Journal of Physics: Conference Series **628** (2015) 012088 doi:10.1088/1742-6596/628/1/012088

Autoregressive modelling for rolling element bearing fault diagnosis

H Al-Bugharbee*[1,2] and I Trendafilova[1]

[1]Department of Mechanical and Aerospace engineering

University of Strathclyde, James Weir Building, 75 Montrose Street, Glasgow G1 1XJ, United Kingdom

[2 On leave] Department of Mechanical engineering, University of Wasit, Wasit, Iraq

E-mail: hussein.al-bugharbee@strath.ac.uk

Abstract. In this study, time series analysis and pattern recognition analysis are used effectively for the purposes of rolling bearing fault diagnosis. The main part of the suggested methodology is the autoregressive (AR) modelling of the measured vibration signals. This study suggests the use of a linear AR model applied to the signals after they are stationarized. The obtained coefficients of the AR model are further used to form pattern vectors which are in turn subjected to pattern recognition for differentiating among different faults and different fault sizes. This study explores the behavior of the AR coefficients and their changes with the introduction and the growth of different faults. The idea is to gain more understanding about the process of AR modelling for roller element bearing signatures and the relation of the coefficients to the vibratory behavior of the bearings and their condition.

1. Introduction

Diagnosis of faults in rolling element bearings is a very essential task in industry. There are so many methods are used for the purpose of fault diagnosis in rolling element bearings.

Mainly, the methods focused on analyzing the vibration signals in its various domains time, frequency and time-frequency. Several papers are published on reviewing and critical evaluation of the methods used [1, 2]. The principles of time series analysis is also used by some researchers to model the vibration signal acquired from the bearing house. Linear autoregressive is a very common model used for the purpose of fault diagnosis in machinery diagnosis. It is simply related each data point of the signal to a set of past data points using some weighted values (i.e. coefficients). However one of the challenges of these with using such models is that existence of non-stationary part in the signal. One of the solutions to cover this problem is use adaptive (time variant) autoregressive in which the model coefficients are evolving with time [3-5]. However, this solution is not that direct use as it requires the assumption of a suitable initial set of model coefficients as well as the assumption of the way of coefficients model evolving. Another way is to transform the non-stationary signals to non-stationary to

a stationary using a suitable stationarization technique and then to use the autoregressive of time invariant coefficients. It is important before starting of the modeling process is to test if the signal is stationary in order to decide if transformation is required or not. One of the most common tests is Kwiatkowski,Phillips, Schmidt, and Shin (KPSS) test [6].

In both cases time (variant or invariant) model, the model coefficients are very important features as they reflect the condition of the signal which provides the chance to use them as feature vectors of the signal. They can also be used to obtain the parametric spectrum of the signal.

Several papers are published in using autoregressive modeling for the machinery fault diagnosis problem [7-10]. However, according to the knowledge of the authors none of these papers analyse the model coefficients individually and that is the part what this study aims to do.

This study suggests the use of a linear AR model applied to the signals after achieving the local stationarity of the signal by testing and transformation if necessary. The obtained coefficients of the AR model are further used to form features vectors which are in turn subjected to pattern recognition for distinguishing among different faults and different fault sizes. This study also explores the behavior of the AR coefficients individually and their changes with the introduction and the growth of different faults. The goal is to gain more understanding about the relation of the of AR coefficients to the vibratory behavior of the bearings and their condition. The study is divided in three parts, namely: signal pre-treatment, diagnosis method and exploration of AR coefficients.

The paper divided in a couple of sections as follow: - section 2: signal pre-treatment. Section 3. Diagnosis method. Section 4. Exploration of AR coefficients. Section 5. Method verification**.** Section 6. Results and discussion. Section7. Conclusion.

2. Signal pre-treatment

The stage of signal pre-treatments includes a number of processes that starts by signal segmentation and finishes with obtaining AR model's coefficients.

2.1. Signal segmentation

Signal segmentation is essential step to provide more segments for analysis purposes especially when it is difficult to acquire more signals experimentally. In addition, it can be useful in enhance the stationarity of the signal where the whole signal can be non-stationarity but some/all of its segment are locally stationary [11].

2.2. Stationarity test and stationarization.

As we mentioned above that linear autoregressive model is suitable for only stationary signal (*i.e signals that the four statistical moments are time invariant*). It is necessary to investigate the stationarity of the signal before subjecting to modeling. This is important to decide whether the signal needs for stationarization transformation or not. The existence of stationarity can simply investigated by the Kwiatkowski, Phillips, Schmidt and Shin (KPSS) test which proposes a null hypothesis that the observed signal is stationary around its deterministic trend [6]. In case that a segment is discovered to be non-stationarity, then it can be subjected to a differencing technique which transforms it to stationarity. Differencing can be described as [12]:-

$$x_i^{dif} = x_i - x_{i-1} \quad , where\ i = 2,3,\dots n \tag{1}$$

Where x_i^{dif} are the components of the new stationarized segment, x_i and x_{i-1} are adjacent signal data points and n is the length of the original segment.

2.3. Linear autoregressive modeling.

A linear Autoregressive (LAR) model is that representation where the output variable is predicted based on linearly depending on its own previous values. In this research, it is used to predict the stationary segmented signals. The mathematical structure of a linear AR model can be described as follow [13]

$$x_i = a_o + a_1.x_{i-1} + a_3.x_{i-2} + \cdots + a_p.x_{i-p} + \varepsilon_i \quad (2)$$

where
x_i is the predicted output value at time i which is linearly related to (p) past points,
p is the order of the model
a_i (i =0,1,2…,p) are weighting coefficients (i.e model parameters).
ε_i ,the error term, is a white noise process, which represents the difference between the actual and linearly predicted values. In this research, the model order is determined based on final prediction error (FPE). The model coefficients are estimated using the least square algorithm. The goodness of model's fit to the experimental data is evaluated by the Normalized Mean Square Error (NMSE) value which given by the equation below: -.

$$Goodness\ of\ Fit|_{NMSE} = \left(1 - \frac{\left(\sum_{i=1}^{i=n}\left(\mathrm{x}_{pr}(i) - \mathrm{x}_{ms}(i)\right)^2\right)^{1/2}}{\sum_{i=1}^{i=n}\mathrm{x}_{pr}(i) - \bar{\mathrm{x}}_{ms}}\right) * 100\% \quad (3)$$

where: -
x_{pr} is predicted signal.
x_{ms} is the real time measured signal
$\bar{x}_{ms}$ is the mean value of real time signal.
n is the number of data points (i.e segment length)

3. Diagnosis method

In the suggested methodology, pattern recognition is used for distinguishing among different classes that corresponds to different bearing conditions. The methodology is suggested to detect different fault locations and different fault severity. In the first stage of diagnosis signals are segmented and distributed into two samples training and testing sample. From training sample the four bearing signal categories are made (i.e H, IRF, BF, ORF). Every segment of the training sample is tested for stationarity using KPSS test (see &2.2) and if required the stationarization transformation is done using equation (1). Next, the segments are subjected to AR modeling equation (2) and coefficients are obtained using least square method. These model coefficients are arranged as rows to form the four signal categories mentioned above (i.e H, IRF, BF and ORF).

Up to this step there will be four matrices corresponding to different bearing condition (i.e. H, IRF, BF and ORF) and another three corresponding for fault severity (S, M, L) of each fault location.

The size of each matrix will be (N X p) dimensions. The number of rows N equals to the number of segments of the training sample while the number of column (p) equals to the number of AR model coefficients (i.e the optimum model order). In case there are different model orders (p) corresponding to the different signal categories, the minimum one is considered to be the number of columns of features vectors matrix (H_K) as in below.

$$\mathbf{H}_K = \begin{bmatrix} a_{K11} & a_{K12} & . & a_{K1p} \\ a_{K21} & a_{K22} & . & a_{K2p} \\ . & . & . & . \\ a_{KN1} & a_{KN2} & . & a_{KNp} \end{bmatrix} \quad (4)$$

Now it is required to calculate the mean of the $\mathbf{H}_K$ rows in order to be used later in Mahalanobis distance calculation.
The mean vector is calculated as follows:

$$\mathbf{E}_K = \frac{\sum_{i=1}^{N}(a_{Ki1}\ldots\ldots\ldots\ldots a_{Kip})}{N} \quad (5)$$

Eventually the Mahalanobis distances are calculated between the new tested features vectors (i.e model coefficients vectors from testing sample) and mean feature vectors as defined in Eq. (5):

$$D_K(\mathbf{x}, \mathbf{E}_K) = (\mathbf{x} - \mathbf{E}_K).\mathbf{S}^{-1}.(\mathbf{x} - \mathbf{E}_K)' \quad (6)$$

where
D_K is mahalnobis distance.
$\mathbf{x}$ a feature vector from the testing sample
$\mathbf{E}_K$ mean features vector of category K
$\mathbf{S}^{-1}$ the inverse ofcategory covarience matrix $\mathbf{H}_K$

The prime in the end of equation (6) means the transpose.
Then the new feature vector is classified based on the NN method. According to this method, each new vector is assigned to the category for which the Mahalanobis distance (see Equation (6)) has a minimum. That is **x** belongs to the category *m* for which $D_K(\mathbf{x}, \mathbf{E}_K)$ has minimum over all *K*.

4. Exploration of AR coefficients behavior.

A further exploration of the AR coefficients is also carried out for further understanding of individual coefficients sensitivity to the different bearing conditions. In this regard, reference regions including upper and lower boundaries are built from the healthy values of each coefficient. These values are taken from the features vector matrix of healthy class ($\mathbf{H}_H$). The boundaries are determined using the mean and standard deviations of the healthy values as in equations (7&8). For a coefficient number (*i*), the boundaries formulas will be: -

$$b_{upper} = \mu\big(\boldsymbol{H}_H(:,i)\big) + d * \sigma\big(\boldsymbol{H}_H(:,i)\big) \quad (7)$$

$$b_{lower} = \mu\big(\boldsymbol{H}_H(:,i)\big) - d * \sigma\big(\boldsymbol{H}_H(:,i)\big) \quad (8)$$

$\mu\big(\boldsymbol{H}_H(:,i)\big)$ is the mean value of the column (*i*) of the matrix $\boldsymbol{H}_H$.

$\sigma\big(\boldsymbol{H}_H(:,i)\big)$ is the standard deviation value of column (*i*) of the matrix $\boldsymbol{H}_H$.

d is a constant, which should be properly selected so that at least 80% of the points are within the boundaries.

The boundaries are normalized by computing their percent deviation from the $\mu\big(\boldsymbol{H}_H(:,i)\big)$ as

$$bni = \frac{\mu\big(\boldsymbol{H}_H(:,i)\big) - bi}{\mu\big(\boldsymbol{H}_H(:,i)\big)} * 100\% \quad (9)$$

Where
bi upper or lower boundary corresponding to coefficient *i.*
bni normalized upper or lower boundary corresponding to coefficient *i.*

Then, to investigate the sensitivity of a parameter a_i to the change of bearing condition (i.e presence or growth of the fault), the values of a_i from other than healthy signals are normalized as in equation (10) to find their percent deviation from the reference values and projected to investigate if they deviate from the reference boundaries (sensitive) or not (insensitive).

$$\%\, dev_{KHi} = \frac{\mu\left(\boldsymbol{H}_H(:,i)\right) - \mu\left(\boldsymbol{H}_K(j,i)\right)}{\mu\left(\boldsymbol{H}_H(:,i)\right)} * 100\% \quad (10)$$

Where $\%\, dev_{KHi}$ is the percent deviation of coefficient *i* of category *K* from the reference value. *j =1:N* is the number of segment.

In order to reduce the effect of false alarm, coefficient a_i is assigned as a fault sensitive one if the $\%dev_{KHi}$values are out the previous reference region for several successive segments. When it is so, and new reference boundaries are built from the new sensitive a_i.

Finally, the common coefficients that are sensitive to fault presence and growth in different fault locations are determined. Eventually, the first three coefficients that showing higher values of deviation among the signal categories are selected for comparison.

5. Method verification.

Bearing vibration data provided by the Case Western Reserve University (CWRU) are used for the purpose of validation[14]. The data considered in this research are shown in Table 1. The raw signal is segmented equally into 2048 points sub signals.

The segmentation process gives 232 segments (4 categories*58 segments for each category) for detection of fault location and 174 segments (3 categories * 58 segments for each category) for detection of fault severity. These are divided equally into training and testing sample. All the segments are checked to be stationary before subjected to modeling.

6. Results and Discussion.

Table 1. The List of data used for analysis. H-healthy, IRF-inner race fault, BF- ball fault, ORF-outer race fault, S-small fault 0.007inch (fault diameter), M=Medium size fault 0.014'', B-big fault 0.021''.

Motor speed(rpm)	Signal Category for fault identification (for every motor speed)	Signal Category for fault severity estimation (for every motor speed)
1797 1772 1750 1730	Data mix set (Healthy, (IRF, BF,ORF)$_{\text{fault size 0.007''}}$)	• IRF set (S, M, B) • BF set (S, M, B) • ORF set (S, M, B)

Tables 2&3 show the correct classification rates of the testing segments for detection of fault locations and detection of fault severity respectively. In Table 2 all apart from those segments which recorded at speed of 1772 rpm are totally correctly classified. In Table 3, all the segments are perfectly classified for all the speeds considered in this analysis.

The first three most fault sensitive model coefficients are shown in Table 4. It is clear that these coefficients sets are not identical. However, some remarks can be highlighted. Among all the cases, the coefficient number 3 (i.e a_3) are existed within the first three higher sensitivity for the change of bearing conditions.

This can be helpful in using it as condition monitoring index. The a_4 are also present as a sensitive coefficient for fault location change or fault growth in the inner and outer races. However this is not completely true when the speed is further increased to 1797 or when the fault growth is monitored on the ball.

Table 2. Classification rate of testing samples.

Motor speed (rpm)	Signal Category used in classification	No.of Test samples	Diagnosis Accuracy %
1797	Healthy, $(IRF, BF,ORF)_{0.007''}$	116	100%
1772	Healthy, $(IRF, BF,ORF)_{0.007''}$	116	97.4%
1750	Healthy, $(IRF, BF,ORF)_{0.007''}$	116	100%
1730	Healthy, $(IRF, BF,ORF)_{0.007''}$	116	100%

Table 3. Classification rate of fault size estimation.

Motor speed (rpm)	Signal Category used in classification	No.of Test samples	IRF_{SMB}	BF_{SMB}	ORF_{SMB}
1797	**S**mall, **M**edium, **B**ig	174	100%	100%	100%
1772	**S**mall, **M**edium, **B**ig	174	100%	100%	100%
1750	**S**mall, **M**edium, **B**ig	174	100%	100%	100%
1730	**S**mall, **M**edium, **B**ig	174	100%	100%	100%

Table 4. Model coefficients of first three higher fault sensitivity ad *d=1.5 as in equation (7&8).*

Motor speed	**Sensitive coefficients set**			
(rpm)	**Data Mixed**	**IRF_{SMB}**	**BF_{SMB}**	**ORF_{SMB}**
1797	[8 3 6]	[3 8 2]	[2 4 3]	[4 3 2]
1772	[4 3 9]	[4 3 5]	[3 1 2]	[3 4 2]
1750	[4 9 3]	[4 3 5]	[3 6 5]	[4 3 2]
1730	[4 3 5]	[4 3 5]	[4 3 5]	[4 3 2]

Figure 1 shows an example of how the coefficient a_4 responses to different bearing fault location at 1730 rpm.

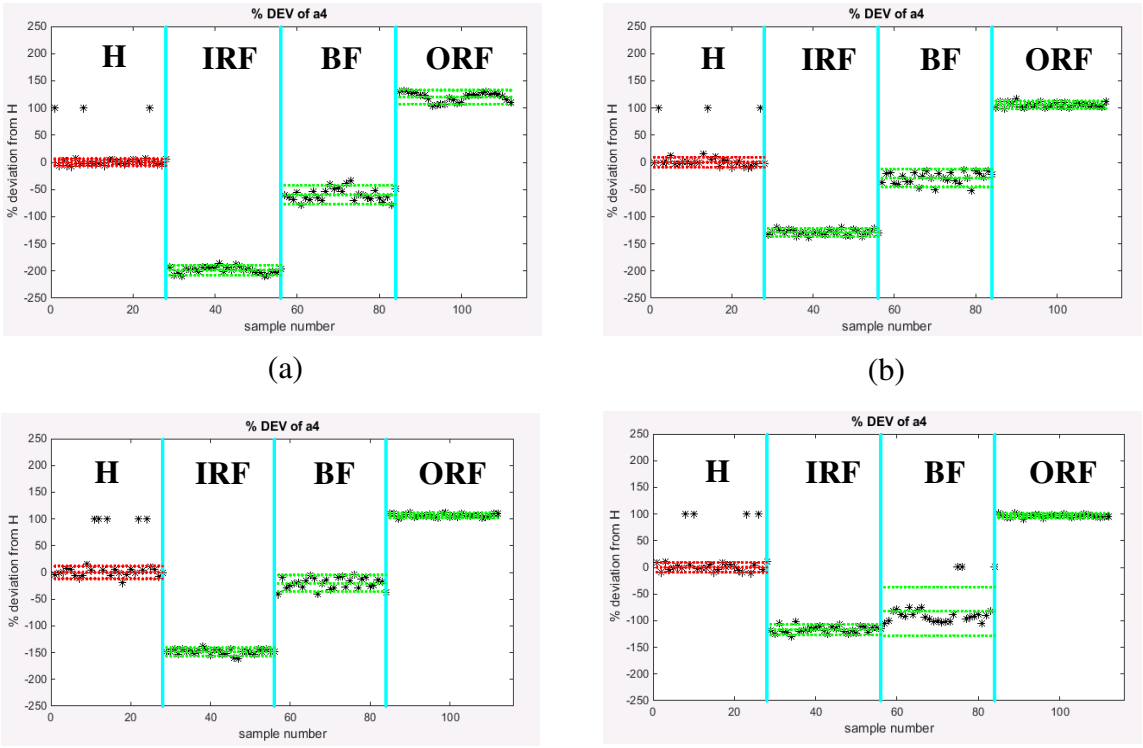

Figure 1. The behaviour of coefficient (a_4) in correspondence to the change of bearing condition (i.e fault location). a) 1797 rpm b) 1772 rpm c) 1750 rpm d)1730 rpm.

7. Conclusion.

In this study, a methodology for fault diagnosis in rolling element bearings is introduced. It shows a very good performance in detection different types and severity of faults. The further analysis of model coefficients shows that there are certain coefficients which are more sensitive to the fault presence and growth when compared to their reference values. However, these sets of coefficients are not unique for all the cases but some of them are common for different bearing conditions. This study is useful in extraction some indices that can be used for bearing condition monitoring.

References

[1] Tandon N, Choudhury A 1999. *Tribology international*. A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings.**32**(8).469-80.

[2] Nataraj C, Kappaganthu K 2011. *13 World Congress in Mechanism and Machine Science;* Vibration on- based diagnostics of rolling element bearings: state of the art and challenges; Guanajuato, Mexico.

[3] Arnold M, Milner X, Witte H, Bauer R, Braun C 1998. *Biomedical Engineering, IEEE Transactions on*. Adaptive AR modeling of nonstationary time series by means of Kalman filtering.**45**(5).553-62.

[4] Zhan Y, Jardine A 2005. *Journal of Sound and Vibration*. Adaptive autoregressive modeling of non-stationary vibration signals under distinct gear states. Part 1: modeling.**286**(3).429-50.
[5] Zhan Y, Jardine A 2005. *Journal of sound and vibration*. Adaptive autoregressive modeling of non-stationary vibration signals under distinct gear states. Part 2: experimental analysis.**286**(3).451-76.
[6] Kwiatkowski D, Phillips P, Schmidt P, Shin Y 1992. *Journal of econometrics*. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?;**54**(1).159-78.
[7] Altmann J, Mathew J 2001. *Mechanical systems and signal processing*. Multiple band-pass autoregressive demodulation for rolling-element bearing fault diagnosis.**15**(5).963-77.
[8] Thanagasundram S, Schlindwein F 2006. *Proceedings of ISMA2006 Noise and Vibration Engineering Conference;* Autoregressive based diagnostics scheme for detection of bearing faults.
[9] Thanagasundram S, Spurgeon S, Soares Schlindwein F 2008. *Journal of Sound and Vibration*. A fault detection tool using analysis from an autoregressive model pole trajectory.**317**(3).975-93.
[10] Wang G, Luo Z, Qin X, Leng Y, Wang T 2008. *Mechanical Systems and Signal Processing*. Fault identification and classification of rolling element bearing based on time-varying autoregressive spectrum.**22**(4).934-47.
[11] Pardey J, Roberts S, Tarassenko L 1996. *Medical engineering & physics*. A review of parametric modelling techniques for EEG analysis.**18**(1).2-11.
[12] Mader H 2006. Statistics in volcanology: Geological Society of London.
[13] Box G, Jenkins G, Reinsel G 2013. *Time series analysis: forecasting and control*: John Wiley & Sons.
[14] Website CWRUBDC. [cited 2014 01-10-2014]. Available from: http://csegroups.case.edu/bearingdatacenter/pages/welcome-case-western-reserve-university-bearing-data-center-website.