a tilt mode of the axlebox.

In the outer oriented bearing there are artificial defects made with a grinding machine to simulate cases of these railway bearings for various states of spall size, defined by the different lengths of defects. The defects made have a width of 75 % of the contact in the axial direction of the bearing, and using a series of different defect lengths from small (mm size) to very large (effectively the distance between two subsequent rollers). An example of such an artificial defect (of 2 mm) is shown in **Figure 9**.

From these measurements a few cases are used to verify size estimation results by calculating and comparing excitation response spectrum and normalized response spectrum with the predicted minima in the force spectrum. The experimental verification is done with 3.5 mm and 7.5 mm sized defects (in the rolling direction), where the 3.5 mm should have well defined spectra while the 7.5 mm is large enough to cause the bandwidth of the force to be smaller than the resolution of the amplitude spectrum of spectrogram method.

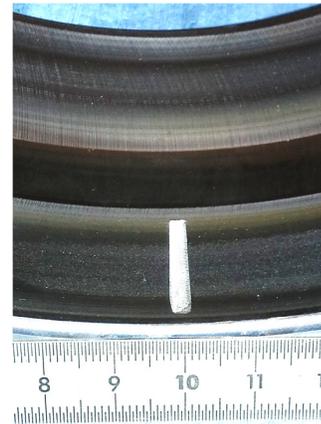


Figure 9. Artificial defect in outer ring of type SKF 229750 test bearing with 2 mm length in the rolling direction. The other defects mentioned are similar to this example.

3.2.3. Response spectra for three sizes of defects

With a 3.5 mm respectively 7.5 mm and 10 mm size defect in a test bearing, the acceleration signals have been observed for running speeds varying from 30 to 160 km/h equivalent speeds. The amplitude spectral density of the spectrogram at 30 km/h is shown in Figure 10 and Figure 11 for respectively the 3.5 and the 7.5 mm defects.

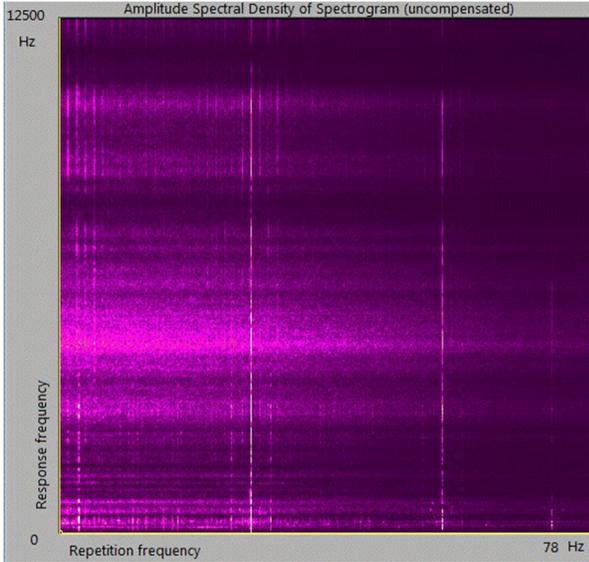


Figure 10. ASDS graph at 30 km/h equivalent, 3.5 mm defect. Note the bright lines appearing at the defect repetition frequency (~28 Hz)

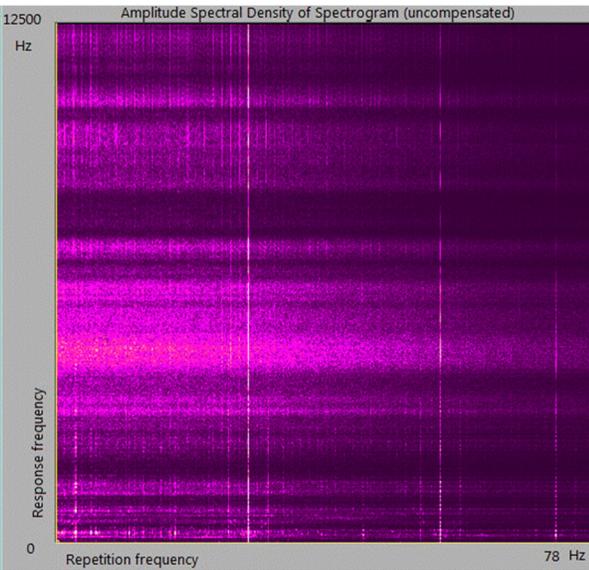


Figure 11. ASDS graph at 30 km/h equivalent, 7.5 mm defect

These graphs show brightly the amplitude of the repetitive impulse responses to rolling through the defect, at 30 km/h this has a repetition frequency of about 28 Hz. Close to the main vertical lines are modulations notably caused by small phase and amplitude differences of the contact forces due various causes, ranging from cage pocket distance variations, roller diameter variations, to not perfect running of the shaft (which has in this test roughly 10 micrometers run-out) and remaining misalignments that cause slight speed difference between the two rows in bearings. At the far right side in the

graphs, the aliased third harmonic of the defect frequency is still visible.

The horizontal “curtains” in the graphs are vibrations caused by random forces and non-repetitive forces.

The response spectra from the acceleration signal are calculated by making the cross section at the defect frequency in each running speed and defect size case. This is shown for the 3.5 mm defect in Figure 12. The first notch is found at 195 Hz, which corresponds to the formula for the largest possible pulse time $T_{h,max}$. The simplified force model would predict $f_h = v/(0.92L) = 248$ Hz for $L=3.5$ mm. Comparing a spectrum of the simplified force model such that the notch frequency is 216 Hz results in the best fit, this would estimate the defect length L for the 3.5 mm case as being 4.0 mm.

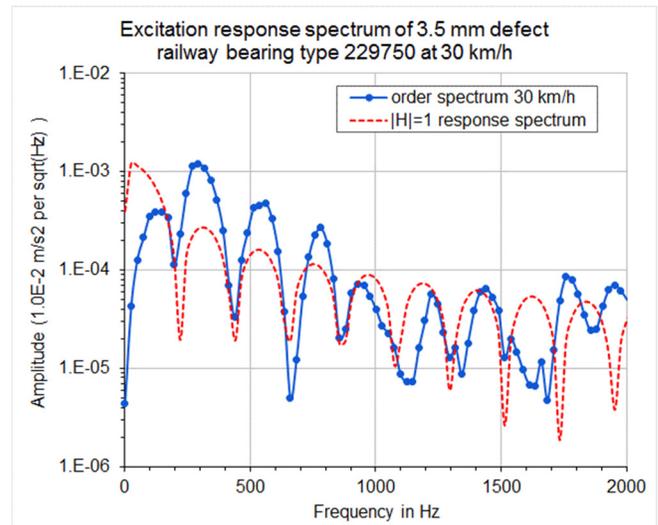


Figure 12. Response spectrum for the 3.5 mm outer ring defect running at 30 km/h equivalent speed.

When the data is shown for various speeds and plotted in orders, it can be shown that the response spectra are having the same local minima, which of course depend on the size of the defect and not on the speed of running. Figure 13 shows order spectra for 30, 50 and 70 km/h test situations and the minima are not very far from each other.

In the field application it is however necessary to have also measurements at real cruising speeds of passenger railway wheelsets which are typically in the 100 – 160 km/h speed range. To observe how well the method is able to follow these speeds are used in the test. The result for the 3.5 mm defect is shown in **Figure 14**. What is clear is that the mass and stiffness of the bearing system is having an impact on the estimate. Between 70 and 120 km/h, the results are spreading more than 30%, mostly to the “short” side of the defect size estimate (i.e. estimate of the bandwidth is too large).

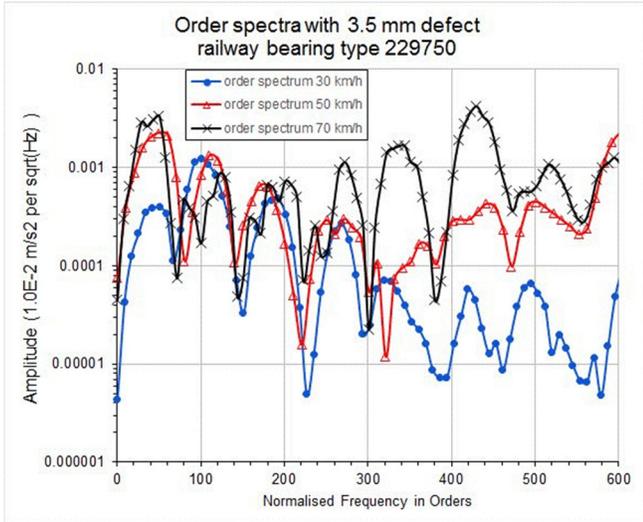


Figure 13. Example order scaled response spectra for three test speeds, 3.5 mm defect size.

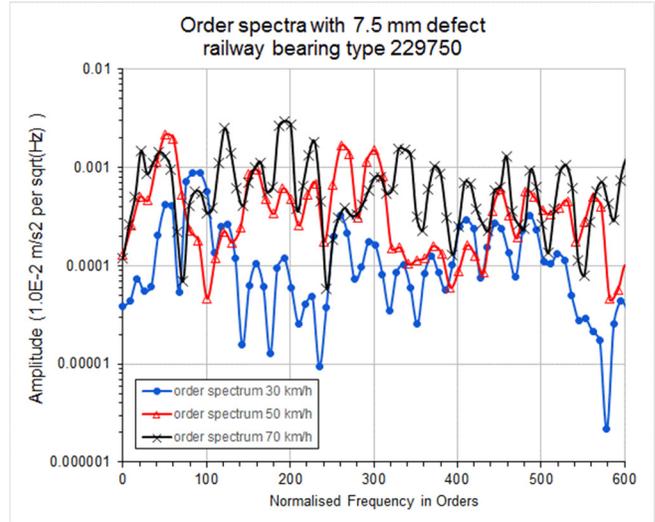


Figure 15. Order spectra for the 7.5 mm defect observed at 30, 50 and 70 km/h equivalent running speed.

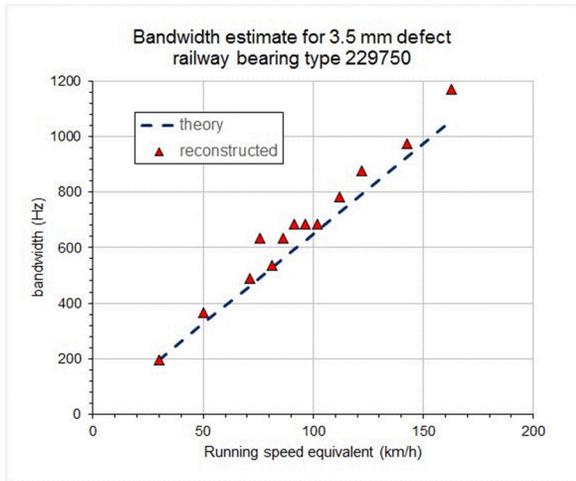


Figure 14. Frequency of the first local minimum as function of running speed.

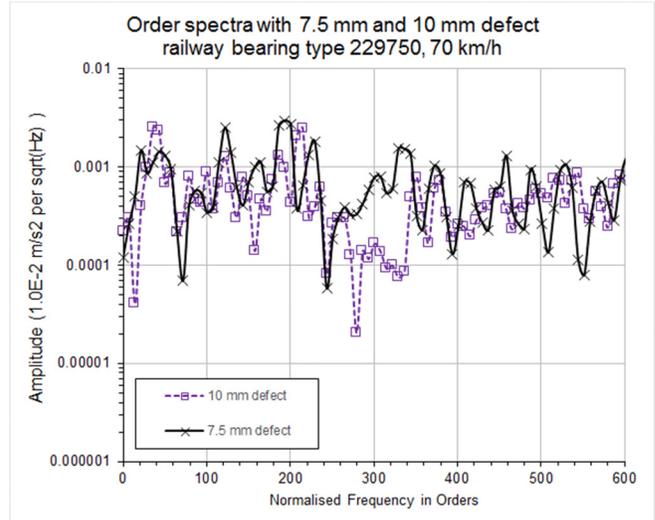


Figure 16. Order spectra of 7.5 and 10 mm defects

The 7.5 mm defect case is less clear to resolve with the amplitude spectral density of the spectrogram. The order spectra of the response spectra at the defect repetition frequency are shown for the most clear cases only, i.e. 30, 50 and 70 km/h, see **Figure 15**. The local minima are not always possible to find, especially the first local minimum is lost for all speeds. The second minimum is also lost for 50 km/h while it is not highly certain that the third minimum for 30 km/h exists. **Figure 16** shows the order spectra of 7.5 and 10 mm defects at 70 km/h, and clearly the result at 10 mm is blurred to the point that it is not possible to have any reliable conclusion about the size of the defect from the minima in the response spectra.

4. DISCUSSION

4.1. Defect Size Determination using ASDS

When observing the order spectra in particular, the variation of signal level can be spanning several orders of magnitude. This complicates the task of finding and judging the local maxima and local minima, which give insight into the potential defect size. In order to make post processing possible to automate the defect size estimation, two questions must be resolved.

1. What is the influence of the mechanical transfer function between the force F inside the bearing, and the acceleration signal observed on the outside of the housing? After all, a mechanical response depends on mass, stiffness and damping and usually many mode shapes are possible. These will lead to local

extremes, both maxima and minima, that may add to the actual local extremes to be found and become a source of uncertainty in the judgement.

2. What is the difference between response spectra of artificial defects, where the modeling and experiment seem to agree reasonably, and the over-rolling of natural spalls? After all, the natural defects may have a different geometry.

4.1.1. Blind Estimation of the Transfer Function

The hypothesis for the estimation of the amplitude of the transfer function is that it is equal to the observed random excitation response $|H_{noise}|$ at a different repetition frequency $\omega_{r,noise}$. The frequency response can be divided out:

$$|A'(\omega_r, \omega_n)| \sim \frac{|A(\omega_r, \omega_n)|}{|H_{noise}(\omega_r, \omega_n)|} = SNR(\omega_r, \omega_n) \quad . \quad (14)$$

There are several problems with this:

- First there must be proof that $|H_{noise}|$ is in principle equivalent to the frequency response function. Assumed are that the random force F_{random} is from the bearing, which in some single-bearing systems like bearing vibration quality control instruments was evidently true. This does not mean however that it is valid for a multiple bearing system, such as in the test shown here. In that case there will be several sources adding, not just the bearing that has a specific defect.
- The random force source may have a specific spectral distribution. If it has a non-uniform excitation spectrum (not a white spectrum) the resulting estimate of $|H_{noise}|$ is in fact a product of F_{random} and the actual different transfer function $|H'_{noise}|$. They cannot then be seen as separate.
- The defect signal itself has randomness in its timing and amplitude. Subsequent force pulses will yield slightly different spectra, and will generally not repeat. This becomes visible as noisy background in the ASDS map, and due to its large power compared to other random sources it can easily blot out the noise response spectrum originally looked for.

A simple experiment illustrates this problematic way of ad-hoc compensation. Figure 17 shows the result of the normalization of order spectra with the 3.5 mm defect by dividing them with the background noise spectra. Instead of becoming better shaped, Figure 17 actually is less clear than Figure 13. More local minima appeared and the variation in defect size estimation increases.

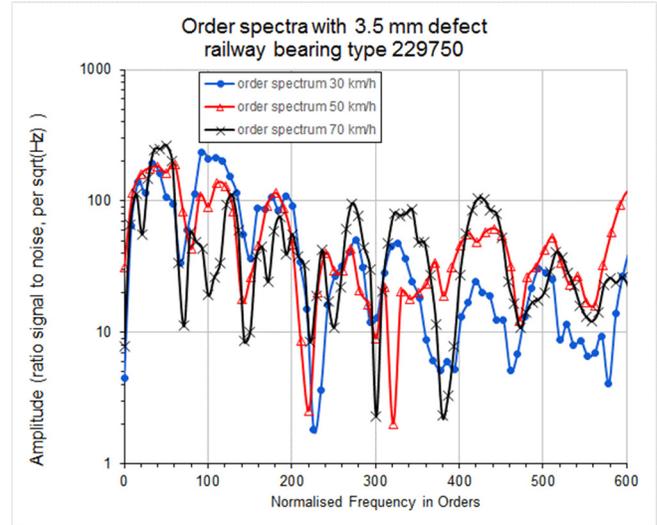


Figure 17. Result of division of order spectra by the random noise response spectra

The estimate from the ASDS can also be compared with a direct estimation of the transfer function. First the frequency response function due to a defect in the bearing has been estimated by means of a run-up/run-down experiment with the 3.5 mm defect. The frequency spectrum of the time signal is calculated for many closely spaced machine running speeds. This yields a collection of spectra that are used to reconstruct the actual transfer, assuming that for each harmonic the displacement value is the theoretical level times the transfer function at that frequency. To illustrate the process, one of the spectra is shown in **Figure 18** highlighting the result at 30 km/h equivalent train speed.

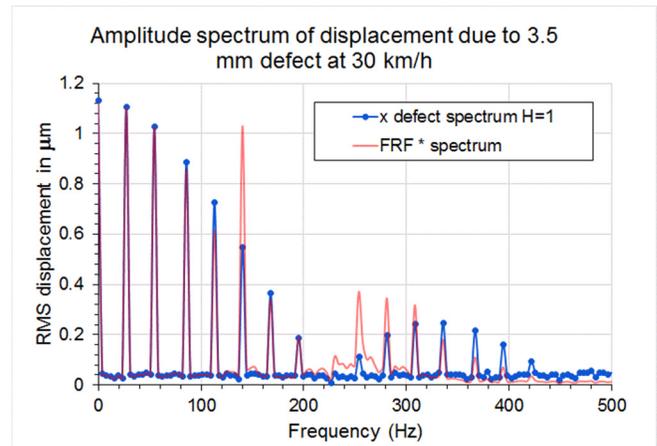


Figure 18. In Blue the modeled “ideal” displacement amplitude spectrum (with transfer function $|H|=1$) and in red the actual observed spectrum due to transfer function

In blue, the calculated spectrum is plotted for $|H|=1$. In red, the actual result is plotted from the measurement at 30 km/h.

Repeating this for many subsequent speeds makes it possible to estimate the transfer function $|H(\omega)|$.

To compare it with the ASDS result, the random vibration was observed for a number of speeds and averaged. The result of both is plotted in Figure 19. The estimated resonance frequencies seem to be close, but the estimated response does not match in its amplitude. The result seems insufficient for estimating the transfer function between the force from the local defect in the bearing, and the accelerometer as observer on the housing.

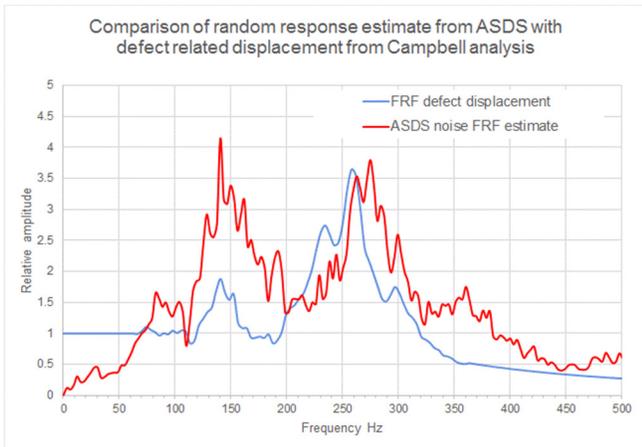


Figure 19. ASDS random noise spectrum versus FRF derived from run-up experiment.

4.1.2. Natural Defects Versus Artificial Defects

A basic question is whether natural defects can be compared with artificial defects. From both modeling and practical observations, there is evidence that this is the case. However, natural defects will have a different remaining life than artificial defects if only because natural defects occur in significantly aged steel while artificial ones are generally made in new bearings. A single case of proof will be shown here that the amplitude spectral density of spectrogram can be successfully used for natural defect size estimation.

A test could be done with a different rolling bearing which had already a significant defect. The bearing used was a tapered roller bearing unit that failed after many years of operation in a rail vehicle. This bearing does not fit in the housing used for SKF 229750 bearings, and a different housing has been used. The defect's shape was irregular, with course edges and the length in the rolling direction varied from approximately 3.5 to 7 mm, see Figure 20.

In this experiment, no suitable spring was available for completing the particular suspension. Instead a rigid pillar was replacing the spring in order to put the cylinder's load on the axlebox. First a test was done with an aluminum cylinder, followed by a test with an additional rubber block of 4 cm height in order to add some compliance and damping.

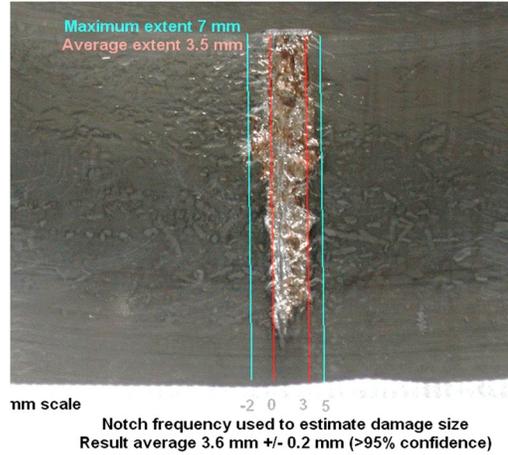


Figure 20. A natural defect in the outer ring running surface of a tapered bearing unit for a railway wheel.

From the pattern of minima in the response spectra, the defect was tested between 10 and 208 km/h the size as estimated as a mean of 3.6 mm with variation between -0.7 to +2.5 mm from this mean value. 95% of the measurement results were within +/- 0.2 mm of the mean, see Figure 21.

Note that mass and stiffness of the suspension was varied to observe disturbing effects from major rigid body resonances, which showed the peak errors on 95 km/h for the weaker suspension (rubber parts added in between) and 121 km/h for the case where a metal part was used instead of a spring.

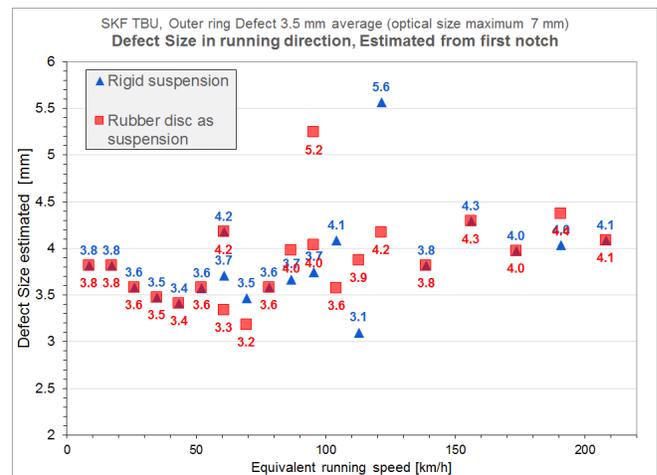


Figure 21. ASDS based defect size estimation result for the natural defect in the outer ring of an SKF double row tapered roller wheel bearing for two different mechanical compliances.

4.2. Growth Experiment to demonstrate Remaining Life

Bearing defect growth has been explored, modeled and tested in several applications (Morales-Espejel and Gabelli 2015, Gabelli and Morales-Espejel 2017, Bolander et.al, 2009 and

Rosado, Forster, Thompson and Cooke, 2010). The small defects were initially stable and it took a while before they started to grow. Once having a significant size, the growth accelerates but it will be depending on the load at what time scale one must be acting.

Railway bearings are generally running at fairly large ratios of load capacity to actual load (C/P), typically more than 10. The rated life is in the order of 4 million km of operation. Once a significant defect has appeared, the question is of course what time frame is reasonable in order to act before catastrophic failure may happen. In this work, the bearing modeling has been limited to estimation of the forces that act on the leading edge of the defect (when the roller has to enter the raceway again) and this indicated forces 3 – 5 times larger than the roller force in the good bearing.

At least one experiment could be done using a worn bearing from a railway customer that showed that even a significant defect is not growing very fast: the case shown had run for 30000 km equivalent in the test machine without major problems, though it was clearly damaged reading from the nature of the vibrations. Figure 22 shows the same defect in two stages of wear. The left photo is before the test distance has done, and the right photo after the test distance. The latter is the same picture as in Figure 20. At its start, the defect was smaller but not very small, its size was about 1.5 mm in the running direction, and about 20 mm in the width of the ring.

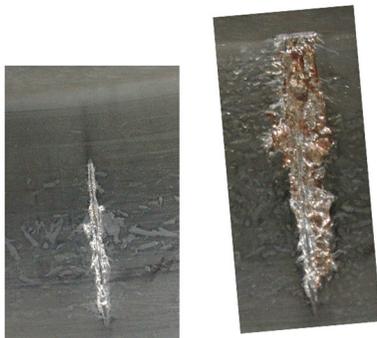


Figure 22. Collage of a defect before and after running a simulated test distance

After the 30000 km run the size had extended to more course shaped defect, 3 to 7 mm in the running direction, and to nearly its full roller-ring contact size in the width of the ring. This lead time shown here is usually adequate for a maintenance action to be planned and executed without large time pressure and conflicts of priority.

From earlier work on various defects in bearings, it is known that the ASDS will detect the defect and determine the size also for the defect stage that the bearing had at the start of the test (Figure 22 left). Unfortunately however, at that time no signals were collected and can therefore not be offered as further proof.

5. CONCLUSIONS

A method of analysis for vibrations has been presented, and discussed, the following conclusions can be drawn:

- Size estimation of defects in railway wheel bearings makes perfect sense as the lead time between a small defect and a catastrophic failure is in many cases long enough to act, and proven to be more than 3 months of use. The size estimate being in millimeters makes it possible to actually follow the growth of a defect and plan corrective actions timely and cost effectively, avoiding conflicts in priority for maintenance workplaces.
- Defects from very small to spanning up to about 20% of the roller spacing in the running direction can be detected and quantified with respect to their size.
- Defects larger than 20% of the roller spacing can be detected but no longer quantified in terms of the length by the ASDS method. The bandwidth of the force from the bearing becomes in the order of the resolution of the response spectrum determined with this method.
- Railway bearings were used as test objects to quantify the size estimation, and proof is presented of the relatively slow progression of defect sizes in the early stages of bearing failure. A lead time of 30000 km for planning maintenance seems possible.

6. REFERENCES

- Antoni, J. (2007). Cyclic spectral analysis of rolling-element bearing signals: Facts and fictions. *Journal of Sound and Vibration*, 304, pp 497–529
- Antoni J., Abboud D., & Xin G. (2016). Cyclostationarity in Condition Monitoring: 10 years after. *Proceedings of International Conference on Noise and Vibration Engineering ISMA2016*, 19 to 21 September 2016, Leuven, Belgium, pp 2365-2376
- Bolander N., Qiu H., Eklund N., Hindle E., & Rosenfeld T. (2009). Physics-based remaining useful life prediction for aircraft engine bearing prognosis. *Proceedings of the Annual Conference of the Prognostics and Health Management Society*, San Diego, CA, USA, September 27 – October 1, 2009.
- Ferreira J.L.A., Balthazar J.C., & Araujo A.P.N. (2003). An investigation of rail bearing reliability under real conditions of use. *Engineering Failure Analysis*, 10, pp 745–758.
- Gabelli A., & Morales-Espejel G.E. (2017). Improved fatigue life analysis of pre-dented raceways used in bearing material testing. *Bearing Steel Technologies 11th Volume*, “Progress in Steel Technologies and Bearing

- Steel Quality Assurance*", ASTM STP1600, pp. 167–191, doi: 10.1520/STP160020160154.
- Harris T.A. (2001). *Rolling Bearing Analysis*. 4th edition, John Wiley and Sons, New York, NY, USA, ISBN 0-371-35457-0.
- Moazen-Ahmadi, A., & Howard C.Q. (2016). A defect size estimation method based on operational speed and path of rolling elements in defective bearings. *Journal of Sound and Vibration*, 385, pp 138–148.
- Mol, H.A., & van Nijen, G.C. (1997). *Method for Analysing Regularly Recurring Mechanical Vibrations*. Patent US5698788. Published December 16, 1997.
- Morales-Espejel G.E., & Gabelli A. (2015). The progression of surface rolling contact fatigue damage of rolling bearings with artificial dents. *Tribology Trans.* (2015); 58: pp. 418–431.
- Müller R., Gratacos P., Mora P., Nielsen J., Feng J., & Cervello S., (2013). Definition of Wheel Maintenance Measures For Reducing Ground Vibration. Deliverable D2.7, EU FP7 project 'Railway induced vibration abatement solutions (RIVAS)', document code RIVAS_UIC_WP2-4_D2_7_V04. http://www.rivas-project.eu/fileadmin/documents/D2.7-Definition_of_wheel_maintenance_measures_for_reducing_ground_vibration.pdf.
- van Nijen, G.C. (1990). *Condition monitoring of bearings*. MSc Thesis, University of Twente, Faculty of Mechanical Engineering, February 1990.
- Oppenheim A.V., & Schaffer R.W. (1975). *Digital Signal Processing*. Englewood Cliffs, NJ: Prentice Hall.
- Rice D.A., Venkatachalam V., & Wegmann M.J. (1988). A simple envelope detector. *IEEE transactions on instrumentation and measurement*, vol 37, no. 2, pp 223–226.
- Rosado L., Forster N.H., Thompson K.L., & Cooke J.W. (2010). Rolling contact fatigue life and spall propagation of AISI M50, M50NiL, and AISI 52100, Part I: Experimental results. *Tribology Transactions*, 53, pp 29–41.
- Rycerz P., Olver A., & Kadiric A. (2017): Propagation of surface initiated rolling contact fatigue cracks in bearing steel, *International Journal of Fatigue*, 97, pp 29–38.
- Saleh J., & Marais K. (2006). Reliability: how much is it worth? beyond its estimation or prediction, the net present value of reliability. *Reliability Engineering & System Safety*, vol 91 pp 665–673.
- Sawalhi N., & Randall R.B. (2011). Vibration response of spalled rolling element bearings: Observations, simulations and signal processing techniques to track the spall size. *Mechanical Systems and Signal Processing*, 25, pp 846–870.
- SKF (2018), *General Catalogue 2018*.
- Wensing J.A. (1998). *On the Dynamics of Ball Bearings*. Doctoral Dissertation, University of Twente, the Netherlands, ISBN 90-36512298
- ten Wolde M., & Ghobbar A.A. (2013). Optimizing Inspection intervals - Reliability and availability in terms of a cost model: a case study on railway carriers. *Reliability Engineering and System Safety*, 114 (2013), pp 137–147.

Acknowledgement

The authors thank Stefan Lammens, SKF Director of Research and Technology Development, for his kind permission to publish this article.

BIOGRAPHIES

Henk Mol is scientist working at SKF on sensing and diagnostics of rolling bearings. Born in Rotterdam, the Netherlands, 5 June 1961. Graduated as MSc of Electrical Engineering at Technical University Delft in 1987 on the subject of solid state tunable oscillators. Since 1987 working for SKF in the research and development area, in jobs ranging from electronic sensing system design and functional integration into bearings for load, angle or vibration measurement, bearing quality control, through to bearing health and bearing application life monitoring. He is author of several patents on sensors, rolling bearing sensing integration and prognostics algorithms.

Antonio Gabelli is senior scientist working at SKF on tribology, prognostics and rolling contact fatigue life of rolling bearings. He joined the Research & Technology Development of SKF in Netherlands in 1981, since then he published many research papers on fatigue and lubrication of rolling bearings. He was one of the key contributors to the new bearing life theory, developed by SKF in the 90's. He is also author of several patents on rolling bearings. He holds a mechanical engineering degree from the University of Padua (Italy) and a PhD degree in from Chalmers Technical University (Sweden). For his contribution to rolling bearing technical literature he was the recipient of two STLE awards in 2010 and 2013.