Vibration signal processing for spall size estimation in rolling element bearings using autoregressive inverse filtration combined with bearing signal synchronous averaging

Nader Sawalhi¹, Wenyi Wang² and Andrew Becker²

Abstract
The main challenging area in the health monitoring of rolling element bearings is the quantification of the spall size using vibration data analysis. This is very crucial for maintenance planning and management decisions. In this article, we present a signal processing scheme for estimating spall sizes in rolling element bearings using autoregressive inverse filtration combined with bearing signal synchronous averaging. The squared envelope of the synchronously averaged signal and its autocorrelation function are used to estimate the spall size. The focus of the preprocessing algorithm using autoregressive inverse filtration resides in enhancing the weak step response events originating from the entry of a rolling element into the spalled region and balancing these with prominent impulse responses which occur when a rolling element strikes the trailing edge. Preprocessing is attained through whitening the shaft order tracked (angular resampled) signal using an autoregressive model based on the shaft synchronously averaged part (autoregressive inverse filtration). Autoregressive inverse filtration is compared to autoregressive filtration based on the raw vibration signal. The selection of the autoregressive model order is realized using Akaike criterion. The efficacy of the two autoregressive filtration algorithms is established by comparing time-domain signals, bearing signal synchronous averages, and their squared envelopes and autocorrelation function. This is done on simulated signals with well-known characteristics and on two sizes of naturally originated and propagated inner race spalls from a high-speed test rig. The sizes of these faults were large in a sense that the rolling element did not bridge over the spall, and this required an adjustment to the size quantification equation to fit this case, which has not been presented before. The combination of autoregressive inverse filtration and the squared envelope of the bearing synchronous averaging gives a superior enhancement to the step response and balances it with the impulse response. This provides the best accuracy in estimating the size of the spall, and unlike other existing algorithms, there exist no need for further processing using wavelets for instance.

Keywords
Rolling element bearing, fault size estimation, autoregressive inverse filtration, pre-whitening, bearing signal synchronous average, order tracking, squared enveloped signal, autocorrelation

Date received: 19 December 2016; accepted: 13 March 2017

Academic Editor: Dong Wang

Introduction
Rolling element bearing fault diagnosis, using vibration data collected via accelerometers, has been the focus of research and industry for decades, where a great

¹Department of Mechanical Engineering, Prince Mohammad Bin Fahd University, Al Khobar, Saudi Arabia
²Aerospace Division, Defence Science and Technology Group, Fishermans Bend, VIC, Australia

Corresponding author:
Nader Sawalhi, Department of Mechanical Engineering, Prince Mohammad Bin Fahd University, Al Khobar 31952, Saudi Arabia.
Email: nsawalhi@pmu.edu.sa
number of signal processing algorithms have been
developed and implemented successfully. A widely used
technique is the envelope analysis or the high-frequency
resonance technique (HFRT). Although diagnosing a
bearing fault can be done with ease and confidence due
to the number of robust algorithms developed, it
may still be a challenging task in some cases, as it is a
function of a number of variables, which may mask the
bearing defect signal. These include the location of the
accelerometer (transfer path between the source of the
fault and the sensor); the speed fluctuations; and, in
particular, variable speed machines and the loading of
the bearings.

A more challenging area is the quantification of the
fault size using the vibration data. This is very crucial
for maintenance planning and management decisions.
Through the novel work of Epps in 1991, it has been
established that the vibration response of a defective
rolling element bearing (localized faults such as spalls)
comprises two main parts (although argued at some
instances that a third part exists, but this was not clearly
established). The first part of the response comes as the
rolling element rolls into the spall. The change of curva-
ture causes a step response in acceleration which results
in a low-frequency transient event with a negative slope.
The second part of the response results from the impact
of the rolling element with the trailing edge of the spall.
This causes a sudden change in the rolling direction,
which gives a step response in velocity and results in an
impulse response in acceleration. The third event, which
results from the fully departure of the rolling element
and regaining contact with raceway (described as a low-
frequency transient with a positive slope—opposite to
the entry), was not clearly distinguished due to the
dominance of the impact event. Epps showed that the
time between the rolling in and the impact (denoted as
the time to impact (TTI)) is proportional to the size of
the spall and can be used to measure the spall width. A
typical example from Epp’s thesis showing the two
parts of the vibration signal is shown in Figure 1.

In 2011, Sawalhi and Randall conducted a number
of tests with seeded inner race and outer race faults
two sizes of 0.6 and 1.2 mm). The test rig (bladed fan
test rig) was operated at speeds between 800 and 2400 r/
min. A typical result from their work for an outer race
spall is shown in Figure 2. The step response, associated
with the ball entry into the spall, was observed, and the
separation between the entry and impact could be
directly linked to the size of the fault. It has been indi-
cated that “the entry into the fault could be classified as
a step response, with mainly low-frequency content,
while the impact on exit excites a much broader band
impulse response, with higher frequency content.”
Analytical simulation and signal preprocessing and
enhancements were proposed to enable quantifying the
size of the spall. In addition to providing an analytical
model for the two parts, Sawalhi and Randall proposed
a number of signal processing alternatives to enhance the weak entry event and estimate the size of the spall. In particular, they proposed the use of signal
pre-whitening using autoregressive (AR) modeling for
enhancing the entry event, Morlet wavelet analysis to
balance the energy of the entry and exit events, and the
cepstrum to estimate the spacing between the two
events. As the size of the spall was very small compared
to the rolling element diameter, it was presumed that
the rolling element will bridge over the spall and that
the impact occurs when the rolling element is half way
through the spall. An equation to estimate the size of
the spall based on the number of samples estimated
between the entry and impact (TTI) was also derived.

A number of researchers have since examined vibration
signals for defective bearings at different running
speeds and with different spall sizes to enable estimating
the fault size. In a 2012 study by Jena and Panigrahi, the
authors used analytical wavelet transform (AWT)
followed by time marginal integration (TMI) to provide
an estimate of a 2.1-mm defect in the inner race at a
speed of 1500 r/min. Singh and Kumar used a prepro-
cessing step (multiply the amplitude of the vibration sig-

nal by the absolute value) to enhance the step response
followed by wavelet analysis to detect defect width over
the range of 0.4399–1.4854 mm. Accuracy of estimates
were observed to increase with the loading of the bear-
ing. More recently, Moustafa et al., in 2014, intro-
duced instantaneous angular speed (IAS) to aid bearing
prognosis. The introduced IAS was tested at very low
speed (10–60 r/min) to estimate the fault size of 0.7–
6 mm. IAS was found to provide accurate size estimation
for normal to high radial loading. Moazenahmadi et al., in
their 2014 article, provided insights and more
understanding to the path of rolling elements in the
defect zone for developing an algorithm to estimate the
size of a bearing defect. Their method produced more
accurate results that were less biased with operational
speed. Ismail and Sawalhi, in 2016, introduced a new
technique for extracting the spall size and determining
the geometric factor between the extracted size and the
actual size without comparing with any reference data.
The technique consists of two energy (or root mean
square (RMS)) envelopes for the vibration signal and
after numerically differentiating this signal using adjus-
table Savitzky–Golay. Finally Wang et al., in their
2016 work, proposed a system of estimating bearing
spall size using a tacho-less synchronous signal aver-
ging (SSA) with respect to the bearing fault character-
istic frequency combined with envelope and wavelet
analyses of the averaged signals. The proposed algo-
rithm, which was applied directly to the raw vibration
signal without preprocessing, was validated using the
vibration data from naturally spalled bearings (4.2 and
6.2 mm) in a high-speed bearing test rig (7700 r/min).
The presented results showed that the technique is effective in revealing the entry and exit features needed for the size estimation of naturally occurring bearing faults. No preprocessing was attempted before applying the SSA scheme.

In this article, we present a signal processing algorithm to estimate the size of spalls in rolling element bearing using vibration signals. The proposed algorithm includes a whitening process of the shaft order tracked signal using autoregressive inverse filtration (ARIF; i.e. filter coefficients are derived from the shaft synchronous averaged signal). The residual of the ARIF process is then utilized to extract a bearing signal synchronous average (BSSA), whose squared envelope is used to estimate the size of the spall. The selection of AR model order and the validation of the algorithm are demonstrated on simulated signals and on the two naturally originated inner race spalls used in Wang.
et al.\textsuperscript{16} A number of comparisons are made to demonstrate the advantages of the ARIF processes compared to AR filtration on the raw signal and the importance of signal preprocessing prior to using BSSA.

This article is organized as follows: In section “Processing scheme to estimate spall size,” the proposed processing scheme to estimate spall size in rolling element bearings is discussed. This involves presenting the ARIF of the order tracked residual signal and the BSSA procedure. The bearing test rig used to collect experimental data, the type of faults introduced in bearings, and the characteristics of the collected vibration data are discussed in section “Bearing test rig and order tracking.” This section also provides calculation of the spall size in samples. Section “Simulations” presents analytical simulations which are based on characteristics similar to that seen in experimental data. This section also presents the results obtained from applying the processing scheme presented in section “Processing scheme to estimate spall size” on the simulated signals. Section “Experimental results” examines the experimental data when subjected to the processing scheme. Finally, conclusions are given in section “Conclusion.”

Processing scheme to estimate spall size

The proposed processing scheme consists of three main stages of processing. The first stage involves preprocessing an angular sampled (order tracked) version of the raw vibration signal using ARIF. The aim of this preprocessing is to remove the effect of the transfer path and whiten the vibration signal which then enhances the weak step response and minimizes the dominance of the impulse response. ARIF processing algorithm is discussed and illustrated in section “ARIF and whitening of the shaft order tracked signal.” Results obtained using this proposed algorithm are compared to AR processing applied directly on the raw vibration signal. In the second stage, the residual of the ARIF algorithm is processed to extract a BSSA with respect to a bearing fault characteristic frequency. The essence of this processing is discussed in section “Tacho-less BSSA.” The final stage of processing includes calculating the squared envelope of the BSSA obtained from step 2 for three or more periods and examining it and its autocorrelation function to estimate the size of the spall.

**ARIF and whitening of the shaft order tracked signal**

Figure 3 describes schematically the ARIF process to enhance the step response and improve the presence and localization of the impulse response.

In this approach, an AR model is derived using the shaft time synchronous averaged (TSA) signal. The first step in the process includes using the tachometer signal to order track the vibration signal, thus removing any speed fluctuations. The TSA is then obtained through finding the ensemble average of a number of shaft rotations.

The TSA contains all the harmonics of the shaft frequency (shaft orders) and the ringing effects (sinusoidal components) of the structural resonances in the system. This TSA is used to build an AR model using equation (1) presented below. In an AR model, a value at time $t$ is based on a linear combination of prior values (forward prediction), a combination of subsequent values (backward prediction), or both (forward–backward prediction). If $x$ is a data series (zero mean stationary process) of length $N$ and $a$ is the AR parameter array of order $p$, an AR model $y$ can then be defined in equation (1) as follows\textsuperscript{17}

$$Y_k = - \sum_{i=1}^{p} a_i x_{k-i} + e_k \quad (1)$$

where $p$ is the order of the model and $a_i (i = 1, 2, ..., p)$ are weighting coefficients. The error term $e_k$ (residual signal) is a white noise process, with a variance $\sigma^2$, which represents the difference between the actual and linearly predicted values. Residual signal contains additive white noise and nonstationarities in the form of impulses, both of which are delta correlated, with a white spectrum.

Several criteria have been proposed to select the optimum (minimum) model order ($p$) The Akaike Information Criterion (AIC)—one of the most popular measures in the literature—determines the model order $p$ by minimizing an information theoretic function of $p$;\textsuperscript{18} AIC($p$) is defined by equation (2)

$$\text{AIC}(p) = N \left( \ln \left( \sigma_p^2 \right) \right) + 2p \quad (2)$$

where $N$ is the number of samples and $\sigma_p^2$ is the estimated variance of the driving noise (i.e. the prediction error). The term $2p$ is a “penalty” for the use of extra AR coefficients that do not substantially reduce the prediction error.

The “AIC” is only one of many criteria proposed for the selection of the AR order. Another popular criterion is the final prediction error (FPE), which selects the model order $p$ by minimizing the function FPE($p$),\textsuperscript{19} defined as

$$\text{FPE}(p) = \left( \sigma_p^2 \right) \times \frac{N + p + 1}{N - p - 1} \quad (3)$$

The term $(N + p + 1/N - p - 1)$ increases with $p$ and represents the inaccuracies in estimating the AR parameters.\textsuperscript{20}
In this work, we have adopted AIC to guide the selection of the model order as it is a widely used one. In addition to the filter length criterion mentioned above, the choice of the filter length \(p\) could be guided subjectively by the duration (length) of the impulse response in samples to capture the excited resonances within the model. More discussion and insights are provided when presenting simulated signals in section “Experimental results.”

Note that the ARIF whitening process ensures that a true AR model is derived for the signal and also ensures a more efficient removal of the masking deterministic components (shaft harmonics and resonance ringing sinusoidal components). Thus, it is anticipated that the ARIF whitening process should provide a better enhancement to the entry event and a better localization to the exit event. The limitation of the ARIF whitening resides in the need of a tachometer signal for order tracking and synchronous average signal calculation. However, this is not seen as an obstacle as order tracking can be realized using phase demodulation or through pseudo tachometer extraction as has been discussed in a number of publications.20,21

Sawalhi and Randall7 used signal pre-whitening using AR modeling to enhance the step response. The whitening was achieved based on deriving an AR filter coefficients using the raw vibration signal itself. The residual signal (difference between the raw signal and an AR-estimated version) is white in the sense of containing noise and impulses, which has a relatively flat spectrum. This helps in lifting the power of low energy events and thus results in gaining some enhancement. In both simulated and experimental sections, ARIF and AR whitening of the raw vibration signals are compared.

**Tacho-less BSSA**

Time synchronous averaging (TSA) method is widely used in gear fault diagnosis.17 This was also used to aid bearing fault diagnostics by applying it on the envelope signal to enhance the bearing fault characteristic features.22,23 Using TSA with rolling element bearings requires a speed reference signal (tachometer) from both the shaft and the cage of the bearing, which is impractical for most machines.23 In 2005, Bonnardot et al.20 introduced a method of extracting TSA without any tachometer reference. This was possible using the vibration signals' instantaneous phase information at a dominant shaft order.20

Tacho-less BSSA method was mainly utilized in the detection and diagnosis of bearing faults.24,25 The only reference of applying the tacho-less BSSA method to size estimation of spalls has been published recently,16 where the authors used phase demodulation to extract a speed reference from the inner race ball pass frequency and used this to find the squared envelope of the BSSA to estimate the size of two naturally grown inner race spalls from a test rig running at a high speed. The authors then used band pass filtration and wavelet analysis to estimate the spall size.

In this work, a period timing method (pseudo encoder)21 is used to extract a tacho reference. In this process, a buffer (filled with zeros) of a size equal to the FFT size of the signals is created. The complex spectrum of interest (band around the characteristic bearing frequency) is then transferred to this buffer (filling in the same lines as in the original spectrum). Finally, the buffer is inversely transformed to the time domain to obtain the reference signal. This signal is a sinusoidal
signal whose periods represent the period of passage of the rolling elements across either the inner race or the outer race. The fluctuations between each rolling element passage can be traced using the zero crossing of the consecutive periods, which can be achieved by setting a trigger and identifying the zero-crossings. This signal is then used to order track the signal and obtain a synchronous average.

**Bearing test rig and vibration data**

**Test rig and two naturally grown large spall sizes**

The test rig used to generate the data for this work is pictured in Figure 4. It consists of two bearing housings. One of these housings contains the test bearing (angular contact bearing), which is loaded axially through screwing a large nut to the housing. The other bearing housing (reaction bearing housing) contains two angular contact bearings, arranged in tandem, which react to the test load. The test rig is driven by a constant speed motor through a belt-pulley system to give a nominal running speed of around 7700 r/min. The test rig was fitted with a vertical–radial accelerometer (as shown in Figure 4), an axial accelerometer, and a tachometer to get a speed reference. Data were acquired at a sampling rate of 200,000 samples/s (Hz).

The information about the test bearing is listed in Table 1. In total, two main fault sizes were used in the testing as shown in Figure 8(b) and (c). The spalls were initiated from a 0.25-mm notch (Figure 8(a)) across the middle of the inner race (depth of 0.1 mm and a width of 2 mm). The seeded notch was created using electrical discharge machining (EDM) and allowed to propagate circumferentially along the raceway in accelerated endurance tests. The final sizes of the spall (lengths) were about 6.2 (AC3) and 4.2 mm (AC8). Sizes were estimated by counting the number of grids on the pictures shown in Figure 8.

![Figure 4. Vibration bearing test rig.](image)

**Table 1. Test bearing parameters.**

| Parameter                        | Value               |
|----------------------------------|---------------------|
| Number of rolling elements N     | 15                  |
| Contact angle (β)                | 15°                 |
| Bore                             | 30 mm               |
| Diameter of outer raceway        | 50.19 mm            |
| Diameter of inner raceway (Di)   | 35.80 mm            |
| Ball diameter (Db)               | 7.14 mm             |
| Maximum radial load              | 9450 N (static)     |
| Maximum axial load               | 4000 N              |
| Pitch diameter (Dp)              | 42.93 mm            |

The rate at which a ball passes a point on the inner race (BPFI) can be found using equation (4)

\[
BPFI = \frac{N}{2} f_r \left(1 + \frac{D_b}{D_p} \cos(\beta)\right) \tag{4}
\]

where \(f_r\) is the shaft rotational speed, \(D_p\) is the pitch diameter, \(D_b\) is the ball diameter, \(N\) is the number of rolling elements, and \(\beta\) is the load contact angle.

Using the values of Table 1, the BPFI can be approximated as 8.7 times the shaft rotational speed. For the AC3 bearing test, the bearing shaft and the inner race were running at a speed of 6270 r/min (104.5 Hz). The BPFI was estimated at 909.1 Hz, which corresponds to an impact period of about 220 samples (i.e. 200,000/909.1). The AC8 bearing was running at a nominal speed of 7500 r/min (125 Hz), which is significantly higher than the AC3 bearing’s running speed, thus giving a shorter impact period estimated at 184 samples (i.e. 200,000/(8.7 × 125)). The spall lengths of AC8 (4.2 mm) and AC3 (6.2 mm) correspond to about 56% and 82%, respectively, of the pitch distance between two balls (7.49 mm (i.e. \((\pi D_i)/15\); Figure 5).

**Spall size estimation based on the TTI (number of samples between the roll in and impact)**

When the size of the spall is large, the rolling element does not bridge over the spall, and equation (5) presented in Sawalhi and Randall\(^7\) does not give accurate estimation and has to be adjusted. Instead, equations (6) and (7) are to be used for accurate estimates. A schematic comparison between the case presented in Sawalhi and Randall\(^7\) under which equation (5) was derived and a large spall size (as the ones used in this work) is shown in Figure 6.

Note that equation (6) has “2” in the denominator and an extra term “x” which is a function of the ball diameter (\(D_b\)) and the depth of the fault (\(\delta\)). The equation for the extra term “x” is given in equation (3). This adjustment comes as a result of the rolling element passage over the spall region compared to bridging over it.
in the early stages of the fault (small spall widths). For a typical fault depth of 0.2 mm, the adjustment value \( x \) was estimated at 1.2 mm. However, as the fault size increased from 4.2 to 6.2 mm, we can assume the faulted area deepened from 0.2 to 0.3 mm; hence, the adjustment will be 1.5 mm.

\[
\begin{align*}
\ell_o &= \pi f_r \left( D_p^2 - D_b^2 \right) / D_p f_s \quad \text{(5)} \\
\ell_o &= \pi f_r \left( D_p^2 - D_b^2 \right) / 2D_p f_s \quad \text{(6)}
\end{align*}
\]

where \( \ell_o \) is the spall width (mm), \( D_p \) is the pitch diameter (mm), \( D_b \) is the ball diameter (mm), \( f_r \) is the shaft rotational speed (Hz), \( f_s \) is the sampling frequency (Hz), \( sp \) is the number of samples between the entry (step response) and the impact, and \( \delta \) is the fault depth.

Based on the measured lengths of spalls and through adjusting with 1.2 mm for the 4.2-mm spall (AC8) and 1.5 mm for the larger 6.2-mm AC3 spall, the spacing between the step and impulse response is estimated at 88 samples and 130 samples, respectively.

\[
x = \sqrt{D_p \delta + \delta^2} 
\]  

(7)
Characteristics of the measured signals

Raw vibration signals from the AC8 and AC3 bearings are shown in Figure 7, where the impacts generated by the ball passage over the spall can be clearly seen.

The amplitude spectra for both AC8 and AC3 signals are shown in Figure 8 with the inner race fault frequency of 1079 Hz (AC8) and 912 Hz (AC3; as marked by the data tip) absolutely dominating. The resonance frequency of the accelerometer is also obvious at around 50 kHz. It is also noted that some resonances are excited below 10 kHz. The AC8 bearing was running at a nominal speed of 7500 r/min, which was significantly higher than the AC3 bearing’s running speed.

Simulations

Signal generation

Figure 9 displays a zoomed in version of Figure 7(b), containing three impact periods. As can be seen, there is a lot of noise and interferences in the raw signal...
between two impacts, so that it is impossible to extract features from the raw signature with the balls entry into and exit from the spalled area on the inner raceway. Because of the high speed used in the test rig (125 Hz), and although the sampling frequency was high (200 kHz), the number of samples between two consecutive impacts are still quite small (around 180 samples). It is most likely that the entry and exit features overlap because the entry-related response does not completely disappear before the exit event starts. This would be similar to the case for the impulse response at exit as it may not die out completely before the step response from the next ball.

In order to better understand the analysis process, a signal simulation was carried out. The experimental characteristics observed in section “Bearing test rig and vibration data” are mimicked using analytical simulations. Table 2 lists the simulation parameters used when creating simulated signals as described in Table 3. Steps 1–4 of Table 3 are detailed in Sawalhi and Randall.\textsuperscript{7} Step 5 was introduced to include a modulation effect. This was realized using a hanning window.

Figures 10 and 11 show the time domain of the simulated signal and its frequency content, respectively.

**Simulated signal processing**

Figure 12 shows the shaft TSA and its frequency content, which contains the shaft frequency and the two excited resonances. The implication of this is that the AR model derived from this signal will capture both the shaft frequency and the resonances in the system.

The optimum AR model orders based on Akaike criterion for the ARIF and the one based on the raw signal are plotted in Figure 13. Deriving AR coefficients based on ARIF requires a model order of 23 samples, which translates to about 12% of the impact period. Doing this based on the raw signal requires a filter length of 223 samples, which appears to be of the same order as the impact period.

The residuals of ARIF and AR based on the raw signal over one shaft rotation are plotted in Figure 14. It is seen that both perform well in removing the ringing effect and localizing both the step and impulse responses. However, the ARIF gives a slightly better enhancement for the step response in amplitude and appearance when compared to AR based on the raw signal.

The performance of preprocessing using AR is tested next on the BSSA of the signals. The comparisons between BSSA with no preprocessing to these obtained from the two AR approaches are shown in Figure 15. It is noted that all of the three periods of BSSA extract the step and impulse responses. BSSA over the raw vibration signal (no preprocessing) extracts both the step and impulse responses but with keeping the ringing effect; thus, the localization in terms of the starting positions is missing. AR on the raw signal enhances the step response, localizes the impulse, and balances both events but gives less clear results compared to ARIF. ARIF enhances the step response, localizes the impulse response, and gives a balance between the two responses. Thus, among the three periods of BSSA, the one based on ARIF gives the best result. This can be seen clearly through inspecting the squared envelope signals of the BSSA as shown in Figure 16, where it can be seen clearly that ARIF gives the best result in enhancing the step response and localizing both events, so that spacing can be easily measured to estimate the size of the spall. In the case of no preprocessing and ARIF, further balancing of the events and low-pass filtration for the high-frequency components attached with the impulse events are usually sought as has been the case in Sawalhi and Randall\textsuperscript{7} and Wang et al.\textsuperscript{16} in which wavelets were employed to aid fault estimation. The one-sided zero-centered autocorrelation of the squared envelope of the BSSA is examined in Figure 17. The 200 and its multiples indicate one impact period, which coincide with

| Table 2. Simulation parameters. |
|----------------------------------|
| **Sampling frequency** $f_s$ (Hz) | 200,000 |
| **Shaft speed** $f_r$ (Hz) | 125 |
| **BPFI (Hz)** | 1000 |
| **Period between impact (BP; samples)** | 200 |
| **Impulse frequency = accelerometer resonance** $f_1$ | 50,000 |
| **Step response frequency** $f_2 = f_1/6$ | 8333 |
| **Spall length** $x_p$ (samples; 60% of impact period) | 120 |
| **Damping constant** $\tau$ (s) | 0.0002 |
the impulse period. The step response location can be seen around sample number 84, giving a spall size of 116 samples, which can be seen as a hump around the location sample 117 in the ARIF and 114 in the AR base. The error in size estimation using ARIF is around 3% for ARIF and 5% for AR based on the raw signal.

**Experimental results**

**Preprocessing using AR and ARIF**

Figure 18 examines the optimum AR filter order for the AC3 and AC8 for ARIF and AR based on the raw signals. ARIF is optimum at 10 and 49 for AC8 and AC3, respectively. This translates to 6% and 22% of the impacts periods for AC8 and AC3, respectively. No minimum was achieved over the 500 orders selected for AR based on the raw signal, as the Akaike values kept decreasing, but it was noted that there is a very minimal change beyond order 200, so it was used for both AC3 and AC8. As observed earlier in the simulated results, ARIF gives a minimum value (optimum) to select the best filter order, while AR based on the raw signal tends to be the best based on the filter length of the same order as the impact period.

Figures 19 and 20 show zoom-in views (about six or seven impacts) from residual signals obtained after preprocessing of AC8 and AC3 signals using AR filtration (both ARIF and AR based on the raw signal). Step responses are noted to be made clear at few instances using both methods of AR and ARIF. However ARIF is observed to give better visual results, and the step responses are obvious at more locations compared to AR based on the raw signals.

**Table 3.** Simulation steps and summary.

| Steps | Equation (mathematical process) | Time-domain signal |
|-------|---------------------------------|--------------------|
| 1 Impulse response | $y = e^{-t/T} \sin(2\pi ft)$ | ![Impulse response](image1) |
| 2 Step response: scaled by | $y = e^{-t/T} \sin(2\pi ft)$ | ![Step response](image2) |
| 3 Repeat the impulse and the step responses with 2% fluctuation of the period | ![Repeating responses](image3) |
| 4 Scale the step response (divide it by 8, this will make its amplitude = 0.25 times the amplitude of the impulse) Introduce a shift to the impulse and add the responses | ![Scaling responses](image4) |
| 5 Create modulation by shaft speed Add noise (SNR = 20) | ![Modulation](image5) |

SNR: signal-to-noise ratio.
Figure 10. One shaft rotation of (a) step response, (b) impulse response, and (c) step and impulse responses added with 120 samples delay, SNR = 20 dB. (d) Signal c modulated by the shaft speed.

Figure 11. Frequency content of simulated signal shown in Figure 10(d): (a) full bandwidth (0–100 kHz) showing excited resonances and (b) frequencies 0–1000 Hz showing shaft frequency and the first four harmonics of the BPFI.
Signals obtained in Figures 19 and 20 were further processed to obtain their BSSA. The results of BSSA are presented in Figures 21 and 22 for AC8 and AC3, respectively. One impact duration, bounded by two impulses, is indicated using a double arrow. The location of the step response is shown in these figures. The signal preprocessed using ARIF provides a much clearer identification for the location of the step response compared to the case of no preprocessing and the one where preprocessing was done using AR based on the raw signal. This can be seen and affirmed further by examining the squared enveloped signals of the BSSA in Figures 23 and 24 and their one-sided zero-centered autocorrelation functions in Figures 25 and 26 for AC8 and AC3 signals, respectively. Of particular interest to the aid of spall size are the squared enveloped signals of the BSSA obtained after AR preprocessing as shown.
in Figure 23 for AC8 and Figure 25 for AC3. ARIF result outperforms the AR processing over the raw signal and the case of no preprocessing. This is clear in particular for the AC3 spall but can also be seen noticeably in the AC8 spall. The use of autocorrelation to estimate the spall size, using ARIF preprocessing, follows a similar trend to what have been discussed in section “Simulations,” where two peaks appear before the impulse location: One represents the step location and the other gives a direct measure for the size of the spall.

It can be seen through inspecting and comparing these results that ARIF gives the best results in enhancing the step response and balancing and localizing the

![Figure 14](image1.png)

**Figure 14.** One shaft rotation of (a) simulated signal, (b) residual of AR derived from the raw signal, and (c) residual of AR inverse filtration.

![Figure 15](image2.png)

**Figure 15.** Three periods of BSSA (a) calculated on the raw signal, that is, with no preprocessing; (b) using AR filtered on the raw signal; and (c) using AR inverse filtered signal.
Figure 16. Three periods of the squared envelope of BSSA (a) calculated on the raw signal, that is, with no preprocessing; (b) using AR filtered on the raw signal; and (c) using AR inverse filtered signal.

Figure 17. One-sided zero-centered autocorrelation of the squared envelope of the BSSA (a) calculated on the raw signal, that is, with no preprocessing; (b) using AR filtered on the raw signal; and (c) using AR inverse filtered signal.
impulse response. Achieved results are in well agreement with the simulated result presented in the previous section. The spall size of AC8 in samples is estimated at 82 samples (i.e. 408–326 from Figure 23(c) and 186–104 from Figure 25(c)), compared to 88 samples for accurate result (section “Spall size estimation based on the TTI (number of samples between the roll in and impact)”). This gives an error of 4.5%. For the AC3, the spall size in samples is estimated at 129–132 samples (i.e. 332–203 from Figure 24(c) and 220–89 from Figure 26(c)) which matches exactly the expected value of 130 samples as discussed in section “Spall size estimation based on the TTI (number of samples between the roll in and impact).”

Figure 18. AR filter order selection based on Akaike criterion: (a) AC8 AR on the raw signal, (b) AC3 AR on the raw signal, (c) AC8 AR inverse, and (d) AC3 AR inverse.

Figure 19. A zoom-in showing six impact periods of AC8 signal: (a) order tracked signal, (b) residual of AR derived from the raw signal, and (c) residual of AR inverse filtration.
Figure 20. A zoom-in showing seven impact periods of AC3 signal: (a) order tracked signal, (b) residual of AR derived from the raw signal, and (c) residual of AR inverse filtration.

Figure 21. AC8: three periods of BSSA (a) calculated on the raw signal, that is, with no preprocessing; (b) using AR filtered on the raw signal; and (c) using AR inverse filtered signal.
Figure 22. AC3: three periods of BSSA (a) calculated on the raw signal, that is, with no preprocessing; (b) using AR filtered on the raw signal; and (c) using AR inverse filtered signal.

Figure 23. AC8: three periods of (a) bearing synchronous average on the raw signal, (b) bearing synchronous average on the residual of AR raw signal, and (c) bearing synchronous average on the residual of the AR inverse.
Conclusion

A signal processing scheme for estimating the fault size in rolling element bearings has been presented. This is based on using a preprocessing stage using ARIF to remove shaft harmonics and the ringing effects of excited resonances, thus enhancing the step response associated with the spall entries and localizing the impulse response at impact instances. Akaike criterion is utilized to select the AR filter order. The

**Figure 24.** AC3: three periods of (a) bearing synchronous average on the raw signal, (b) bearing synchronous average on the residual of AR raw signal, and (c) bearing synchronous average on the residual of the AR inverse.

**Figure 25.** AC8: one-sided zero-centered autocorrelation of the squared envelope of the BSSA (a) calculated on the raw signal, that is, with no preprocessing; (b) using AR filtered on the raw signal; and (c) using AR inverse filtered signal.

**Conclusion**

A signal processing scheme for estimating the fault size in rolling element bearings has been presented. This is based on using a preprocessing stage using ARIF to remove shaft harmonics and the ringing effects of excited resonances, thus enhancing the step response associated with the spall entries and localizing the impulse response at impact instances. Akaike criterion is utilized to select the AR filter order. The
preprocessing stage is followed by a BSSA stage with respect to the ball pass frequency of the fault. The presented processing scheme has been tested on simulated signals and on two inner race spalls for bearings running at high speed. The inner race spalls were naturally grown, and thus, they contained realistic feature of the spalls.

Simulated results were used to test the processing stages and gain an in-depth understanding of the processing. Comparisons were held between preprocessing using ARIF and AR based on the raw signal. The comparison involved comparing the orders of the filters, the time-domain signals, and the bearing signal synchronous averaged signals and their squared enveloped signals and autocorrelation functions to estimate the size of the spall. Simulated results showed a superior performance of fault estimation using ARIF as a preprocessing step, followed by AR based on the raw signal. The case of no preprocessing gave the least accurate estimate as the ringing effect associated with resonances appeared in the BSSA.

For the experimental data, the weak entry event in the cases of the two spall sizes was very hard to observe in the raw data. The use of an ARIF and AR based on the raw signal made it possible to detect some instances of step entries. This was even made better using BSSA on the preprocessed signals and plotting the squared envelopes of the BSSA signals, where localization of both the step and impulse response was achieved. The result quality was much better when preprocessing the signal using ARIF, and in addition to enhancing the step response, the impulse response was better localized. This, in turn, gave a better result and higher accuracy in estimating the size of the spall which were measured directly from the squared envelope signal and using the autocorrelation of the squared enveloped signals.

Acknowledgements
We would like to acknowledge Mr Peter Stanhope of DST Group for his support in the bearing tests. We also thank Dr David Forrester and Dr Peter Frith of DST Group for their support of this work.

Declaration of conflicting interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The first author acknowledges the research fund he received from Defence Science and Technology Group to conduct this work.

He also acknowledges the financial support received from Prince Mohammad Bin Fahd University (PMU) for paying the APC.

References
1. Darlow MS, Badgley RH and Hogg GW. Application of high frequency resonance techniques for bearing diagnostics in helicopter gearboxes. Technical report, US Army Air Mobility Research and Development Laboratory, Moffett Field, CA, 1974, pp.74–77.
2. Antoni J and Randall RB. The spectral kurtosis: application to the surveillance and diagnostics of rotating machines. *Mech Syst Signal Pr* 2006; 20: 308–331.

3. Antoni J. Fast computation of the kurtogram for the detection of transient faults. *Mech Syst Signal Pr* 2007; 21: 108–124.

4. Sawalhi N and Randall RB. Spectral kurtosis optimization for rolling element bearings. In: *Proceedings of the eighth international symposium on signal processing and its applications*, Sydney, NSW, Australia, 28–31 August 2005.

5. Ho D and Randall RB. Optimisation of bearing diagnostic techniques using simulated and actual bearing fault signals. *Mech Syst Signal Pr* 2000; 14: 763–788.

6. Epps IK. An investigation into vibrations excited by discrete faults in rolling element bearings. PhD Dissertation, University of Canterbury, Christchurch, 1991, http://hdl.handle.net/10092/6025

7. Sawalhi N and Randall RB. Vibration response of spalled rolling element bearings: observations, simulations and signal processing techniques to track the spall size. *Mech Syst Signal Pr* 2011; 25: 846–870, http://www.mpft.org/FaultData/FaultData.htm

8. Sawalhi N and Randall RB. Spectral kurtosis enhancement using autoregressive models. In: *4th Australasian congress on applied mechanics (ACAM 2005)*, Melbourne, VIC, Australia, 16–18 February 2005, pp.231–236.

9. Mallat S. *A wavelet tour of signal processing*. Academic Press, 1999.

10. Childers DG, Skinner DP and Kemerait RC. The cepstrum: a guide to processing. *Proc IEEE* 65: 1428–1443.

11. Ismail M and Sawalhi N. Bearing spall size quantification based on geometric interpretation of vibration envelope energy. *Paper presented at 13th international conference on condition monitoring and machinery failure prevention technologies*, Paris, 10–12 October 2016. London: The British Institute of Non-Destructive Testing.

12. Jena DP and Panigrahi SN. Precise measurement of defect width in tapered roller bearing using vibration signal. *Measurement* 2014; 55: 39–50.

13. Singh M and Kumar R. Thrust bearing groove race defect measurement by wavelet decomposition of pre-processed vibration signal. *Measurement* 2013; 46: 3508–3515.

14. Moustafa W, Cousinard O, Bolaers F, et al. Low speed bearings fault detection and size estimation using instantaneous angular speed. *J Vib Control* 2014; 22: 3413–3425.

15. Moazenahmadi A, Petersen D, Howard C, et al. Defect size estimation and analysis of the path of rolling elements in defective bearings with respect to the operational speed. *Paper presented at 43rd international congress on noise control engineering*, Melbourne, VIC, Australia, 16–19 November 2014. Reston, VA: Institute of Noise Control Engineering.

16. Wang W, Sawalhi N and Becker A. Size estimation for naturally occurring bearing faults using synchronous averaging of vibration signals. *J Vib Acoust: T ASME* 2016; 138: 051015 (10 pp.).

17. Wang W and Wong AK. Autoregressive model-based gear fault diagnosis. *J Vib Acoust: T ASME* 2002; 124: 172–179.

18. Akaike H. Fitting autoregressive for prediction. *Ann I Stat Math* 1969; 21: 243–247.

19. Marple SL. *Digital spectral analysis: with applications*. Englewood Cliffs, NJ: Prentice Hall, 1987.

20. Bonnardot F, El Badouaia M, Randall RB, et al. Use of the acceleration signal of a gearbox in order to perform angular resampling (with limited speed fluctuation). *Mech Syst Signal Pr* 2005; 19: 766–785.

21. Urbanek J, Barszcz T, Sawalhi N, et al. Comparison of amplitude based and phase based methods for speed tracking in application to wind turbines. *Metrol Meas Syst* 2011; 18: 295–304.

22. Sabini EP, Lorenc JA, Henyan O, et al. Bearing defect detection using time synchronous averaging (TSA) of an enveloped accelerometer signal. US Patent no. US6681634B2, 2004.

23. McFadden PD and Toozhy MM. Application of synchronous averaging to vibration monitoring of rolling element bearings. *Mech Syst Signal Pr* 2000; 14: 891–906.

24. Luo HG, Qiu H, Ghanime G, et al. Synthesized synchronous sampling technique for differential bearing damage detection. *J Eng Gas Turb Power: T ASME* 2010; 132: 072501 (8 pp.).

25. Siegel D, Al-Atat H, Shauche V, et al. Novel method for rolling element bearing health assessment—a tachometerless synchronously averaged envelope feature extraction technique. *Mech Syst Signal Pr* 2012; 29: 362–376.