2011 American Control Conference
on O'Farrell Street, San Francisco, CA, USA
June 29 - July 01, 2011

Model-Based Adaptive Frequency Estimator for Gear Crack Fault Detection

Geoff McDonald and Qing Zhao, *Member, IEEE*

***Abstract*— Detection of gear cracks from vibration data is a difficult task. This paper investigates an alternative to the linear predictor residual fault detection based on the nonlinear adaptive control system concept of frequency estimators. The frequency estimator model takes advantage of the sinusoidal nature of vibration and adapts the system model during operation. The low-computational requirements, no-priori knowledge, sinusoidal-based prediction, and on-line model adaption makes this model ideal for on-line gear crack fault detection. Performance is evaluated through both synthetic and experimental data while comparing to the autoregressive linear predictor model.**

I. INTRODUCTION

Gear crack faults in rotating machinery are a serious problem which can cause possible machine catastrophic failure and potential loss of life [1]. Recently in vibration-based fault detection there has been a trend in the field towards detecting these faults through vibration measurements using autoregressive (AR) matched model prediction error signals [2], [3], [4]. AR models are linear models matched to the vibration signal under no-fault conditions, and the model is applied to predict the next vibration sample and compared to the actually measured sample. This error in prediction, the residual, is well-suited to extract the impulse-like features associated with gear crack vibration.

Although a vibration signal is well-modelled as a linear system, such as an autoregressive model, due to the nature of vibration being sinusoidal it can be improved through prediction based on a sinusoid restriction [5]. Recently, adaptive system theory has been applied towards real-time estimation of multiple frequency components of signals [6]. This frequency estimator (FE) is in the form of an adaptive state-space model and forms a nonlinear system. This model is not only able to predict future samples based on a sum of sinusoids model, but also adapts the sinusoidal model during operation. This results in the ability to predict samples on-line even with time-variant sinusoid components and no priori-knowledge requirements such as the data fitting required in the AR approach. Despite the relatively complex operation of the FE method, this paper demonstrates that it is of comparable calculations-per-sample to a low order AR algorithm. Clearly the FE model is well-suited to the nature of vibration signals and poses application in improving the field of gear-crack fault detection and other fields.

G. McDonald and Q. Zhao are with the Advanced Control System Laboratory, Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada. glmcdona@gmail.com & qingzhao@ece.ualberta.ca

This paper investigates the application the adaptive system concept of FE for the detection of gear-crack faults in rotating machinery. Firstly, a discrete-time (DT) version of the continuous-time N-component FE proposed by M. Hou [6] is derived. Next a novel amplitude-invariant adaptive identifier is proposed. Finally, general model parameters are selected for a 2 and 3 component FE using a non-linear least-squares (LSQ) approach. For performance evaluation, the standard AR predictive model is compared to the FE predictive model in complexity, simulation performance, and experimental performance on a controlled gear-crack.

II. AUTOREGRESSIVE (AR) RESIDUAL METHOD

Towards detection of gear tooth faults in on-line vibration data, the autoregressive (AR) method is considered very effective due to the low computational requirements and little priori knowledge requirements [3]. The AR model is a linear time-invariant system of model order M. The AR model parameters are estimated using the windowless Burg's lattice-based approach.

The AR method first involves selecting a model order for the linear model. Similar to other gear crack applications [2], [3], [4], the minimum Akaike Information Criterion (AIC) [7] is applied to select the optimal model order. The AIC simplified for model comparison is calculated as

$$AIC = 2n + K\ln(RSS),$$

where n and K refer to the number of estimated parameters, and number of samples respectively. RSS refers to the residual sum of squares of the data fit, resulting in a balance between minimizing the prediction error while penalizing a high model order. For the duration of this paper, the AIC is calculated for model orders of $M = 1, 6, 11, ..., 146$, and the optimal model order is selected as the order corresponding to the minimum AIC.

Figure 1 illustrates the prediction residual method used for detecting gear cracks. The AR model is fit to the data under no-fault condition, and the AR model is then used to compare the predicted vibration level of each sample with the actual vibration level of each sample during machine operation. Because gear-crack faults are known to develop as impulse-like disturbances, ideally this residual signal will resemble an impulse repeating every gear rotation during a gear crack.

This prediction residual can then be further processed for fault detection.

978-1-4577-0079-8/11/$26.00 ©2011 AACC

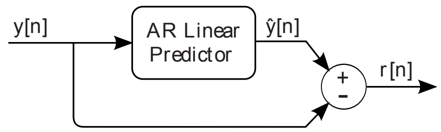

Fig. 1. AR-based linear-prediction error system

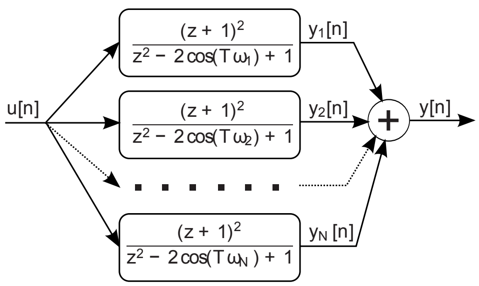

Fig. 2. N-component DT system model

III. N-Component Frequency Estimator (FE)

Based on the N-component FE model by M. Hou [6] for continuous-time, a low complexity discrete-time sample predictor is first derived. The derivation starts from the zero-order hold discretized model of the sum of sinusoid system for N-components, Figure 2.

Forming the transfer function for each sinusoidal component,

$$Y_i(z) = \frac{N(z)}{D_i(z)}U(z) = \frac{(z+1)^2}{z^2 - 2\cos(T\omega_i)z + 1}U(z),$$

the complete transfer function follows as

$$Y(z) = U(z)\sum_{i=1}^{N} Y_i(z) = \frac{P(z)}{Q(z)}U(z),$$

where $P(z)$ and $Q(z)$ refer to the numerator and denominator polynomial of the complete transfer function,

$$\begin{aligned} Q(z) &= \prod_{i=1}^{N}(z^2 - 2\cos(T\omega_i)z + 1) \\ &= z^{2N} + \theta_{N-1}z^{2N-1} + \theta_{N-2}z^{2N-2} + \cdots \\ &\quad + \theta_0 z^N + \cdots + \theta_{N-2}z^2 + \theta_{N-1}z + 1, \qquad (1) \end{aligned}$$

and θ_i's refer to the resulting expanded polynomial coefficients. By introducing a Hurwitz polynomial,

$$\alpha(z) = z^{2N} + \alpha_{2N-1}z^{2N-1} + \cdots + \alpha_1 z + \alpha_0$$

we rewrite the transfer function as

$$Y(z) = \frac{P(z)}{\alpha(z) + (Q(z) - \alpha(z))}U(z).$$

Rearranging we have,

$$Y(z) = \frac{\alpha(z) - Q(z)}{\alpha(z)}Y(z) + \frac{P(z)}{\alpha(z)}U(z)$$

and taking $U(z) = 0$ we have

$$Y(z) = \frac{\alpha(z) - Q(z)}{\alpha(z)}Y(z)$$

with estimator state-space realization of

$$x[n+1] = Ax[n] + By[n]$$

$$\hat{y}[n] = \hat{C}x[n]$$

$$A = \begin{bmatrix} -\alpha_{2N-1} & -\alpha_{2N-2} & \cdots & -\alpha_1 & -\alpha_0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$$

$$B = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \hat{C} = \alpha^T - \hat{\theta}^T V$$

$$\alpha = \begin{bmatrix} \alpha_{N-1} \\ \alpha_{N-2} \\ \vdots \\ \alpha_1 \\ \alpha_0 - 1 \end{bmatrix}, \hat{\theta} = \begin{bmatrix} \hat{\theta}_{N-1} \\ \hat{\theta}_{N-2} \\ \vdots \\ \hat{\theta}_0 \end{bmatrix}$$

$$V = \begin{bmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 & 1 & 0 \\ 0 & 1 & \cdots & 0 & \cdots & 1 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \iddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 0 \end{bmatrix}$$

where the system output, $\hat{y}[n]$, is the signal estimation and $\hat{\theta}$ corresponds to the estimated frequency components $\hat{\omega}$ by Equation 1. The parameter α is tunable and controls the convergence behaviour of the system, and the selection of these parameters are discussed in Section V. The $\hat{\theta}$ estimates are adapted according to the residual as described in the following section.

IV. Adaptive Identifiers

Updating the frequency estimates, $\hat{\omega}$, is achieved through updating the related Hurwitz parameters, $\hat{\theta}$. The $\hat{\theta}$ parameter is updated according to the prediction residual, $r[n] = y[n] - \hat{y}[n]$, such that the system converges $\hat{\theta} \to \theta$. A simple adaptive identifier can be formed by simply adjusting $\hat{\theta}$ such that the residual decreases:

$$\hat{\theta}[n] = \hat{\theta}[n-1] - \gamma V x[n-1] r[n-1]$$

where γ is a tunable scalar parameter controlling the rate at which $\hat{\theta}$ is adapted. A smaller value will result in a longer convergence time but is more robust to noise, while a larger value improves the convergence time but reduces noise tolerance. One drawback of this possible adaptive identifier is the amplitude sensitivity. By multiplying the input signal by a scalar, $y[n] = ks[n]$, we have

$$x[n+1] = Ax[n] + Bks[n]$$

$$\frac{x[n+1]}{k} = \frac{Ax[n]}{k} + Bs[n]$$
$$k\hat{s}[n] = \hat{C}x[n]$$

then redefining $\bar{x}[n] = x[n]/k$, we have the system

$$\bar{x}[n+1] = A\bar{x}[n] + Bs[n]$$
$$\hat{s}[n] = (\alpha^T - \hat{\theta}^T V)\bar{x}[n]$$
$$\hat{\theta}[n] = \hat{\theta}[n-1] - \gamma V k\bar{x}[n-1](ks[n] - k\hat{s}[n])$$

clearly resulting in a k^2 factor increase in the changes to $\hat{\theta}[n]$ each iteration and in turn resulting in significantly larger changes to the estimated frequencies $\hat{\omega}$. Ideally, the change to $\hat{\theta}[n]$ should be independent of the signal amplitude such that the parameter γ does not need to be tuned for each input signal. By introducing a state-normalization factor to the adaptive identifier, the amplitude sensitivity issue can be addressed,

$$\hat{\theta}[n] = \hat{\theta}[n-1] - \frac{\gamma V x[n-1](y[n-1] - \hat{y}[n-1])}{\|x[n-1]\|^2}$$

and following a similar procedure as above with input signal $y[n] = ks[n]$,

$$\begin{aligned}\hat{\theta}[n] &= \hat{\theta}[n-1] - \frac{\gamma V k\bar{x}[n-1](ks[n-1] - k\hat{s}[n-1])}{\|k\bar{x}[n-1]\|^2} \\ &= \hat{\theta}[n-1] - \frac{\gamma V \bar{x}[n-1](s[n-1] - \hat{s}[n-1])}{\|\bar{x}[n-1]\|^2}\end{aligned}$$

where the change to $\hat{\theta}[n]$ is now independent of the input signal amplitude. This proposed change introduces a singularity at $|x[n]| = 0$, and the adaptive identifier is slightly adjusted such that

$$\hat{\theta}[n] = \hat{\theta}[n-1] - \Delta\hat{\theta}[n-1]$$
$$\Delta\hat{\theta}[n] = \begin{cases} \frac{\gamma V x[n](y[n]-\hat{y}[n])}{\|x[n]\|^2} & \text{if} \|x[n]\|^2 > \epsilon \\ 0 & \text{otherwise} \end{cases}$$

where ϵ is a small positive scalar. Although this simplifies the model by being amplitude invariant, the selection of α and γ is still a difficult process due to the non-linear characteristics of the system. For the rest of the paper only this FE State-Normalized adaptive identifier is studied, as the results generalize to any input amplitude. The next section presents a least-squares minimization approach to selecting these parameters.

V. Least-Squares (LSQ) Parameter Selection

Parameter selection in a non-linear system is a difficult process. This problem is approached as a minimization problem where the input parameters, γ and α, are adjusted to minimize the error in the estimated $\hat{\theta}[n]$. Minimizing the estimated frequency error, $\hat{\omega}[n]$, is not considered due to the complexity involved with calculating $\hat{\omega}[n]$ from $\hat{\theta}[n]$. LSQ minimization is performed as

$$\min_{\gamma,\alpha} \|f_\theta\|_2^2$$

where f_θ is defined as the error in θ estimates in K-sample trajectories for L randomly generated input signals. Each input signal includes additive white Gaussian noise and N frequency components with uniformly-distributed frequency and initial phase. The L input signals are known and their corresponding correct θ is calculated by Equation 1. This forms the cost matrix,

$$f_\theta = \begin{bmatrix} e_\theta[0] & e_\theta[1] & \cdots & e_\theta[K-1] \end{bmatrix}$$
$$e_\theta[n] = \begin{bmatrix} \theta_{y_0} - \hat{\theta}_{y_0}[n] \\ \theta_{y_1} - \hat{\theta}_{y_1}[n] \\ \vdots \\ \theta_{y_{L-1}} - \hat{\theta}_{y_{L-1}}[n] \end{bmatrix}$$

used for the LSQ minimization. Minimizing this least-squares non-linear minimization problem is approached using the Levenberg-Marquardt with line-searching algorithm [8], [9]. LSQ initial parameters of $\alpha = \begin{bmatrix} 0 & 0 & \cdots & 0 & -1 \end{bmatrix}^T$ and $\gamma = 0$ are used for the adaption and simulation initial conditions of $x[0] = 0$, and $\hat{\theta}[0] = 0$ are used. Table I indicates the final FE parameter values after 10 iterations with $K = 2000$, $M = 500$, and additive Gaussian white noise of zero mean and 0.1 standard deviation.

TABLE I
Final values for FE parameters γ and α by LSQ minimization

N	γ	α
2	0.522	$\begin{bmatrix} -0.163 & 0.255 & -0.0840 & -0.911 \end{bmatrix}^T$
3	1.06	$\begin{bmatrix} 0.227 & 0.834 & 0.135 & 0.357 & -0.0607 & -1.09 \end{bmatrix}^T$

Figure 3 plots the convergence behaviour of $\hat{\theta}$ for the first four input signals of the 3-component FE after iterative parameter selection. It can be seen that the estimates typically converge correctly in a relatively short number of samples.

VI. Computational Requirements

The computational requirements of both the M-order AR model and the N-component FE is important for implementation. For the FE, the back-calculation of the $\hat{\omega}$ from $\hat{\theta}[n]$ is not included because this information is not required in the context of the proposed fault detection. Table II below indicates the computational requirements of the M-order AR model and N-component FE using both adaptive identifiers. All models were implemented in controllable canonical form and computations per sample were calculated.

For performance comparison purposes, an AR model of order $M = 9$ is included in the comparison as it is of similar complexity to the 3-component FE model.

VII. Simulation Results

For validation, a simulated gear-crack vibration signal is generated with varied levels of fault. The vibration signal is formed as harmonics of the motor vibration of 60 Hz

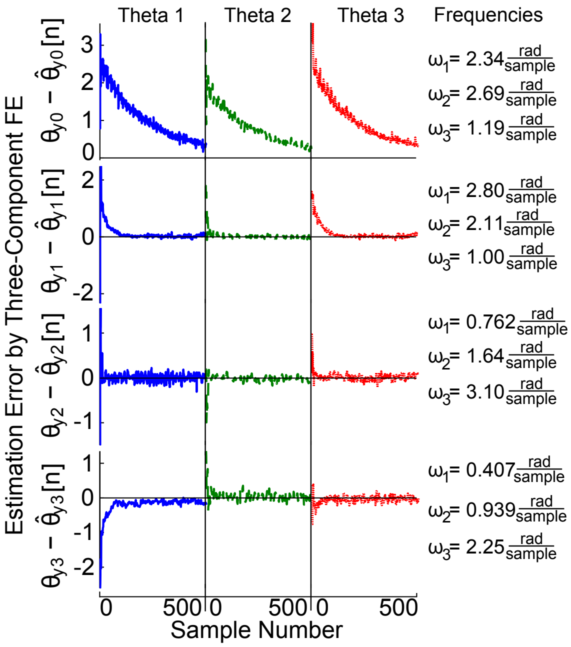

Fig. 3. Convergence of θ estimates for the three-component FE for the first four input signals each composed of three sinusoidal components at indicated frequencies.

TABLE II

FLOATING POINT CALCULATIONS REQUIRED FOR EACH SAMPLE OF PREDICTED OUTPUT AS A FUNCTION OF MODEL ORDER, M, OR NUMBER OF FREQUENCY COMPONENTS, N.

Method	Additions per sample	Multiplications per sample
AR	$2M-1$	$2M$
FE Classical	$5N-1$	$3N+2$
FE State-Normalized	$6N-2$	$5N+3$

plus fault modelled as time-localized decaying exponential enveloped sinusoid vibrations repeating at the gear rotational period. The motor harmonics of 60 Hz, 120 Hz, ..., 360 Hz have amplitudes of 0.5, 0.2, 0.1, 0.02, 0.03, and 0.01 respectively and initial phase of 1.04, 1.82, 6.60, 2.83, 4.08, and 0.00 rad respectively. Additive white Gaussian noise with zero mean and 0.001 standard deviation is included. The disturbance gear crack vibration signal is added to the signal at a period of 83 milliseconds (12 Hz) and is composed of a decaying exponential envelope with time constant 2e-3 and carrier frequency of 360 Hz. Figure 4 plots the simulated vibration signal for a gear-crack peak disturbance of 0.362 and sampling rate of 10 kHz.

Figure 5 presents the residual results when model-matching the AR models to the no-fault vibration data, and applying the FE and resulting AR models to the fault vibration signal in Figure 4.

Kurtosis is defined as the fourth standardized moment,

$$k = \frac{E(x-\mu)^4}{\sigma^4}$$

where $E()$ denotes the statistical expectation, x the signal, μ the signal mean, and σ the signal standard deviation.

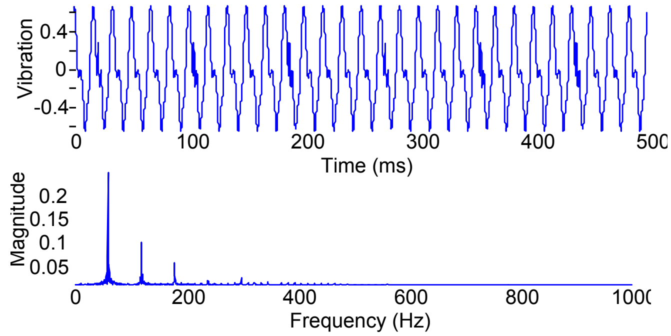

Fig. 4. Simulated gear crack vibration signal and spectrum.

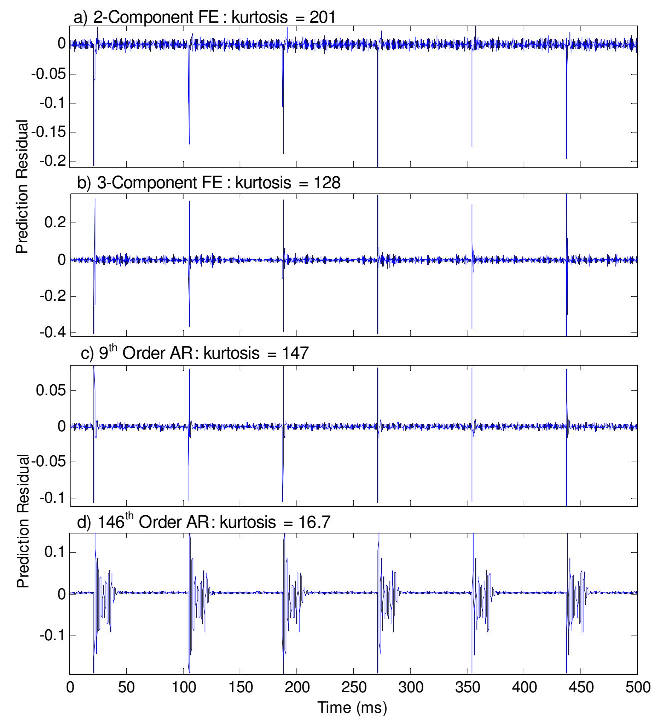

Fig. 5. Prediction residual and corresponding kurtosis using the a) 2-component FE, b) 3-component FE, c) 9th order AR, and d) 146th order AR model.

Kurtosis is a measure of how outlier-prone a signal is, and is commonly used as a fault indicator in gear-related faults due to the impulse-like vibration manifestation. The kurtosis values are analyzed to evaluate the FE and AR methods' ability to detect the gear crack fault. Comparing the fault peak disturbance versus the kurtosis, Figure 6, the 2-component FE performs the best for higher fault amplitudes while the 9th order AR model performs the best at small amplitudes. However, these results vary drastically according to the composition of the vibration signal and should not be interpreted as a general performance trend.

VIII. EXPERIMENTAL DESIGN AND RESULTS

For validation and comparison, vibration data collected from a controlled gear tooth crack experiment under varying degrees of crack severity is analyzed. The machine configu-

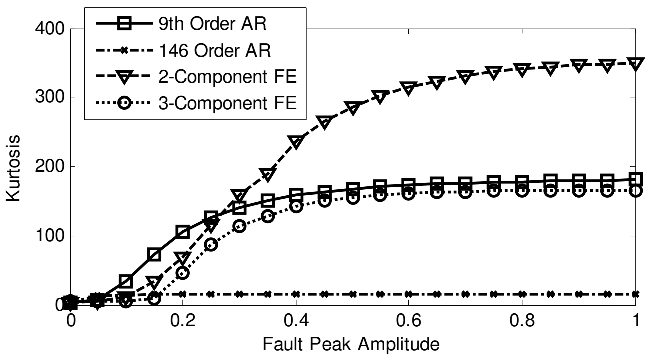

Fig. 6. Kurtosis of model prediction residual plotted versus fault peak disturbance.

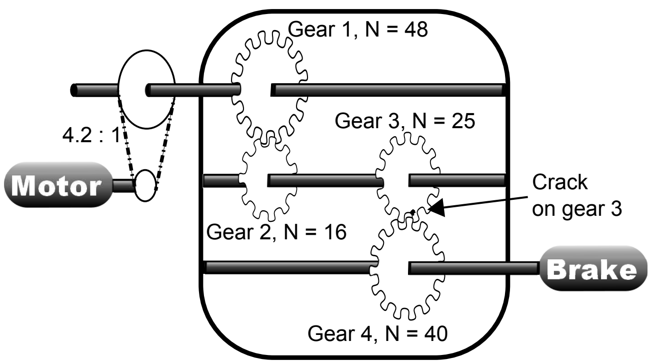

Fig. 7. Experimental machine configuration.

ration, Figure 7, is composed of a motor, gearbox, and break.

A tooth crack is introduced in Gear 3 at varying levels using a cutting knife of width 0.4 mm controlling the depth and length of the cut. The gears for fault levels of 25%, 50%, 75%, and 100% can be seen in Figure 8.

The experiment consists of recording of the vibration along the axis of the impact force for three separate measurements during three separate machine start-ups under each operating condition. The motor operating conditions are varied between full-load, half-load, and no-load, with speeds ranging from 800 to 3000 rpm at 200 rpm intervals. In total, this results in 540 vibration measurements with approximately 8000 samples each. The sampling frequency is dependent on the rotation speed of the motor. Figure 9 illustrates the averaged rms vibration of the machine versus fault level. It can be seen that the vibration trend does not increase proportionally to the fault level; this is likely as a result of properties of the five different gears used or slight changes to the machine as the gear swapping is performed.

For the 540 datasets, the general data processing procedure is as follows:

1) Subtract the signal mean from each dataset.
2) Fit the AIC selected order and 9th order AR models to a dataset at no-fault condition at each rpm and load condition.
3) For each dataset, both AR models are used to predict the samples using the dataset's corresponding no-fault fit model. For example, a dataset measured at 3000 rpm, full-load, 25% fault level would use the AR models fit to the datasets at 3000 rpm, full-load, 0% fault level.
4) For each dataset, both FE models are simulated and the prediction residual calculated.
5) Calculate the kurtosis of each prediction residual signal while ignoring the first 200 samples to allow for the AR and FE models to converge.

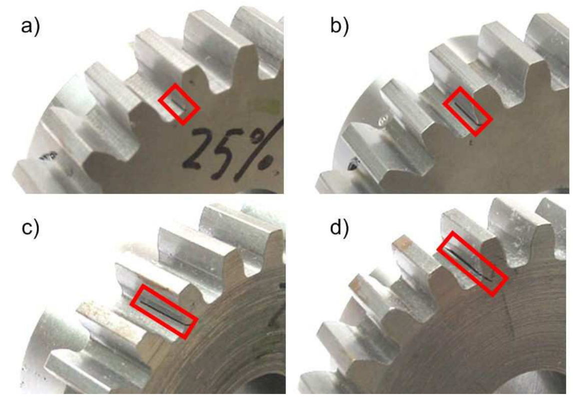

Fig. 8. Experimental gears with tooth cracks at levels of a) 25%, b) 50%, c) 75%, and d) 100% [10].

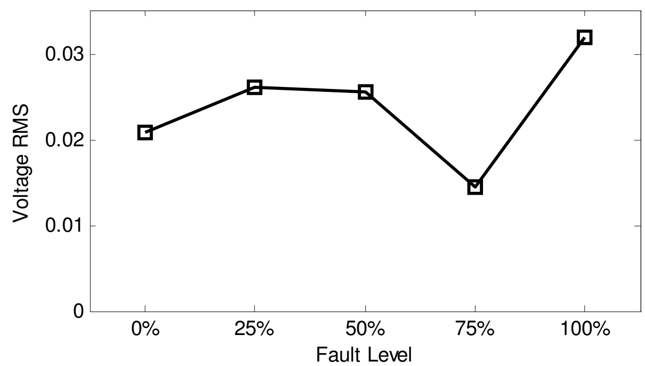

Fig. 9. RMS of measured acceleration according to fault level.

From this procedure, it is clear that one big advantage of the FE-based method is the fact that it is a general model requiring no training to different datasets, different machines, different motor speeds, or even different applications. As a result the FE model will perform better under varied machine conditions. Figure 10 presents the prediction error for all four methods with the machine at 3000 rpm, full load, and 100% fault level.

Figure 11 presents the no-fault normalized kurtosis results for the four predictive models under full-load, half-load, and no-load. It can be seen that the 2 and 3 component FE provide a better correlation between increasing fault level and higher kurtosis values. It is clear that the FE results in a better gear fault indicator over both AR models, with the 3-component FE performing the best.

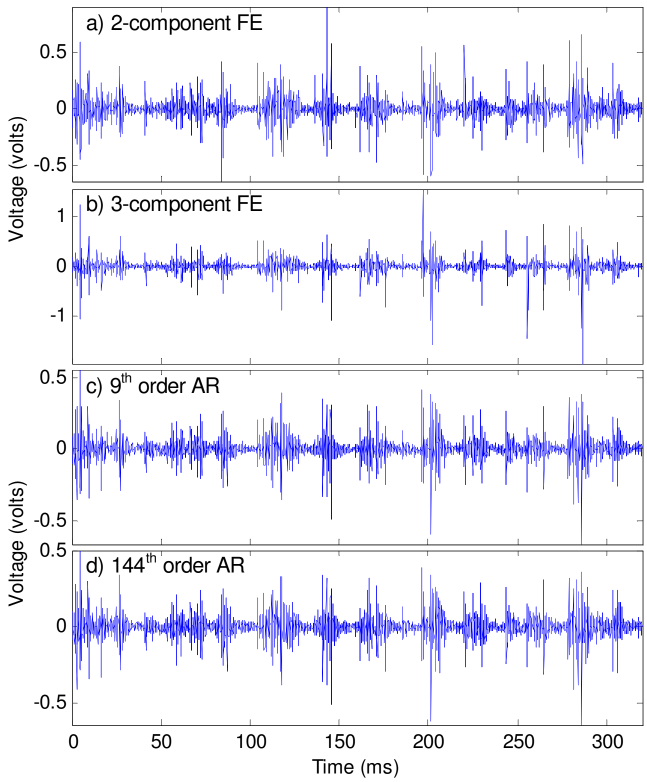

Fig. 10. Prediction error for the experimental setup at 3000 rpm, full load, and 100% fault level using a a) 2-component FE, b) 3-component FE, c) 9th order AR model, and d) an AIC selected 146th order AR model.

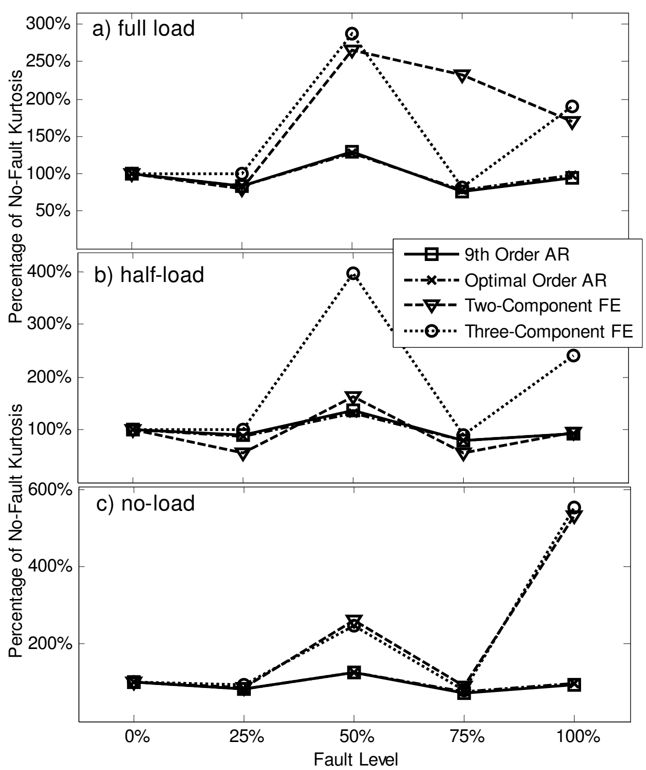

Fig. 11. Kurtosis of the signals averaged across all rpm settings and plotted as a percentage of no-load fault kurtosis under a) full load, b) half-load, and c) no load.

IX. Conclusion

In conclusion, this paper suggests the FE-based method not only detects gear crack faults better than the AR model based approach, but it requires no data fitting and is of similar computational requirements to a low order AR model. The experimental data indicates that the FE-based method outperforms the AR-based method in gear crack detection, while the simulation data suggests that the FE model only performs better at a higher fault level. Neither of these results are conclusive, but the FE model results proves a promising no priori-knowledge alternative to the AR model.

Further work should investigate the FE-based method on additional experimental setups, and additional simulation models. The maximum entropy deconvolution has been shown to greatly improve to the AR residual results [4], and similarly it can be applied to the FE residual. Better parameter selection, convergence analysis, and higher order frequency estimators can be investigated for the FE model approach.

Acknowledgment

The authors would like to thank Dr. Ming J Zuo, Xianfeng Fan, Yao Wang, Wei Li, Siyan Wu, and Jian Qu from the Reliability Research Lab, Mechanical Engineering University of Alberta, for their support and experimental data on the gear crack vibration experiment. We also wish to thank NSERC, Canada and Syncrude Canada Ltd. for making this possible through their funding support.

References

[1] P. D. Samuel and D. J. Pines, "A review of vibration-based techniques for helicopter transmission diagnostics," *Journal of Sound and Vibration*, vol.282, no.1–2, pp.475–508, April 2005

[2] W. Wang and A. Wong, "Autoregressive model-based gear fault diagnosis," *Journal of Vibration and Acoustics* vol.124, no.2, April 2002

[3] W. Wang, "An evaluation of some emerging techniques for gear fault detection," *The International Journal of Structural Health Monitoring*, vol. 2, no.3, pp. 225242, September 2003

[4] H. Endo and R. Randall, "Enhancement of autoregressive model based gear tooth fault detection technique by the use of minimum entropy deconvolution filter," *Mechanical Systems and Signal Processing* vol.21, no.2, February 2007

[5] A. Gerasimov, O. Morozov, E. Soldatov and V. Fidelman, "Using the method of modified linear prediction for robust acoustic coding of speech," *Journal of Radiophysics and Quantum Electronics*, vol. 49, no.7, pp.535-539, July 2006

[6] M. Hou, "Estimation of sinusoidal frequencies and amplitudes using adaptive identifier and observer," *IEEE Transactions on Automatic Control*, vol.52, no.3, March 2007

[7] S. Kay, *Modern spectral estimation: theory and application*, Prentice Hall pp. 106-118, 161-178, 228-237, 1988

[8] K. Levenberg, "A method for the solution of certain problems in least-squares," *Quarterly Applied Math* vol.2, pp.164168, 1944

[9] D. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *SIAM Journal Applied Math.* vol.11, pp.431441, 1963

[10] X. Fan, Y. Wang, W. Li, S. Wu, and M. Zuo, "Experiment design of gear crack growth," Reliability Research Lab, Dept. Mech. Eng., Univ. of Alberta, Edmonton, AB, Tech. Rep., August 2005