Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Sound and Vibration 286 (2005) 429–450
JOURNAL OF SOUND AND VIBRATION
www.elsevier.com/locate/jsvi

Adaptive autoregressive modeling of non-stationary vibration signals under distinct gear states. Part 1: modeling

Y.M. Zhan, A.K.S. Jardine*

Department of Mechanical and Industrial Engineering, University of Toronto, 5 King's College Road, Toronto, Ont., Canada M5S 3G8

Received 26 January 2004; received in revised form 24 September 2004; accepted 6 October 2004
Available online 28 December 2004

Abstract

Non-parametric time–frequency techniques are increasingly developed and employed to process non-stationary vibration signals of rotating machinery in a great deal of condition monitoring literature. However, their capacity to reveal power variations in the time–frequency space as precisely as possible becomes a hard constraint when the aim is that of monitoring the occurrence of mechanical faults. Therefore, for an early diagnosis, it is imperative to utilize methods with high temporal resolution, aiming at detecting spectral variations occurring in a very short time. This paper proposes three new adaptive parametric models transformed from time-varying vector-autoregressive model with their parameters estimated by means of noise-adaptive Kalman filter, extended Kalman filter and modified extended Kalman filter, respectively, on the basis of different assumptions. The performance analysis of the proposed adaptive parametric models is demonstrated using numerically generated non-stationary test signals. The results suggest that the proposed models possess appealing advantages in processing non-stationary signals and thus are able to provide reliable time–frequency domain information for condition monitoring.
© 2004 Elsevier Ltd. All rights reserved.

1. Introduction

By monitoring machine condition, either continuously or at regular intervals, changes in machine condition can be detected and repairs made in a timely manner, before breakdown

*Corresponding author. Tel.: +1 416 978 2921; fax: +1 416 946 5462.
E-mail addresses: yimin.zhan@utoronto.ca (Y.M. Zhan), jardine@mie.utoronto.ca (A.K.S. Jardine).

0022-460X/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jsv.2004.10.024

occurs. Any major piece of industrial machinery requires a certain degree of maintenance; often, this maintenance is condition based, that is, decisions regarding the repair or replacement of a machine part, overhauls, and standard maintenance are made on the basis of the actual condition of the machine. Thus proper machine condition monitoring procedures can result in lower maintenance costs and prolonged machine life. The most common family of machine condition monitoring methods is based on the analysis of vibration and acoustic signals, measured using a range of sensing techniques [1].

The machine condition signals often demonstrate a highly non-stationary property due to the fact that defects and incipient failures often manifest themselves in the form of changes in the spectrum of a measured signal. This phenomenon has increasingly impelled the applications of non-parametric joint time–frequency (T–F) methods to the analysis of non-stationary machine vibration signals since they are able to produce an overall view of the behavior of non-stationary signals by means of the so-called time-varying spectrum which is defined in the T–F space and represents the evolution of signal power as a function of both time and frequency [2]. It has seen a surge of such activities in the last decade. The most extensively employed non-parametric techniques involve the short-time Fourier transform (STFT), Wigner–Ville (WVD) and Choi–Williams distributions (CWD), wavelet transform and their enhanced derivatives [3–9].

The most promising application field of such methodology is for the analysis of highly transient phenomena in machinery. The capacity to reveal power variations in the T–F space as precisely as possible becomes a hard constraint for non-parametric T–F techniques when the aim is that of monitoring the occurrence of mechanical faults. At their early stages, faults start as almost impulsive events and determine a change in the 'signature' of the signal in the T–F space [2]. Therefore, for an early diagnosis, it is compulsory to utilize methods with high temporal resolution, aiming at detecting spectral variations occurring in a very short time. However, the non-parametric T–F methods suffer from various limitations, respectively. If the number of sampling points is fixed, the effectiveness of STFT is limited by the fact that it is unlikely to achieve good resolutions in both time and frequency domains simultaneously due to 'uncertainty principle'. In addition, the resolution of STFT partly depends on the type of the window applied. The performance of WVD and CWD of Cohen class is seriously influenced by a so-called 'cross-terms' which indicates some spurious components that appear among the real frequency components and detrimentally affect the interpretation of the T–F distribution. The wavelet transform attracting extensive interests is widely documented in the applications to machine condition monitoring. Nevertheless, it is found that it can only recover the relatively stronger frequency components from the vibration signals and gives rise to significant error when confronting higher frequency components [4,10]. Therefore, it has not yet found its way into the mainstream of machinery diagnostics.

In comparison, the modern spectral analysis method is more effective. In the modern spectral analysis, the techniques of time series modeling (AR, MA and ARMA, etc.), known as parametric spectrum analysis methods, have been applied to vibration signals analysis of rotating machinery by using time-invariant coefficients [11–13]. As a consequence, both accuracy and resolution can be significantly improved. It is noteworthy that autoregressive integrated moving average (ARIMA) model as a powerful technique for analyzing non-stationary signals has been widely used in forecasting [14–16] and a variety of engineering subjects [17–19]. However, this

methodology cannot be employed to generate T–F representation since the non-stationary time series needs to be differenced first until it is stationary and thus constant ARMA model coefficients are obtained by model identification.

Usually, the autoregressive model (AR) or its multivariate derivative vector autoregressive model (VAR) is most preferred since it is the best compromise between temporal representation and speed, efficiency and simplicity of algorithms enabling the estimation of model parameters. In practice, the spectrum of ARMA process could even be represented purely in terms of the AR coefficients without resort to compute the MA coefficients [20]. In view of the time-varying frequency components and magnitudes of non-stationary multivariate vibration signals, it is natural to assume the coefficient matrices of VAR model to be time-varying. Therefore, the conventional Levinson–Durbin recursion-based estimation procedure utilized in Ref. [13] constrains their AR model to some specific stationary conditions only and therefore does not possess an inherent structure for processing non-stationary vibration signals. Conforto and D'Alessio [2] presented a comparative study of time-varying univariate AR model and non-parametric T–F distributions, i.e. STFT and CWD. A time-varying univariate AR model based on the same parameter estimation mechanism is also proposed and applied to process biological signals [21]. However, they both presume a deterministic evolution of the coefficients by making use of deterministic time-dependent basis functions, which, from a theoretical point of view, is not adequate in applications where a pure stochastic system is assumed. Further, the predetermined noise components of their models will dramatically reduce the robust property of model under non-stationary conditions.

Up to now, little attention has been focused on time-varying VAR models where the evolution law of time-varying coefficients is assumed to be stochastic, whereas the parameter estimation of time-varying multivariate time series models on the basis of advanced adaptive filtering theory for optimum condition-based maintenance (CBM) purposes in the sense of providing highly precise T–F domain information is rarely investigated. Therefore, on-line parametric modeling of rotating machinery subject to vibration monitoring is highly desirable for presenting accurate and high-resolution T–F representations so as to satisfy the requirements of CBM. Previous research related to Kalman filter-based time-varying AR model can be found from different fields, but without an efficient on-line estimation scheme for the noise covariance, e.g. Refs. [22–24]. Such a drawback will make the time-varying AR model unable to yield accurate estimation under highly noisy circumstances. It has not seen previous research addressing the potential of such a methodology for condition monitoring in the literature. Our previous study proposed two multivariate state space models transformed from a time-varying VAR model with a noise-adaptive Kalman filtering algorithm (NAKF) [25] and a modified extended Kalman filtering algorithm (MEKF) [26], respectively. In this study we advance our previous work by adding a new state space model on the basis of an extended Kalman filter (EKF) and present a thorough and comprehensive investigation under a variety of gear states.

The remainder of this paper is organized as follows. Section 2 presents three state space models transformed from a time-varying vector autoregressive model by means of NAKF, EKF and MEKF under three different model assumptions, and related aspects. The model evaluation using non-stationary simulated signals with abrupt changes is presented in Section 3. Conclusions are drawn in Section 4.

2. Description of models

2.1. Motivations

As we know, stationary AR model can only fit stationary time series. However, in machine condition monitoring practices, the vibration signals are usually non-stationary due to a variety of reasons, for example, the geometric irregularities or faults of some machinery components. Therefore, a time-varying AR model characterized by time-dependent model coefficients should be employed to fit the non-stationary vibration signals. But, how time-dependent are the AR model coefficients? Consequently, different assumptions can be made to describe the time-dependent feature of the coefficients of the time-varying AR model. The random walk assumption that assumes that the AR model coefficients are only disturbed by a white Gaussian noise results in the model I proposed in Section 2.2. Model II proposed in Section 2.3 considers a more general assumption that the evolution law of the coefficients of the time-varying AR model is subject to not only a white Gaussian noise but an unknown constant state coefficient matrix. Model III in Section 2.4 assumes that the unknown state coefficient is not constant but time-varying. From the viewpoint of engineering, the three models present three different assumptions to describe how the time-dependent coefficients of the time-varying AR model evolve when it is employed to fit non-stationary vibration signals.

2.2. Model I based on NAKF

In this section we present a state space model transformed from VAR model with time-dependent coefficients and provide recursive algorithms for the implementation of Kalman filter. A VAR process is a discrete-time multivariate linear stochastic process given by

$$y_i = \sum_{k=1}^{p} A_k y_{i-k} + \varepsilon_i \tag{1}$$

for $i = 1, 2, \ldots, N$, that is, the time series can be considered as the output of a linear all-poles filter driven by a white-noise signal with a flat spectrum, where N is the sample size, p the order of VAR model, y_i the ith measurement vector of dimension $d \times 1$, A_k the kth $d \times d$ coefficient matrix of the measurement y_{i-k}, and ε_i an $d \times 1$ sequence of zero-mean white Gaussian measurement noise. Considering the non-stationary property of vibration signatures we assume the coefficient matrices of the above VAR model to be time-varying

$$y_i = \sum_{k=1}^{p} A_k(i) y_{i-k} + \varepsilon_i. \tag{2}$$

To make use of the Kalman filtering algorithm, it is necessary to develop a state space representation of model (2). This can be achieved by rearranging the elements of the matrices of coefficients in a vector form using the vec-operator, which stacks the columns of a matrix on top of each other from the left to right side. Then, with the following notation:

$$a_i = \mathrm{vec}([A_1(i), A_2(i), \ldots, A_p(i)]^{\mathrm{T}}), \tag{3}$$

$$Y_i = (y_i^{\mathrm{T}}, y_{i-1}^{\mathrm{T}}, \ldots, y_{i-p+1}^{\mathrm{T}}), \tag{4}$$

$$C_i = I_d \otimes Y_i^{\mathrm{T}}, \tag{5}$$

where I_d is an $d \times d$ identity matrix, $\otimes$ is the Kronecker product, an appropriate state space representation of the VAR model with stochastic coefficients can be given by

$$a_{i+1} = f_i(a_i) + v_i, \tag{6}$$

$$y_i = C_{i-1}^{\mathrm{T}} a_i + \varepsilon_i, \tag{7}$$

where a_i is the $pd^2 \times 1$ state vector, v_i is an $pd^2 \times 1$ sequence of zero-mean white Gaussian state noise, uncorrelated with a_1 and ε_i, ε_i is the same as in Eq. (1) and uncorrelated with a_1 and v_i, and Eq. (7) has an adaptive time-varying coefficient C_{i-1}^{T} of dimension $d \times pd^2$. We also have $\hat{a}_0 (= E(a_0))$, the Gaussian $pd^2 \times 1$ initial state vector with covariance matrix$P_{0|0} (= \mathrm{Cov}(a_0))$, and the noise covariance matrices:

$$E\{v_k v_i^{\mathrm{T}}\} = Q_k \delta_{k-i}, \tag{8}$$

$$E\{\varepsilon_k \varepsilon_i^{\mathrm{T}}\} = R_k \delta_{k-i}, \tag{9}$$

where T denotes transposition, δ denotes the Kronecker delta sequence, Q_k denotes the covariance matrix of state noise and R_k denotes the covariance matrix of measurement noise. For convenient implementation, $\hat{a}_0$ and $P_{0|0}$ are arbitrarily chosen in model I and also in the following models II and III. Suppose that the evolution law of the state vector a_i is a random walk process which results in a state space representation below instead of Eqs. (6) and (7)

$$a_{i+1} = a_i + v_i, \tag{10}$$

$$y_i = C_{i-1}^{\mathrm{T}} a_i + \varepsilon_i. \tag{11}$$

Thus, with the aid of a standard Kalman filter the following recursive prediction equations:

$$P_{i|i-1} = P_{i-1|i-1} + Q_{i-1}, \tag{12}$$

$$\hat{a}_{i|i-1} = \hat{a}_{i-1}, \tag{13}$$

where $P_{i|i-1}$ is the one-step ahead prediction of the state covariance matrix, $P_{i-1|i-1}$ is the estimate (error) covariance matrix of the state vector, $\hat{a}_{i|i-1}$ is the one-step ahead prediction of the state vector, $\hat{a}_{i-1}$ is the optimal filtering estimate of the state vector a_{i-1} and updating equations

$$G_i = P_{i|i-1} C_{i-1} [C_{i-1}^{\mathrm{T}} P_{i|i-1} C_{i-1} + \hat{R}_i]^{-1}, \tag{14}$$

$$P_{i|i} = [I - G_i C_{i-1}^{\mathrm{T}}] P_{i|i-1}, \tag{15}$$

$$\hat{a}_{i|i} := \hat{a}_i = \hat{a}_{i|i-1} + G_i (y_i - C_{i-1}^{\mathrm{T}} \hat{a}_{i|i-1}) \tag{16}$$

can be obtained for $i = 1, 2, \ldots, N$, where G_i is the Kalman gain and the covariance matrix Q_i of state noise is assumed to be known a priori. From the incoming measurement information y_i and the optimal state prediction $\hat{a}_{i|i-1}$ obtained in the previous step, the innovations sequence

is defined to be

$$z_i := y_i - C_{i-1}^{\mathrm{T}} \hat{a}_{i|i-1}, \tag{17}$$

which leads to the estimate of R_i, using the i most recent residuals, given by

$$\hat{R}_i = \frac{1}{i-1} \sum_{k=1}^{i} (z_k - \bar{z})(z_k - \bar{z})^{\mathrm{T}}, \tag{18}$$

where

$$\bar{z} = \frac{1}{i} \sum_{k=1}^{i} z_k \tag{19}$$

is the mean of the innovations up to the ith time instant. Note that the estimation approach (18) requires the white Gaussian property of z_i [27]. Obviously, this property can be obtained if the Kalman filter operates in its optimum state. The optimum filter behavior will be examined by the test described in Section 2.6. Therefore, an adaptive Kalman filter for estimating the state as well as the noise covariance matrices can be called an NAKF.

It should be pointed out that if the model is of non-stationarity and in the absence of any prior information on the initial state, $\hat{a}_0 (= E(a_0))$ can be set to zero and $P_{0|0} (= \mathrm{Cov}(a_0))$ to K times the identity matrix where K is a very large number [28]. This large covariance matrix indicates that little or nothing is known of the initial state. In effect we can also use the first few measurements to estimate the start-up values, so again the few prediction errors and variances should be omitted from the likelihood function. These two approaches are both applicable for the situation of vibration signals. However, this issue should not be a great concern since a recursive filtering scheme based on the periodic feature of vibration signals of rotating machinery will allow one to choose arbitrary initial model parameters.

2.3. Model II based on EKF

Suppose that a linear system with state space description below instead of Eqs. (10) and (11)

$$a_{i+1} = M_i a_i + v_i, \tag{20}$$

$$y_i = C_{i-1}^{\mathrm{T}} a_i + \varepsilon_i \tag{21}$$

is being considered, where, assuming $n = pd^2$ for simplicity, $M_i = \mathrm{diag}[m_1, \ldots, m_n]$ is an unknown constant diagonal coefficient matrix and, as before, $a_i \in \Re^n$, $v_i \in \Re^n$, $\varepsilon_i \in \Re^d$, and v_i and ε_i are uncorrelated white Gaussian noise sequences. Let us assume a vector θ to represent the unknown constant elements m_i for $i = 1, \ldots, n$, namely $\theta = (m_1, \ldots, m_n)^{\mathrm{T}}$. The objective is to estimate a_i and identify θ which must be treated as a random vector such as

$$\theta_{i+1} = \theta_i + \zeta_i, \tag{22}$$

where ζ_i is any zero-mean white Gaussian noise sequence uncorrelated with ε_i and with preassigned positive definite covariances $\mathrm{Cov}(\zeta_i) = W_i$. Otherwise, this assumption would not lead us anywhere since its value cannot be updated. In applications, we may choose $W_i = W$ for all i, where W is an arbitrary constant $n \times n$ matrix. Now the system equations (20) and (21)

together with assumption (22) can be reformulated as the nonlinear model

$$\begin{bmatrix} a_{i+1} \\ \theta_{i+1} \end{bmatrix} = \begin{bmatrix} M_i(\theta_i)a_i \\ \theta_i \end{bmatrix} + \begin{bmatrix} v_i \\ \zeta_i \end{bmatrix}, \tag{23}$$

$$y_i = [C_{i-1}^{\mathrm{T}} \quad 0] \begin{bmatrix} a_i \\ \theta_i \end{bmatrix} + \varepsilon_i, \tag{24}$$

where the parameters are treated as additional states form an augmented state vector. The EKF procedure which takes real-time linear Taylor approximation can then be applied to estimate the state vector which contains θ_i as its components. That is, θ_i is estimated optimally in an adaptive way. Thus, the following EKF algorithm can be derived for $i = 1, \ldots, n$.

Set

$$\begin{bmatrix} \hat{a}_0 \\ \hat{\theta}_0 \end{bmatrix} = \begin{bmatrix} E(a_0) \\ \hat{\theta}_0 \end{bmatrix} \quad \text{and} \quad P_0 = \begin{bmatrix} \mathrm{Cov}(a_0) & 0 \\ 0 & W_0 \end{bmatrix} \tag{25}$$

the recursive prediction equations:

$$P_{i|i-1} = \begin{bmatrix} M_{i-1}(\hat{\theta}_{i-1}) & \dfrac{\partial}{\partial\theta}[M_{i-1}(\hat{\theta}_{i-1})\hat{a}_{i-1}] \\ 0 & I \end{bmatrix} P_{i-1|i-1} \begin{bmatrix} M_{i-1}(\hat{\theta}_{i-1}) & \dfrac{\partial}{\partial\theta}[M_{i-1}(\hat{\theta}_{i-1})\hat{a}_{i-1}] \\ 0 & I \end{bmatrix}^{\mathrm{T}} + \begin{bmatrix} Q_{i-1} & 0 \\ 0 & W_{i-1} \end{bmatrix}, \tag{26}$$

$$\begin{bmatrix} \hat{a}_{i|i-1} \\ \hat{\theta}_{i|i-1} \end{bmatrix} = \begin{bmatrix} M_{i-1}(\hat{\theta}_{i-1})\hat{a}_{i-1} \\ \hat{\theta}_{i-1} \end{bmatrix} \tag{27}$$

and the updating equations:

$$G_i = P_{i|i-1}[C_{i-1}^{\mathrm{T}} \quad 0]^{\mathrm{T}}[[C_{i-1}^{\mathrm{T}} \quad 0]P_{i|i-1}[C_{i-1}^{\mathrm{T}} \quad 0]^{\mathrm{T}} + \hat{R}_i]^{-1}, \tag{28}$$

$$P_{i|i} = [I - G_i[C_{i-1}^{\mathrm{T}} \quad 0]]P_{i|i-1}, \tag{29}$$

$$\begin{bmatrix} \hat{a}_i \\ \hat{\theta}_i \end{bmatrix} = \begin{bmatrix} \hat{a}_{i|i-1} \\ \hat{\theta}_{i|i-1} \end{bmatrix} + G_i(y_i - C_{i-1}^{\mathrm{T}}\hat{a}_{i|i-1}), \tag{30}$$

where Q_i is assumed to be known a priori and $\hat{R}_i = \mathrm{Cov}(\varepsilon_i)$ which can also be adaptively estimated by Eq. (18).

2.4. Model III based on MEKF

An identical state space representation to Eqs. (20) and (21) is proposed here. However, the diagonal coefficient matrix $M_i = \mathrm{diag}[m_1, \ldots, m_n]$ is assumed to be time-varying system parameters rather than a constant vector. The MEKF introduces a very efficient parallel

computational scheme for system parameters identification [29]. The modification is achieved by an improved linearization procedure, which results in that the MEKF algorithm can be applied to real-time system parameter identification even for time-varying stochastic systems. The MEKF algorithm consists of two sub-systems. Algorithm I, which deals with model (31) and (32) below, is a modification of the EKF, where the real-time linear Taylor approximation is not taken at the previous estimate. Instead, in order to improve the performance, it is taken at the optimal estimate of state vector a_i given by a standard Kalman filtering algorithm called Algorithm II, which deals with model (33) and (34) below. Therefore, the system can be reformulated as the nonlinear stochastic system

$$\begin{bmatrix} a_{i+1} \\ \theta_{i+1} \end{bmatrix} = \begin{bmatrix} M_i(\theta_i)a_i \\ \theta_i \end{bmatrix} + \begin{bmatrix} v_i \\ \zeta_i \end{bmatrix}, \tag{31}$$

$$y_i = [C_{i-1}^{\mathrm{T}} \quad 0] \begin{bmatrix} a_i \\ \theta_i \end{bmatrix} + \varepsilon_i, \tag{32}$$

to which the Algorithm I can be applied, and subsystem

$$a_{i+1} = M_i(\tilde{\theta}_i)a_i + v_i, \tag{33}$$

$$y_i = C_{i-1}^{\mathrm{T}} a_i + \varepsilon_i, \tag{34}$$

to which the Algorithm II can be applied. The two algorithms are applied in parallel starting with the same initial estimate, where Algorithm I is used for yielding the estimate $[\tilde{a}_i \ \hat{\theta}_i]^{\mathrm{T}}$ with input $\hat{a}_{i-1}$ obtained from Algorithm II, which is used for yielding the estimate $\hat{a}_i$ with the input $[\tilde{a}_{i-1} \ \hat{\theta}_{i-1}]^{\mathrm{T}}$ obtained from Algorithm I. The two-algorithm procedure listed below is called the parallel algorithm.

Algorithm I. Set

$$\begin{bmatrix} \tilde{a}_0 \\ \tilde{\theta}_0 \end{bmatrix} = \begin{bmatrix} E(a_0) \\ E(\theta_0) \end{bmatrix} \quad \text{and} \quad P_0 = \mathrm{Cov}\left(\begin{bmatrix} a_0 \\ \theta_0 \end{bmatrix} \right). \tag{35}$$

For $i = 1, \ldots, N$, compute the recursive prediction equations:

$$P_{i|i-1} = \left[\frac{\partial}{\partial \begin{bmatrix} a_{i-1} \\ \theta_{i-1} \end{bmatrix}} \begin{bmatrix} M_{i-1}(\tilde{\theta}_{i-1})\hat{a}_{i-1} \\ \tilde{\theta}_{i-1} \end{bmatrix} \right] P_{i-1} \left[\frac{\partial}{\partial \begin{bmatrix} a_{i-1} \\ \theta_{i-1} \end{bmatrix}} \begin{bmatrix} M_{i-1}(\tilde{\theta}_{i-1})\hat{a}_{i-1} \\ \tilde{\theta}_{i-1} \end{bmatrix} \right]^{\mathrm{T}} + Q_{i-1}, \tag{36}$$

$$\begin{bmatrix} \tilde{a}_{i|i-1} \\ \tilde{\theta}_{i|i-1} \end{bmatrix} = \begin{bmatrix} M_{i-1}(\tilde{\theta}_{i-1})\hat{a}_{i-1} \\ \tilde{\theta}_{i-1} \end{bmatrix} \tag{37}$$

and the updating equations:

$$G_i = P_{i|i-1}[C_{i-1}^{\mathrm{T}} \quad 0]^{\mathrm{T}}[[[C_{i-1}^{\mathrm{T}} \quad 0]P_{i|i-1}[C_{i-1}^{\mathrm{T}} \quad 0]^{\mathrm{T}} + \hat{R}_i]]^{-1}, \tag{38}$$

$$P_{i|i} = [I - G_i[C_{i-1}^{\mathrm{T}} \quad 0]]P_{i|i-1}, \tag{39}$$

$$\begin{bmatrix} \tilde{a}_i \\ \tilde{\theta}_i \end{bmatrix} = \begin{bmatrix} \tilde{a}_{i|i-1} \\ \tilde{\theta}_{i|i-1} \end{bmatrix} + G_i(y_i - C_{i-1}^{\mathrm{T}}\tilde{a}_{i|i-1}), \tag{40}$$

where $Q_i = Q_1 = \mathrm{Cov}\left(\begin{bmatrix} v_i \\ \zeta_i \end{bmatrix}\right)$, $\hat{R}_i = \mathrm{Cov}(\varepsilon_i)$, and $\hat{a}_{i-1}$ is obtained by the following Algorithm II.

Algorithm II. Set

$$\hat{a}_0 = E(a_0) \quad \text{and} \quad P_0 = \mathrm{Cov}(a_0). \tag{41}$$

For $i = 1, \ldots, N$, compute the recursive prediction equations:

$$P_{i|i-1} = [M_{i-1}(\tilde{\theta}_{i-1})]P_{i-1|i-1}[M_{i-1}(\tilde{\theta}_{i-1})]^{\mathrm{T}} + Q_{i-1}, \tag{42}$$

$$\hat{a}_{i|i-1} = M_{i-1}(\tilde{\theta}_{i-1})\hat{a}_{i-1} \tag{43}$$

and the updating equations:

$$G_i = P_{i|i-1}C_{i-1}[C_{i-1}^{\mathrm{T}}P_{i|i-1}C_{i-1} + \hat{R}_i]^{-1}, \tag{44}$$

$$P_{i|i} = [I - G_iC_{i-1}^{\mathrm{T}}]P_{i|i-1}, \tag{45}$$

$$\hat{a}_i = \hat{a}_{i|i-1} + G_i(y_i - C_{i-1}^{\mathrm{T}}\hat{a}_{i|i-1}), \tag{46}$$

where $Q_i = Q_2 = \mathrm{Cov}(v_i)$, $\hat{R}_i = \mathrm{Cov}(\varepsilon_i)$, and $\tilde{\theta}_{i-1}$ is obtained from Algorithm I.

However, in order to apply the above MEKF process, we still need an initial estimate $\hat{\theta}_0 := \hat{\theta}_{0|0}$, which should be properly chosen so as to be convergent to the true underlying process and the recursive filtering policy particularly for the periodic signal of rotating machinery will greatly reduce such a requirement and significantly stabilize the filtering estimation. Evidently, the parallelism we have considered here is fundamentally motivated by the need for an evaluation of the Jacobian matrix of the (nonlinear) vector-valued function $M_{i-1}(\theta_{i-1})a_{i-1}$ and the prediction term $[\tilde{a}_{i|i-1} \ \tilde{\theta}_{i|i-1}]^{\mathrm{T}}$ at the optimal position $\hat{a}_{i-1}$ at each time instant. Relevant work addressing efficient systolic implementation for Kalman filter can be found in Ref. [30]. As well, it must be pointed out that all existing EKF algorithms are ad hoc schemes since different linearizations have to be used to derive the results [31,32]. Hence, there is no rigorous theory to guarantee the optimality of the EKF or MEKF algorithm in general [29]. The performance of EKF or MEKF is dependent on the status of the vibration data to be analyzed and also the effectiveness of linearization.

2.5. Model order selection criteria

A major concern with parametric spectral analysis methods is the selection of the order of the AR model (or prediction error filter). If the order is too small, resonances in the data cannot be

resolved and the spectral estimates will be biased and smoothed and some information will be lost. On the other hand, if the model order is too large, spurious peaks (instabilities) occur in the spectral estimates which result in a large error variance and might be misleading and cause a diagnosis error [12,33]. Recently, a number of AR model order selection criteria are available, e.g. the Akaike information criterion (AIC), final prediction error (FPE), coefficient of determination (COD), minimum description length (MDL), etc. Their advantages and disadvantages have been well addressed in many literatures, e.g. Ref. [34].

However, it should be kept in mind that they should be used as a guide only, since extensive experimental results reported in the literature indicate that the above model order selection criteria do not yield definitive results in every case [33]. Further study of AR model order selection is an interesting and important area of investigation, but beyond the focus of this study. Our experiments using on-line gearbox vibration signals showed that analysis based on AIC, FPE and COD together can provide the reliable guidance for selecting AR model order in this study.

The most commonly used criterion AIC, using maximum likelihood principles and given by

$$\mathrm{AIC}(p) = \sum_{i=1}^{N} \ln |\hat{R}_i| + 2pd^2, \tag{47}$$

where p is the model order, N is the number of data points, $\hat{R}_i$ is the covariance matrix of the measurement noise and d is the dimension of the time series under investigation, gives a compromise between model complexity and goodness of fit, and tracks both the decreasing error power and the increasing spectral variance with respect to an increasing model order. Usually, the AIC tends to underestimate the model order with increasing signal-to-noise ratio (SNR).

The FPE criterion for determining the order of AR models is perhaps the most widely known weighted residual error method, and is given by

$$\mathrm{FPE}(p) = \frac{N+p+1}{N-p-1}\hat{R}_p, \tag{48}$$

where $\hat{R}_p$ is the estimate of the covariance matrix of measurement noise under order p and will be replaced by $\hat{R}_N$ obtained at the final time instant using Eq. (18). The FPE gives excellent results for processes which can be adequately described by an AR model. However, it is known to underestimate the model order with increasing SNR.

The COD represents the proportion of variation in the dependent variable that has been explained or accounted for by the regression line and is given by

$$\mathrm{COD}(p) = 1 - \frac{\sigma_e^2}{\sigma_y^2}, \tag{49}$$

where σ_e^2 is the variance of residuals and will be replaced by $\hat{R}_N$ obtained at the final time instant using Eq. (18) and σ_y^2 is the variance of the measurand y_i, $i = 1, 2, \ldots, N$. The value of the COD may vary from zero to one. A COD of zero indicates that none of the variation in the time series of interest is explained by the regression equation; whereas a COD of one indicates that 100% of the variation of the time series of interest has been explained by the regression equation. This method is insensitive to changes in model order as p is increased, but it can give sufficiently reliable results in some applications.

2.6. Test for optimality

The models having been identified and the parameters estimated, diagnostic checks are then applied to the fitted models. The test to determine whether the innovations series z_i is a white sequence, thus indicating optimum filter behavior, is based upon an estimate $\hat{V}_k$ of the autocorrelation sequence V_k. Data is processed in batches of N samples. For a given batch, N_s samples of $\hat{V}_k$ given by

$$\hat{V}_k = \frac{1}{N}\sum_{i=k}^{N-1} z_i z_{i-k}^{\mathrm{T}}, \tag{50}$$

for $k = 0, 1, \ldots, N_s - 1$ are calculated ($N_s < N$). The above is an asymptotically unbiased estimate with mean and (approximate) covariance given by

$$E\{\hat{V}_k\} = \left(1 - \frac{k}{N}\right) V_k, \tag{51}$$

$$\mathrm{Cov}([\hat{V}_k]_{i,j}, [\hat{V}_l]_{m,n}) \cong \frac{1}{N}\sum_{t=-\infty}^{\infty} ([V_t]_{i,m}[V_{t+l-k}]_{j,n} + [V_{t+l}]_{i,n}[V_{t-k}]_{j,m}), \tag{52}$$

where $\mathrm{Cov}(a, b) = E\{[a - E\{a\}] \cdot [b - E\{b\}]\}$, and $[\cdot]_{i,j}$ denotes the element in row i and column j of the matrix. The estimate can also be shown to be consistent because the summation in Eq. (52) is finite and asymptotically normal. It is the Gaussian property of $\hat{V}_k$ that we use to test for whiteness of the innovations sequence z_i. From the 95% confidence limit test for a random variable X with Gaussian distribution,

$$P\{-X_0 \leqslant X \leqslant X_0\} = 0.95 \quad \text{for } X_0 \cong 1.96\sigma_x. \tag{53}$$

This can be applied to, for example, elements in the main diagonal of $\hat{V}_k$. For z_i a white sequence, from Eq. (52) it follows that

$$\mathrm{Cov}([\hat{V}_k]_{i,j}, [\hat{V}_l]_{m,n}) = \frac{1}{N}[V_0]_{i,m}[V_0]_{j,n} \quad \text{for } k = l > 0. \tag{54}$$

The variance of a diagonal element $[\hat{V}_k]_{i,i}$ is then given by

$$\mathrm{var}([\hat{V}_k]_{i,i}) = \frac{1}{N}[\hat{V}_0]_{i,i}^2, \quad k > 0 \tag{55}$$

and substituting Eq. (55) into Eq. (53) results in the limit(s) for the test

$$X_0 = \frac{1.96}{\sqrt{N}}[\hat{V}_0]_{i,i}. \tag{56}$$

The actual test is performed by determining the percentage of values of $[\hat{V}_k]_{i,i}$ for $k = 1, 2, \ldots, N_s - 1$, which fall outside of the range X_0. If this is less than 5%, then z_i is considered white. The test can be done on several of the main diagonal elements of $\hat{V}_k$ and simulations have shown good results using the two extreme elements [35]. The Kolmogorov–Smirnov goodness-of-fit test (K–S test, cf. Appendix A) under the significance level of 5% together with the present optimality check will be applied to test whether the innovations series z_i is a white sequence in this

study. However, the whiteness of innovations sequence is a sufficient condition only under the assumption that transition coefficient matrix M_i is known. When M_i is unknown, an additional condition for filter optimality, namely that the innovations have zero mean, i.e.,

$$E\{z_i\} = 0 \tag{57}$$

for $i = 0, 1, \ldots, N$ must be imposed on the model so as to suffice for optimality testing [35]. This additional condition should be applied to models II and III.

2.7. Parametric spectral analysis in T–F domain

It can be shown that the parametric spectrum of the signal depends on the estimated parameters of time series models. In fact, the relationship is given by

$$P(f) = \frac{P_N(f)}{|H(f)|^2}, \tag{58}$$

that is, the signal power spectrum density (PSD) depends on what can be expressed as the product of $P_N(f)$, PDS of the white noise ($P_N(f) = P_{NO}$), and $H(f)$, frequency response of the linear filter. After estimating the coefficient matrices of the VAR model, an instantaneous estimating of the spectral density function, which in the multivariate case is a matrix valued function of frequency, can be given in terms of the VAR coefficient matrices

$$P_i(f) = [H_i^{-1}(f)]\hat{R}_i[H_i^{-1}(f)]^*, \tag{59}$$

where the asterisk mark denotes the conjugate complex and

$$H_i(f) = I_d - \sum_{k=1}^{p} \hat{A}_k(i)\mathrm{e}^{-\mathrm{j}2\pi f T_s k}. \tag{60}$$

The PSD function (59) is adaptive because each coefficient matrix $\hat{A}_k(i)$ is considered to be subject to a stochastic process in order to fit the time-varying spectral characteristics of the vibration signals. In this manner, the model can, in principle, follow rapidly varying spectra because it has an inherently non-stationary structure. Accordingly, the PSD function, expressing the spectrum by using the time-varying VAR parameters, becomes a function of two variables, time i and frequency f in the same way as for all T–F distributions [2].

3. Model evaluation using simulated signals

Two simulated signals are tested in this section. The results of optimality tests are generalized in Table 1. To provide a convenient manner, the construction of simulated signal and the resulting T–F maps will be presented first in each case. It is then followed by the description of model order selection procedure and the optimality tests for white Gaussian assumption and zero mean of innovations.

Table 1
Generalization of model evaluation using simulated signals[a]

Signal	Model	Order p	K–S test	Pr (%)	Zero mean test of innovations in four intervals			
					1st	2nd	3rd	4th
Exponentially decaying signal	I		1	13.26			—	
	II	100	1	12.46	1.4319e-02	2.7542e-03	1.6372e-03	3.0795e-02
	III		1	12.41	1.4528e-02	2.7996e-03	1.6640e-03	3.0791e-03
Simulated signal of gear fault	I		1	0.88			—	
	II	25	1	0.78	1.0159e-01	1.3704e-02	9.3447e-03	6.7116e-03
	III		1	0.78	1.3168e-01	1.9204e-02	1.1662e-02	8.4520e-03

[a]Pr (%) denotes the outlier percentage. The intervals of the exponentially decaying signal consist of $[0, 0.5]$, $[0.5, 1.0]$, $[1.0, 1.5]$ and $[1.5, 2.0]$ s. And, the intervals of the simulated signal of gear fault consist of $[0°, 90°]$, $[90°, 180°]$, $[180°, 270°]$ and $[270°, 360°]$. K–S test denotes the Kolmogorov–Smirnov goodness-of-fit test, where 0 denotes that the null hypothesis that innovations can be adjusted to Gaussian distribution is accepted under 5% significance level and 1 denotes rejection of null hypothesis.

3.1. Exponentially decaying signal

The exponentially decaying signal with abrupt changes in frequency composition considered in this section is artificially obtained and composed of three distinct frequency components which overlap with each other and have different decay factors. However, in order to examine the adaptation and T–F resolution capabilities of the proposed models, abrupt changes in spectral contents are introduced into this exponentially decaying signal. This artificial signal is expressed by

$$S(n) = \begin{cases} F_1, & 0 \leqslant t < 0.5\,\text{s}, \\ F_1 + F_2, & 0.5\,\text{s} \leqslant t < 1.5\,\text{s}, \\ F_1 + F_2 + F_3, & 1.5\,\text{s} \leqslant t < 2.0\,\text{s}, \end{cases}$$

where

$$F_1 = \text{e}^{-n/2}\cos(2\pi f_1 n) = \text{e}^{-n/2}\cos(2\pi 15 n),$$

$$F_2 = \text{e}^{-n}\cos(2\pi f_2 n) = \text{e}^{-n}\cos(2\pi 35 n),$$

$$F_3 = \text{e}^{-n/4}\cos(2\pi f_3 n) = \text{e}^{-n/4}\cos(2\pi 25 n)$$

and $f_1 = 15\,\text{Hz}$, $f_2 = 35\,\text{Hz}$ and $f_3 = 25\,\text{Hz}$ are the three distinct frequency components, $\text{e}^{-n/2}$, e^{-n} and $\text{e}^{-n/4}$ the three decay factors, $n = [0 : 0.001 : (2 - 0.001)]$ and sampling frequency is then 1 kHz (Nyquist frequency is 500 Hz). Therefore, abrupt changes in the major frequency composition take place at time instants 0.5 and 1.5 s, respectively.

The signal and the resulting T–F maps are presented in Fig. 1, where the abrupt changes of frequency composition can be observed in either the time-domain signal as shown in Fig. 1(a) or the T–F maps as shown in Figs. 1(b)–(d). As can be seen, very consistent T–F representations are

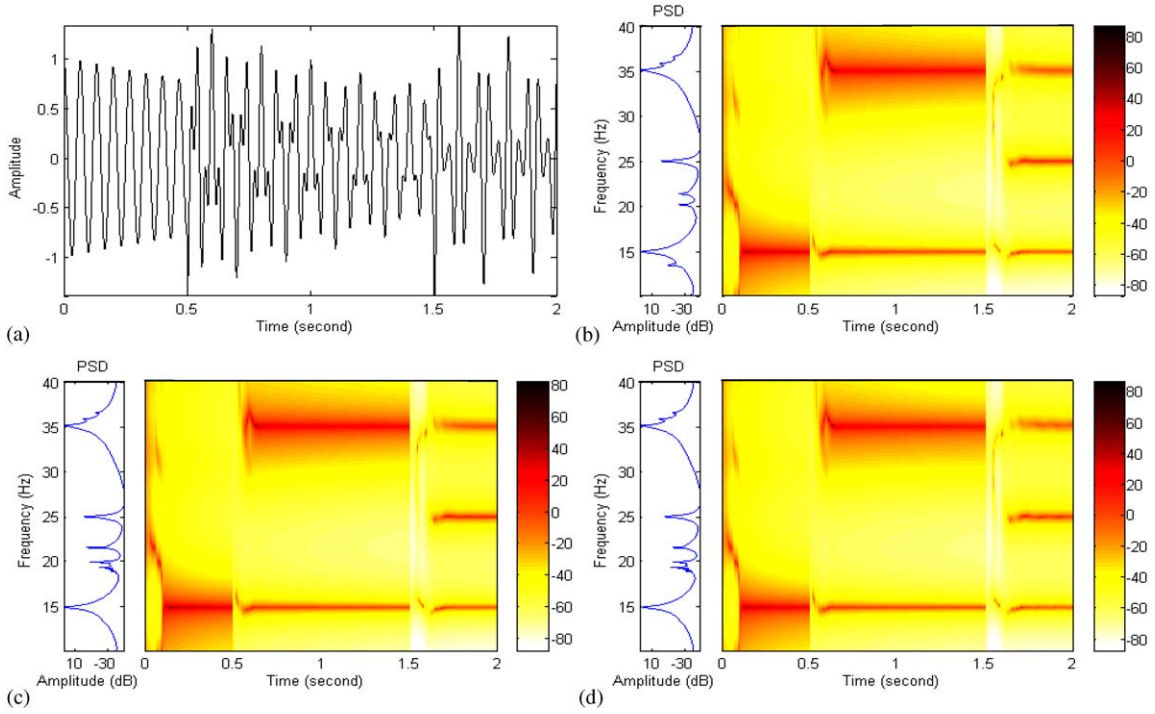

Fig. 1. Exponentially decaying signal and T–F maps, where the sampling frequency f_s is 1 kHz, time resolution is 1 mini-second and frequency resolution is 0.1 Hz. (a) Exponentially decaying signal, (b) T–F map of model I, (c) T–F map of model II, (d) T–F map of model III.

presented by all models which correctly capture the actual frequency components after a certain period of unstable adaptation at the beginning of each of [0, 0.5], [0.5, 1.5] and [1.5, 2.0] s intervals, respectively.

The optimum order indicated by the AIC and FPE is always $p = 205$ for each model as shown in Fig. 2, whereas the COD indicates that $p = 250$ is the optimum one for each model. However, our experiments revealed that model with the order of 205 or higher present a longer unstable adaptation procedure and relatively stronger spurious components especially within the [1.5, 2.0] s interval in comparison with lower order values, whereas model with the orders lower than 80 cannot correctly recover the frequency components within [1.5, 2.0] s and thus results in a low-frequency resolution. Therefore, order $p = 100$ is finally determined for each model based on extensive tests. Table 1 illustrates that the outlier percentages of all models, 13.26%, 12.46% and 12.41%, are larger than 5% and thus are not strictly consistent with the whiteness assumption of innovations sequence.

In addition, all models fail the K–S test for Gaussian assumption at 5% significant level as indicated by either the normal probability plots as shown in Figs. 3(a)–(c), respectively, where the plot of innovations of each model is not linear, or the K–S statistics, 13.6514, 13.5996 and 13.4711, which are larger than the K–S critical value 0.8950. Figs. 3(d) and (e) show that the means of innovations of models II and III tend to zero at a very fast rate at the initial filtering stage and strictly progress forward at levels close to zero over the time horizon.

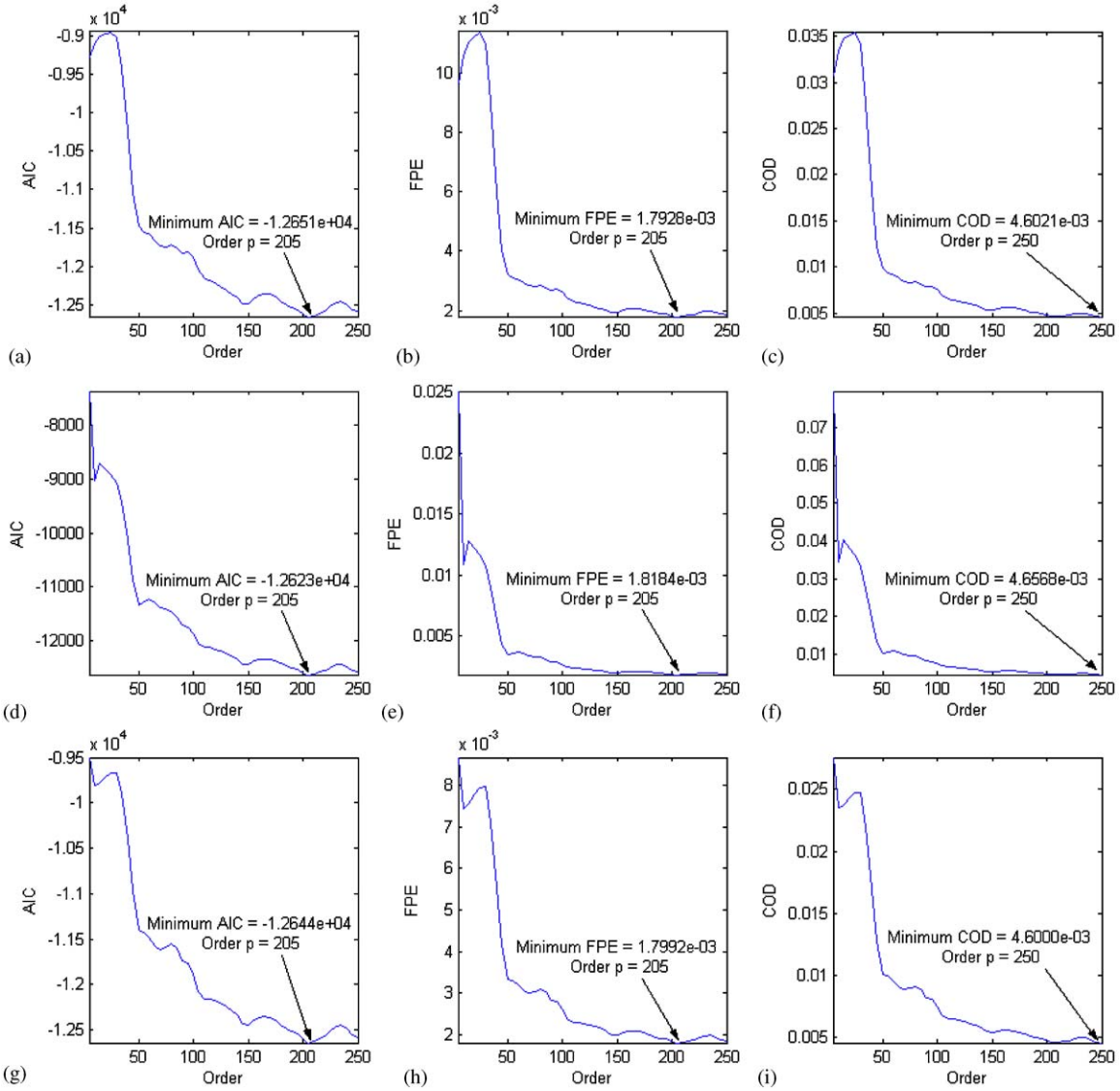

Fig. 2. AIC, FPE and COD order selection criteria for the exponentially decaying signal, where the optimum order indicated by AIC and FPE is $p = 205$ for each model, respectively, and the optimum order indicated by COD is $p = 250$ for each model, respectively. (a) AIC for model I, (b) FPE for model I, (c) COD for model I, (d) AIC for model II, (e) FPE for model II, (f) COD for model II, (g) AIC for model III, (h) FPE for model III, (i) COD for model III.

However, inconsistency with the white Gaussian assumption does not harm the quality of the resulting T–F representations as shown in Figs. 1(b)–(d). Further experiments were performed and showed that model with outlier percentage less than 5% may not necessarily present acceptable T–F representation. As well, models II and III are found to be relatively more sensitive to the initialization of filters, i.e. the selection of initial filter parameters, than model I. Therefore,

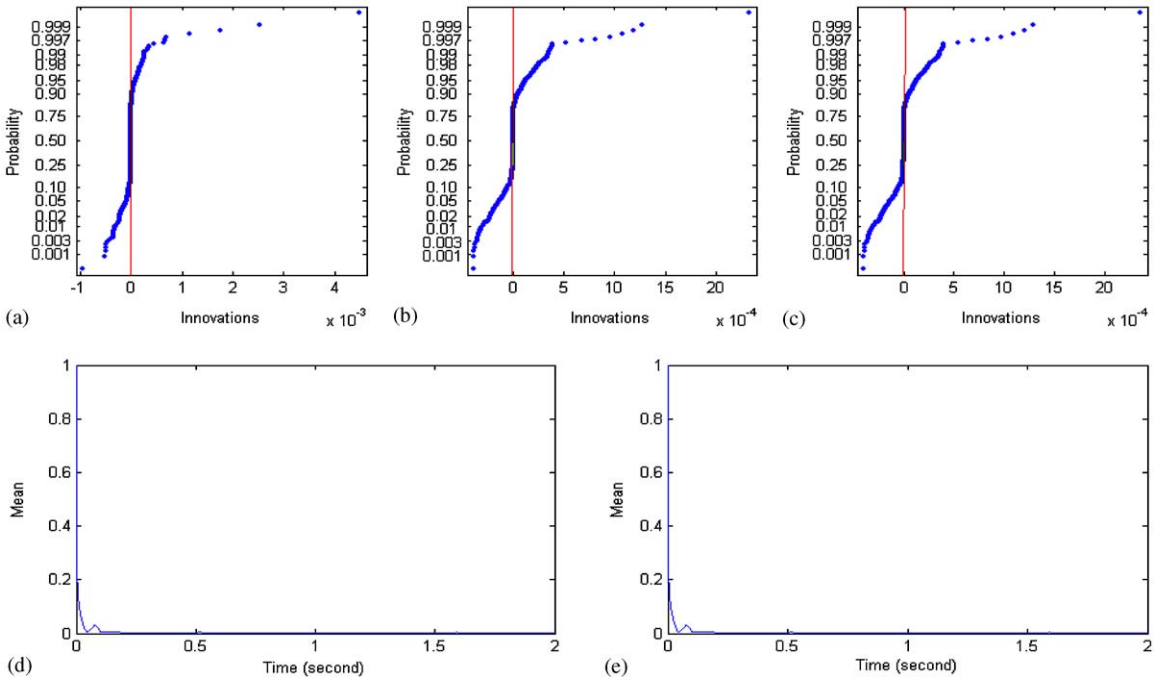

Fig. 3. Normal probability plots and zero mean optimality tests of innovations for the exponentially decaying signal. The K–S statistics of models I, II and III are 13.6514, 13.5996 and 13.4711, respectively. The K–S critical value is 0.8950. (a) Normal probability plot of model I, (b) normal probability plot of model II, (c) normal probability plot of model III, (d) zero mean test of innovations for model II, (e) zero mean test of innovations for model III.

appropriate initialization of filters should be emphasized here for models II and III so as to obtain desirable T–F representations.

3.2. Simulated signal of gear fault

A simulated signal similar to the one used in Ref. [36] is constructed in this section to evaluate the performance of the proposed three models. This example is intended to numerically generate simulated signal which possesses some major features of gear vibration signal under faulty state, say non-adjacent tooth cracks. This simulated signal takes the form of a time-domain average containing 1024 sampling points and has five components, a pure sine wave, an amplitude-modulated sine wave, a frequency-modulated sine wave and two Gaussian impulses, as given by

$$\begin{aligned} S(n) = {} & A_1 \sin(2\pi f_1 n) + (1 + A_2 \cos(2\pi f_2 n)) A_3 \sin(2\pi f_3 n) \\ & + A_4 \sin(2\pi f_4 n + A_5 \sin(2\pi f_5 n)) \\ & + A_6 \exp\left(-2k_1^2\left(n - \frac{n_1}{N}\right)^2\right) + A_7 \exp\left(-2k_2^2\left(n - \frac{n_2}{N}\right)^2\right), \end{aligned}$$

where $f_1 = 80$ Hz, $f_2 = f_5 = 2$ Hz, $f_3 = 120$ Hz, $f_4 = 160$ Hz, $A_1 = A_4 = 1.5$, $A_2 = 0.75$, $A_3 = 1$, $A_5 = 2$, $A_6 = A_7 = 3$, $k_1 = k_2 = 256$, $n_1 = 256$, $n_2 = 512$, $N = 1024$, $n = [0 : T_s : (1 - T_s)]$, $T_s = 0.9766$ mini-second and sampling frequency is then 1.024 kHz.

Fig. 4 shows the signal in time-domain and the resulting T–F maps. It is clear that the resulting T–F maps produced by the three models as shown in Figs. 4(b)–(d) significantly outperform that produced by STFT in Ref. [36] in either capturing correct components or providing highly delicate T–F resolution. The pure sine wave, amplitude-modulated sine wave and frequency-modulated sine wave are successfully identified by each model. Further inspection reveals that models II and III show some advantages over model I with respect to the accuracy of how the two Gaussian impulses, which are supposed to appear as two small patches centered at 90° and 180°, are locally emphasized. In particular, after the identification of the second Gaussian impulse at 180°, model I presents very slightly spurious power distribution close to the zero frequency boundary within the angular interval of [180°, 360°]. On the other hand, similar to the previous case using exponentially decaying signal, it is found that both models II and III have relatively higher requirement for an appropriate initialization of filter in order to obtain an acceptable T–F representation. Therefore, an appropriate model initialization is still necessary for models II and III.

The results of model order selection are presented in Fig. 5. All criteria indicate that order $p = 25$ is the optimum one for each model. Therefore, order $p = 25$ is assigned to each model, respectively. The K–S test shows that the innovations sequence generated by each model is not

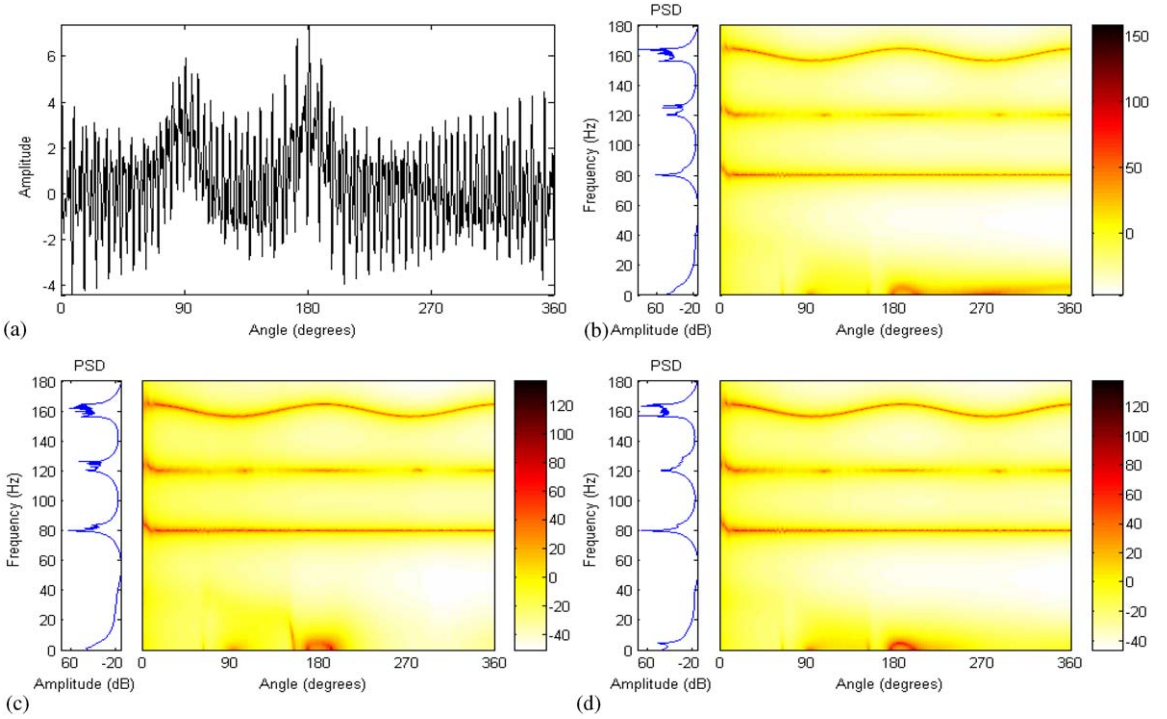

Fig. 4. Simulated signal of gear fault and T–F maps, where the sampling frequency f_s is 1.024 kHz, frequency resolution is 0.1 Hz and angular resolution is 0.3516° for all T–F maps in this figure. (a) Simulated signal of gear fault, (b) T–F map of model I, (c) T–F map of model II, (d) T–F map of model III.

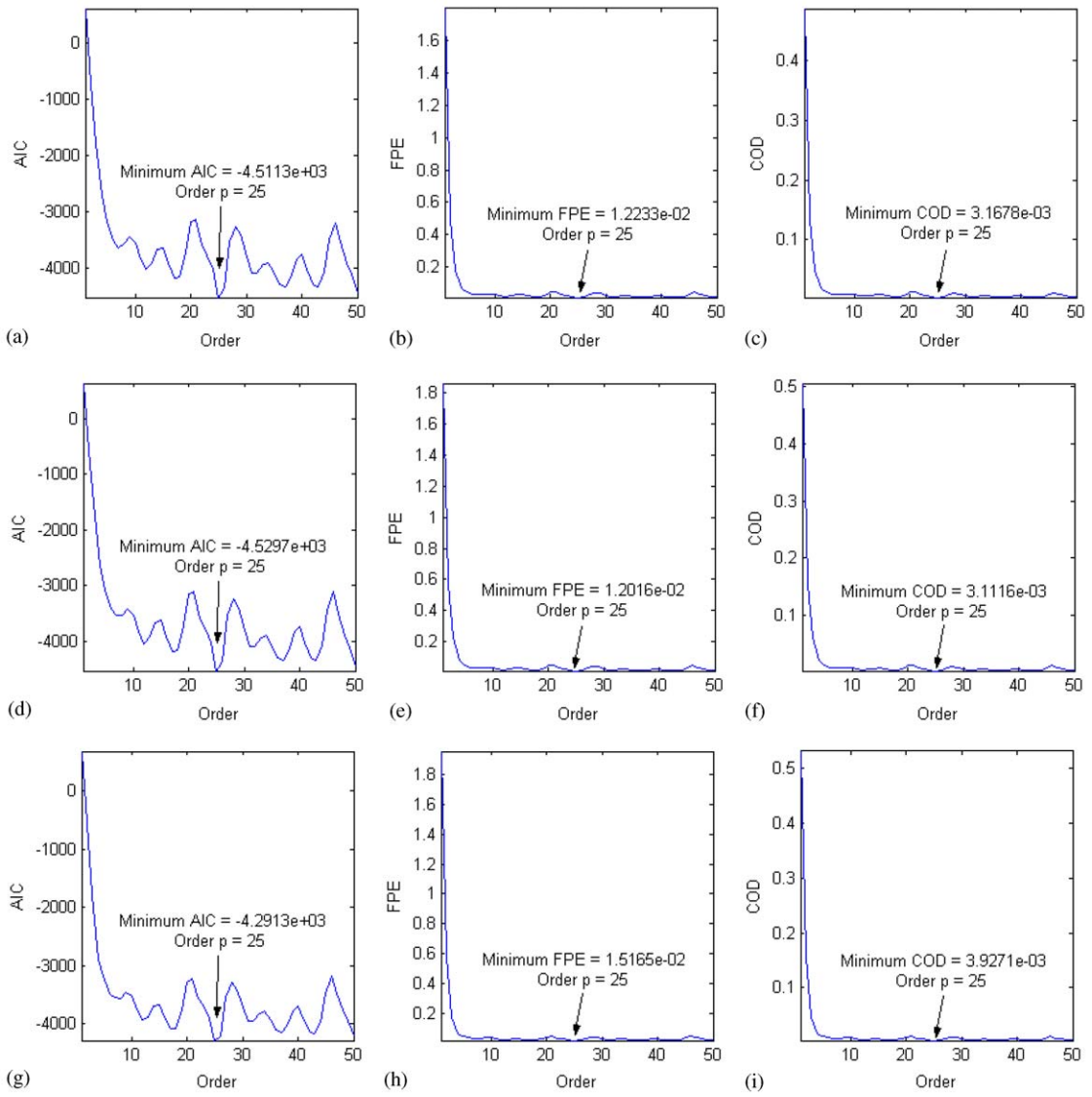

Fig. 5. AIC, FPE and COD order selection criteria for the simulated signal of gear fault, where the optimum order is $p = 25$ as indicated by all criteria for each model, respectively. (a) AIC for model I, (b) FPE for model I, (c) COD for model I, (d) AIC for model II, (e) FPE for model II, (f) COD for model II, (g) AIC for model III, (h) FPE for model III, (i) COD for model III.

Gaussian as the K–S statistics, 11.3499, 11.5429 and 11.4466, are larger than the critical value 0.8950 as shown in Table 1.

In addition, the normal plots of innovations in Figs. 6(a)–(c) are not linear. However, the whiteness assumption of innovations is valid for each model under such an order as indicated by their outlier percentages, 0.88%, 0.78% and 0.78% which are less than 5%. Also, the zero mean tests of innovations in Figs. 6(d) and (e) show that the mean of innovations generated by each

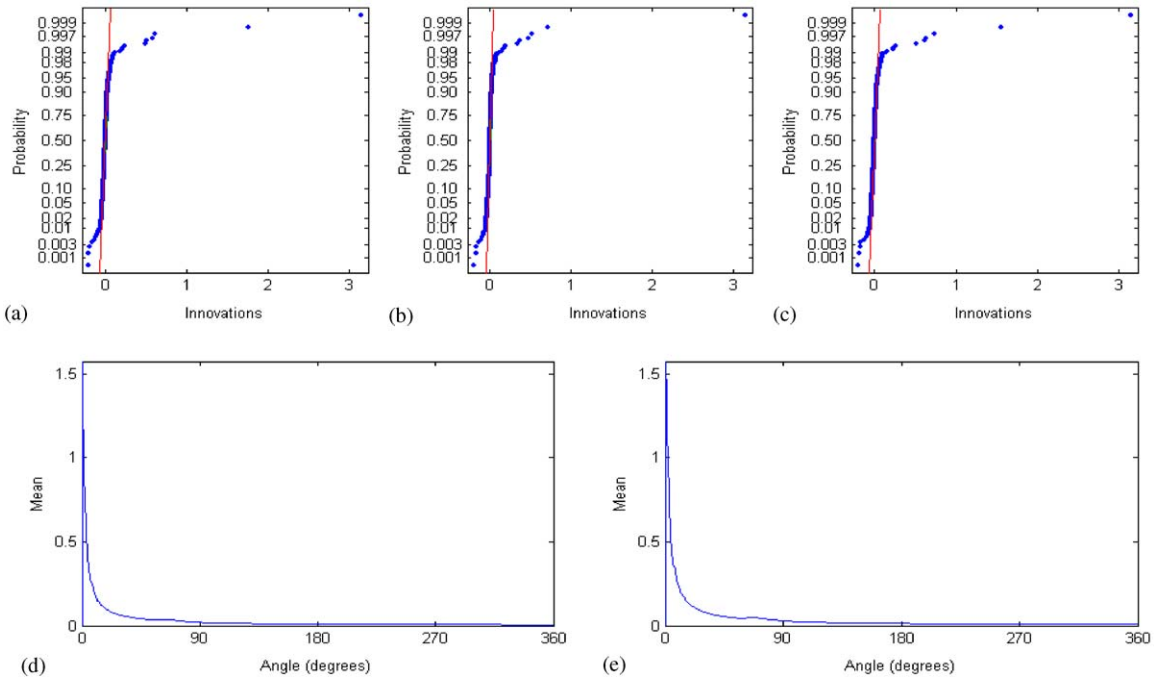

Fig. 6. Normal probability plots and zero mean optimality tests of innovations for the simulated signal of gear fault. The K–S statistics of models I, II and III are 11.3499, 11.5429 and 11.4466, respectively. The K–S critical value is 0.8950. (a) Normal probability plot of model I, (b) normal probability plot of model II, (c) normal probability plot of model III, (d) zero mean test of innovations for model II, (e) zero mean test of innovations for model III.

model converges to level near zero quickly. Therefore, the optimum behavior of filter is obtained by each model in this case.

It is noteworthy that alternative non-parametric T–F transforms such as Wigner–Ville or wavelets could be attempted and compared with the proposed parametric models. However, a relatively acceptable identification of major frequency components is usually accompanied by the presence of spurious components for the Wigner–Ville transform, while for wavelet transform a precise identification in, say, frequency domain is simultaneously at the expenditure of time domain resolution. Thus, under the premise of a constant number of sampling points, the proposed three models are more favorable.

4. Conclusions

This paper proposes three new adaptive parametric models transformed from time-varying vector-autoregressive model with their parameters estimated by means of noise-adaptive Kalman filter, extended Kalman filter and modified extended Kalman filter, respectively, on the basis of different assumptions. The performance analysis of the proposed adaptive parametric models is demonstrated using numerically generated non-stationary test signals. The results indicate that the proposed models possess appealing advantages in processing non-stationary signals and thus

are able to provide reliable T–F domain information for condition monitoring. In particular, the EKF- and MEKF-based models demonstrate appealing advantages in localizing the gear fault-induced features, while the NAKF-based model demonstrates strong adaptability to a wide variety of model initialization. Further analysis using actual on-line vibration signals of gearbox will be presented in Part 2 of this study.

Acknowledgements

We are most grateful to the Applied Research Laboratory at the Pennsylvania State University and the Department of the Navy, Office of the Chief of Naval Research (ONR) for providing the data used to develop this work. We thank Bob Luby at PricewaterhouseCoopers for his help to obtain the data. The authors also wish to thank the Natural Sciences and Engineering Research Council of Canada (NSERC), Material and Manufacturing Ontario of Canada (MMO) and the CBM Consortium companies for their financial supports. Finally, we appreciate very much the valuable comments from the anonymous referees who greatly helped to improve this work.

Appendix A. Kolmogorov–Smirnov goodness-of-fit test

The Kolmogorov–Smirnov (K–S) goodness-of-fit test compares an empirical distribution function with the distribution function of the hypothesized distribution [37].

To define the K–S statistic, we must first define an empirical distribution function. For the K–S test, we define an empirical distribution function $F_n(x)$ from our data $X_1, X_2, \ldots, X_n$ as

$$F_n(x) = \frac{(\text{number of } X_i\text{'s} \leqslant x)}{n} \tag{A.1}$$

for all real numbers x. Thus, $F_n(x)$ is a (right-continuous) step function such that $F_n(x_{(i)}) = i/n$ for $i = 1, 2, \ldots, n$. If $\hat{F}(x)$ is the fitted distribution function, a natural assessment of goodness of fit is some kind of measure of the closeness between the functions F_n and $\hat{F}$. The K–S test statistic D_n is simply the largest (vertical) distance between $F_n(x)$ and $\hat{F}(x)$ for all values of x and is defined formally by

$$D_n = \sup_x \{|F_n(x) - \hat{F}(x)|\}. \tag{A.2}$$

(The 'sup' of a set of numbers A is the smallest value that is greater than or equal to all members of A. The 'sup' is used here instead of the more familiar 'max' since, in some cases, the maximum may not exist. For example, if $A = (0, 1)$, there is no maximum but the 'sup' is 1.) D_n can be computed by calculating

$$D_n^+ = \max_{1 \leqslant i \leqslant n} \left\{ \frac{i}{n} - \hat{F}(x_{(i)}) \right\}, \quad D_n^- = \max_{1 \leqslant i \leqslant n} \left\{ \hat{F}(x_{(i)}) - \frac{i-1}{n} \right\} \tag{A.3}$$

and finally letting

$$D_n = \max\{D_n^+, D_n^-\}. \tag{A.4}$$

Suppose that the hypothesized distribution is $N(\mu, \sigma^2)$ with both μ and σ^2 unknown. We can estimate μ and σ^2 by $\bar{X}(n)$ and $S^2(n)$, respectively, and define the distribution function to be that of the $N(\bar{X}(n), S^2(n))$ distribution; i.e., let $\hat{F}(x) = \Phi\{[x - \bar{X}(n)]/\sqrt{S^2(n)}\}$, where Φ is the distribution function of the standard normal distribution. Using this $\hat{F}$ (which has estimated parameters), D_n is computed in the same way, but different critical points must be used. An accurate approximation, which obviates the need for large tables, is provided by Law and Kelton [37]; namely, we reject H_o if

$$\left(\sqrt{n} - 0.01 + \frac{0.85}{\sqrt{n}}\right) D_n > c'_{1-\alpha}, \tag{A.5}$$

where α is always 0.05 in this study and thus $c'_{1-\alpha}$ takes the value of 0.8950 as shown in Table 6.14 on page 390 by Law and Kelton [37].

References

[1] B. Samimy, G. Rizzoni, Mechanical signature analysis using time–frequency signal processing: application to internal combustion engine knock detection, *Proceedings of the IEEE* 84 (1996) 1330–1343.

[2] S. Conforto, T. D'Alessio, Spectral analysis for non-stationary signals from mechanical measurements: a parametric approach, *Mechanical Systems and Signal Processing* 13 (1999) 395–411.

[3] W.J. Wang, P.D. McFadden, Application of wavelets to gearbox vibration signals for fault detection, *Journal of Sound and Vibration* 192 (1996) 927–939.

[4] S.K. Tang, On the time–frequency analysis of signals that decay exponentially with time, *Journal of Sound and Vibration* 234 (2000) 241–258.

[5] N. Baydar, A. Ball, A comparative study of acoustic and vibration signals in detection of gear failures using Wigner–Ville distribution, *Mechanical Systems and Signal Processing* 15 (2001) 1091–1107.

[6] P.W. Tse, Y.H. Peng, R. Yam, Wavelet analysis and envelope detection for rolling element bearing fault diagnosis—their effectiveness and flexibilities, *Journal of Vibration and Acoustics* 123 (2001) 303–310.

[7] J.H. Lee, J. Kim, H.J. Kim, Development of enhanced Wigner–Ville distribution function, *Mechanical Systems and Signal Processing* 15 (2001) 367–398.

[8] S.U. Lee, D. Robb, C. Besant, The directional Choi–Williams distribution for the analysis of rotor-vibration signals, *Mechanical Systems and Signal Processing* 15 (2001) 789–811.

[9] N.G. Nikolaou, I.A. Antoniadis, Rolling element bearing fault diagnosis using wavelet packets, *NDT&E International* 35 (2002) 197–205.

[10] W.Q. Wang, F. Ismail, M.F. Golnaraghi, Assessment of gear damage monitoring techniques using vibration measurements, *Mechanical Systems and Signal Processing* 15 (2001) 905–922.

[11] D.C. Baillie, J. Mathew, A comparison of autoregressive modeling techniques for fault diagnosis of rolling element bearings, *Mechanical Systems and Signal Processing* 10 (1996) 1–17.

[12] J.P. Dron, L. Rasolofondraibe, F. Bolaers, A. Pavan, High-resolution methods in vibratory analysis: application to ball bearing monitoring and production machine, *International Journal of Solids and Structures* 38 (2001) 4293–4313.

[13] W.Y. Wang, A.K. Wong, Autoregressive model-based gear fault diagnosis, *Journal of Vibration and Acoustics* 124 (2002) 172–179.

[14] A.G. Gray, P.J. Thomson, On a family of finite moving-average trend filters for the ends of series, *Journal of Forecasting* 21 (2002) 125–149.

[15] A.M. Bianco, B.M. Garcia, E.J. Martinez, V.J. Yohai, Outlier detection in regression models with ARIMA errors using robust estimates, *Journal of Forecasting* 20 (2001) 565–579.

[16] S. Makridakis, M. Hibon, ARMA models and the Box–Jenkins methodology, *Journal of Forecasting* 16 (1997) 147–163.

[17] V.R. Prybutok, J. Yi, D. Mitchell, Comparison of neural network models with ARIMA and regression models for prediction of Houston's daily maximum ozone concentrations, *European Journal of Operational Research* 122 (2000) 31–40.
[18] K. Kumar, V.K. Jain, Autoregressive integrated moving averages (ARIMA) modeling of a traffic noise time series, *Applied Acoustics* 58 (1999) 283–294.
[19] W.H. Ip, R. Fung, K.W. Keung, An investigation of stochastic analysis of flexible manufacturing systems simulation, *International Journal of Advanced Manufacturing Technology* 15 (1999) 244–250.
[20] P.S. Naidu, *Modern Spectrum Analysis of Time Series*, CRC Press, Boca Raton, 1996.
[21] J.M. Girault, F. Ossant, A. Ouahabi, D. Kouame, F. Patat, Time-varying autoregressive spectral estimation for ultrasound attenuation in tissue characterization, *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control* 45 (1998) 650–659.
[22] M. Arnold, W.H.R. Miltner, H. Witte, R. Bauer, C. Braun, Adaptive AR modeling of nonstationary time series by means of Kalman filtering, *IEEE Transactions on Biomedical Engineering* 45 (1998) 553–562.
[23] B. Panuthat, T. Funada, N. Kanedera, Speech analysis/synthesis/conversion by using sequential processing, *IEEE International Conference on Acoustics, Speech, and Signal Processing* 1 (1999) 15–19.
[24] K.M. Malladi, R.V.R. Kumar, K.R. Veerabhadra, Gauss–Markov model formulation for the estimation of time-varying signals and systems, *IEEE Region 10 International Conference on Global Connectivity in Energy*, *Computer*, *Communication and Control*, vol. 1, 1999, pp. 166–169.
[25] Y.M. Zhan, V. Makis, A.K.S. Jardine, An adaptive model of rotating machinery subject to vibration monitoring, *IIE Annual Research Conference Proceedings*, Orlando, FL, 2002, Paper No. 2066.
[26] Y.M. Zhan, V. Makis, A.K.S. Jardine, Adaptive model of rotating machinery subject to random deterioration for vibration monitoring, *Journal of Quality in Maintenance Engineering* 9 (2003) 351–375.
[27] A. Moghaddamjoo, R.L. Kirlin, Robust adaptive Kalman filtering with unknown inputs, *IEEE Transactions on Acoustics, Speech, and Signal Processing* 37 (1989) 1166–1175.
[28] G.J. Janacek, L. Swift, *Time Series: Forecasting, Simulation, Applications*, Ellis Horwood, New York, 1993.
[29] C.K. Chui, G. Chen, *Kalman Filtering: With Real-Time Applications*, Springer, New York, 1999.
[30] S.Y. Kung, J.N. Hwang, Systolic array designs for Kalman filtering, *IEEE Transactions on Signal Processing* 39 (1991) 171–182.
[31] N.U. Ahmed, S.M. Radaideh, Modified extended Kalman filtering, *IEEE Transactions on Automatic Control* 39 (1994) 1322–1326.
[32] D.S. Wall, F.M.F. Gaston, Modified extended Kalman filtering, *IEEE 13th International Conference on Digital Signal Processing Proceedings*, vol. 2, 1997, pp. 703–706.
[33] T.M. Romberg, J.L. Black, T.J. Ledwidge, *Signal Processing for Industrial Diagnostics*, Wiley, England, Chichester, 1996.
[34] H. Lütkepohl, *Introduction to Multiple Time Series Analysis*, Springer, Berlin, 1991.
[35] G. Noriega, S. Pasupathy, Adaptive estimation of noise covariance matrices in real-time preprocessing of geophysical data, *IEEE Transactions on Geoscience and Remote Sensing* 35 (1997) 1146–1159.
[36] W.J. Wang, P.D. McFadden, Early detection of gear failure by vibration analysis—I, calculation of the time–frequency distribution, *Mechanical Systems and Signal Processing* 7 (1993) 193–203.
[37] A.M. Law, W.D. Kelton, *Simulation Modeling and Analysis*, McGraw-Hill, New York, 1991, pp. 387–390.