Journal of Sound and Vibration 325 (2009) 629–648

Contents lists available at ScienceDirect

Journal of Sound and Vibration

journal homepage: www.elsevier.com/locate/jsvi

Gearbox vibration monitoring using extended Kalman filters and hypothesis tests

Yimin Shao [a,*], Chris K. Mechefske [b]

[a] The State Key Laboratory of Mechanical Transmission, Chongqing University, Chongqing 400044, PR China
[b] Department of Mechanical and Materials Engineering, Queen's University, Canada

ARTICLE INFO

Article history:
Received 31 August 2008
Received in revised form
18 March 2009
Accepted 21 March 2009
Handling Editor: L.G. Tham
Available online 5 May 2009

ABSTRACT

The efficiency of many maintenance programs is heavily dependent on the detection accuracy of the condition monitoring system. Condition indicators that are sensitive to environmental or operational variables of no interest will inevitably reflect irrelevant fluctuations and thus mislead the subsequent analysis. In consideration of this phenomenon, a fully automatic and robust vibration monitoring system for gearboxes is proposed in this study. The primary objective here is on how to exclude the effects of variable load conditions. The proposed technique features a number of appealing advantages, which include extended Kalman filter-based time-varying autoregressive modeling, automatic autoregressive model order selection with the aid of a non-paired two-sample Satterthwaite's t'-test, a highly effective and robust condition indicator (the means of one-sample Kolmogorov–Smirnov goodness-of-fit test), and an automatic alert generating mechanism for incipient gear faults with the aid of a Wilcoxon rank-sum test. Two sets of entire lifetime gearbox vibration monitoring data with distinct variable load conditions were used for experimental validation. The proposed condition indicator was compared with other well-known and/or recently proposed condition indicators. The results demonstrate excellent performance of the proposed technique in four aspects: the effectiveness of identifying the optimum model order, a minimum number of false alerts, constant behavior under variable load conditions, and to some extent an early alert for incipient gear faults. Furthermore, the proposed condition indicator can be directly employed by condition-based maintenance programs as a condition covariate for operational maintenance decision analysis. It provides a quantitative and more efficient means for exchanging condition information with maintenance programs in comparison with the widely used non-parametric time–frequency techniques such as wavelets, which rely on visual inspection.

1. Introduction

Condition-based maintenance (CBM) programs are now widely used in industry in order to reduce unscheduled downtime and production losses. The condition monitoring system is required to provide condition information that is accurate. However, most condition monitoring techniques are implemented under a certain specified operating conditions. Therefore, consideration of possible operating and/or environmental fluctuations into a condition monitoring system is

* Corresponding author. Tel.: +86 23 65112520; fax: +86 23 65112522.
E-mail addresses: ymshao@cqu.edu.cn (Y. Shao), chrism@me.queensu.ca (C.K. Mechefske).

0022-460X/$ - see front matter © 2009 Published by Elsevier Ltd.
doi:10.1016/j.jsv.2009.03.029

usually neglected in conventional condition monitoring techniques. The currently intensively exploited non-parametric time–frequency techniques (wavelets, short time Fourier transforms) are subject to operational and environmental variables and represent the signal dynamics in time–frequency domain caused by both the object of interest and other irrelevant external sources. In this regard, this study will focus on the deterioration detection of gears under variable load conditions and develop a condition monitoring system characterized by a number of highly desirable properties for practical condition monitoring.

In contrast to advanced non-parametric time–frequency techniques such as wavelets [1–3] and Hilbert–Huang transforms [4–6], which have incurred substantial research efforts in recent years, the potential of parametric-based approaches in the condition monitoring of rotating machinery has not been paid sufficient attention. A few recent studies investigated the performance of time-series models (autoregressive (AR) and mixed autoregressive and moving-average (ARMA)) in the condition monitoring of rotating machinery [7,8]. However, there still remain some critical unresolved issues. Wang and Wong [7] made use of a stationary AR model to fit healthy time-synchronous average (TSA) signals and then calculated the kurtosis statistic of model estimates when the fitted AR model is applied to TSA signals collected under faulty gear conditions. Time synchronous averaging is simply the average of several dynamics signals each recorded for a fixed time period. The time period over which the averaging takes place is typically matched to one gear rotation or one gear cycle. One gear cycle is the time required for the mating pair of gears to return to their initial mating position where the exact same teeth are in contact. The results were claimed to be capable of generating an early warning for incipient gear faults. However, the proposed technique is unable to deal with variable load conditions. The fitted healthy TSA signals and the filtered faulty TSA signals must be of identical load condition. In this connection, it may not be appropriate for the vibration monitoring of actual gearing systems which are subject to variable loads. The work conducted by Wang and Mechefske [8] is devoted to the performance analysis of both time-varying AR and ARMA models when vibration signals with a few well-known features are processed. Further efforts are necessary if a CBM program intends to utilize the resulting time–frequency information provided by the analyzed AR and ARMA models, since it requires a sharp visual discrimination and no quantitative condition indicators are provided.

In this study, we will focus on a time-varying AR modeling technique which assumes that the coefficients of the AR model are time-varying and updated with the incoming vibration measurements at every instant of time. A significant amount of previous research has made use of time-varying AR models in a variety of research areas [9–11]. The work done here is based on a Kalman filter directed AR model which estimates the coefficients of a time-varying AR model adaptively with the progress of time. The original time-varying AR model is expressed by

$$z_t = \sum_{i=1}^{p} \phi_i(t) z_{t-i} + \alpha_t \tag{1}$$

where t is the time index, z_t is the de-meaned series of original vibration measurement series, $\phi_i(t)$ are the time-varying coefficients, α_t is the additive white noise, and p is the model order. This expression can be transformed to a state-space representation whose state vector s_t, composed of the unknown coefficients $\phi_i(t)$, is usually assumed to follow a random walk process [9,10]:

$$s_t = s_{t-1} + \gamma_t \tag{2}$$

Here γ_t stands for the random shock of a white noise series with zero mean and covariance matrix σ_γ^2 corrupting the evolution of s_t. However, the impulsive nature of gear vibration signals (caused by imperfections and irregularities of gear tooth profile and/or instantaneous fluctuations of any regular external excitation over each shaft revolution) leads us to consider a more general assumption. This allows the state vector s_t to be subjected to an unknown constant coefficient matrix M_t as described by

$$s_t = M_t s_{t-1} + \gamma_t \tag{3}$$

In principle, the covariance matrix of σ_γ^2 is an unknown parameter and should also be estimated with every incoming vibration measurement. However, for the sake of simplicity and computational economy, such a condition for σ_γ^2 is usually put aside, and thus, σ_γ^2 is assumed to be a constant. Adaptive estimation of σ_γ^2 in a state-space model is a research topic which has received considerable attention in recent years [12,13] but is beyond the focus of this study. The unknown parameters contained in M_t are termed "system parameters" and are estimated simultaneously with the state vector s_t. Therefore, the extended Kalman filter (EKF) is necessary under such a scenario.

An AR model order must be specified prior to the basic estimation procedure introduced above. However, unlike the conventional time-series model order selection which is usually based upon the Akaike information criteria (AIC) or the like, this study attempts to establish a generalized time-varying AR model which furnishes a compromised model fit to multiple healthy gear motion signals collected under distinct load conditions. Such a scheme stems from a critical finding in previous research work [14], which indicated that the time-series model order is the decisive parameter which determines whether or not a statistical condition indicator extracted from time-series model residuals is able to be independent of load condition. However, the search for optimum AR model order was carried out by means of a subjective assessment of a few performance statistics in Ref. [14], which to a certain degree degrades the real-time applications of the proposed technique. To overcome this disadvantage, this study will propose a novel and automatic model order selection

method without human supervision using a non-paired two-sample Satterthwaite's t'-test, which is specifically for fitting an AR model to multiple healthy gear vibration signals collected under distinct load conditions, where the signals to be processed herein are gear motion residual (GMR) signals extracted from TSA signals. Gear motion residual signals are simply the difference (amplitude as a function of time) between a time synchronous averaged signal and a raw dynamic signal recorded over the same length of time.

To be associated with and to satisfy the needs of a CBM program, generalized and quantitative condition indicators are compulsory. Our literature review indicates that most advanced CBM models recently developed for rotating machinery place no specific restrictions on the relationship of condition indicators with background operating conditions, which are in fact equivalent to assuming a constant operating condition [15–17]. This is to some extent not consistent with realistic scenarios, since variable work conditions are frequently encountered in industry.

On the other hand, advanced condition monitoring techniques like wavelets or alternative sophisticated time–frequency techniques are identical to the fast Fourier transform (FFT) in the sense that both a three-dimensional time–frequency representation and a two-dimensional FFT spectral plot are essentially dependent on signal strength. Any fluctuation in load or other source inputs will result in a remarkably different signal representation in the time–frequency domain or the frequency domain. Expertise is mandatory in the explanation and understanding of the resulting spectral representation, which, however, is extremely difficult to be transformed into numerical indicators for use in a CBM program. Independent condition indicators that give a straightforward and numerical assessment of the condition of the object being monitored are critical for the needs of either condition monitoring or operational maintenance decision analysis. Many researchers have employed kurtosis values for this purpose [7,18–20] which is largely independent of load, since it is based on the statistical distribution of a signal rather than the signal strength. However, a previous study has shown that this is not necessarily true, since different loads can result in significant deviation in the distributions of resulting gear motion signals or time-series model residuals [14].

In this study, a one-sample Kolmogorov–Smirnov (K–S) test statistic (KSS) will be used as a numerical indicator to assess the condition of a target gear by considering that it is able to provide an interval $(0,cv)$ to cover normal variability of gear wear, where cv stands for the critical value of one-sample K–S test. The kurtosis value can only designate normality at a value of 3 and is sensitive to outliers. It should be pointed out that KSS could also demonstrate load-dependent behavior when the statistical distribution of a signal carries strong load-incurred features [21]. Therefore, a specific load independent technique is necessary to remove the influence of load and enable the condition indicator to be exclusively sensitive to the actual condition of the object being monitored. Such an objective is incorporated in this study and is solved when the optimum AR model order is selected by means of a non-paired two-sample Satterthwaite's t'-test.

When an automatic vibration monitoring system is of interest, the identification of effective alerts for incipient gear faults should also be automated and free from human visual inspection. This is similar to the detection of abrupt changes taking place in both the underlying process and noise disturbances, which have been extensively investigated [12,13,22]. In this study, we will propose a specific procedure based on a Wilcoxon rank-sum test for automatic identification of alerts for incipient gear faults based on the traces of condition indicators.

The remainder of this paper is organized as follows. Section 2 presents the theoretical fundamentals of the proposed vibration monitoring methodology. Section 3 presents the experimental set-up for validation of the proposed technique. Section 4 presents an experimental validation using entire lifetime vibration monitoring signals of a gearbox under constant and sinusoidal load conditions. Section 5 presents an experimental validation using entire lifetime vibration monitoring signals of a gearbox under constant load conditions with an abrupt jump. Conclusions are given in Section 6.

2. The proposed vibration monitoring system

2.1. General description

The proposed automatic vibration monitoring system can be illustrated by two separate diagrams, one for the description of off-line modeling and another for the description of actual on-line implementation as shown in Figs. 1 and 2, respectively.

The fundamental principle of the proposed methodology is the optimum AR model established on the basis of healthy-condition GMR signals of a gearbox collected under a variety of possible load conditions that form a potential load space. The resulting optimum AR model order and its model coefficients are able to give a *consistent statistic* (or *statistically equal condition indicator*) calculated based on the AR model residuals of a future healthy-condition GMR signal for the same gearbox collected under arbitrary load condition within the considered load space. A significant deviation will be generated when a faulty-condition GMR signal is filtered.

Our extensive experimental analysis shows that the healthy-condition vibration signals from a gear, collected under the lower and upper limit load conditions are sufficient for searching for the optimum AR model order. The optimum AR model order will allow a normal spread of AR model residuals from future healthy-condition GMR signals which are collected under any load condition within the pre-specified lower and upper load limits. Therefore, as shown in Fig. 1, vibration signals of only lower and upper limit load conditions are considered during the process of modeling. These signals

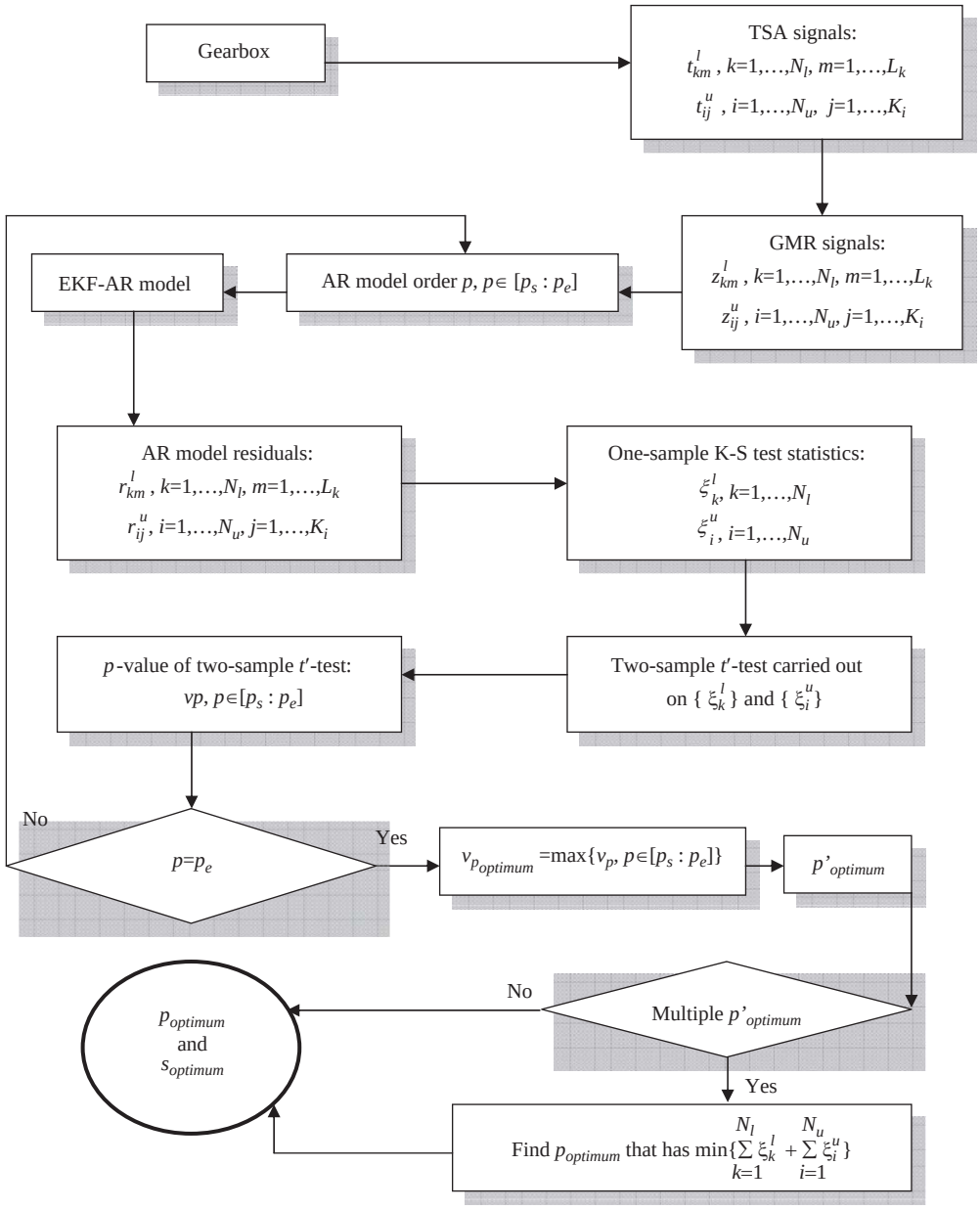

Fig. 1. Diagram of modeling of the proposed vibration monitoring system. The superscripts or subscripts *l* and *u* stand for lower and upper limit load conditions. N_l and N_u stand for the total number of signals collected under lower and upper limit load conditions, respectively. L_k and K_i stand for the number of data points in the *k*th and *i*th TSA or GMR signal of lower and upper load limit conditions, respectively. *ps* and *pe* stand for the interval from which the AR model order *p* is chosen.

consequently result in two sets of AR model residuals, r^l_{km} and r^u_{ij}, and subsequently two sets of KSS, $\{\xi^l_k\}$ and $\{\xi^u_i\}$, corresponding to the lower and upper limit load conditions, respectively.

Under a certain AR model order p, a non-paired two-sample Satterthwaite's t'-test is carried out between the two series $\{\xi^l_k\}$ and $\{\xi^u_i\}$, and generates a p-value v_p. It is known that a higher p-value of non-paired two-sample Satterthwaite's t'-test implies a statistically stronger similarity. By examining all orders within $[p_s:p_e]$ and selecting the order which results in the highest p-value, the best statistical similarity between the condition indicators corresponding to the lower limit load condition and those corresponding to the upper limit load condition can be achieved. As such, the load-dependent behavior of a condition indicator (which is KSS here) remains at the least statistical level. One may argue that the obtained statistical consistency of a condition indicator under both the lower and upper limit load conditions may possibly be of a strong violation of the normality assumption of AR model residuals in the healthy condition of the target gear. However, our extensive experiments demonstrated that the normality assumption is a weak condition. The extracted GMR signals from

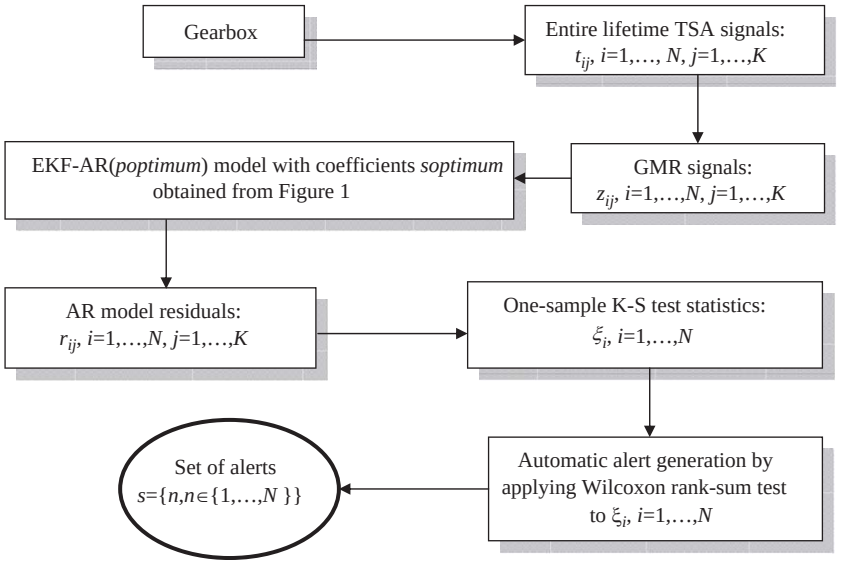

Fig. 2. Diagram of implementation of the proposed vibration monitoring system. N stands for the total number of signals collected during the entire lifetime of the gearbox of interest. K stands for the length of each TSA, GMR, or AR model residual signal. The set of alerts, which is a subset of {1,…,N}, contains the indexes of all one-sample K–S test statistics ξ_i identified by the Wilcoxon rank-sum test as a warning for incipient gear fault.

TSA together with the established AR model greatly lower the possibility of non-normality of AR model residuals. Therefore, priority is given to the statistical equivalence of a condition indicator under arbitrary load conditions.

In the case where limited vibration monitoring data are available for modeling, multiple AR model orders could result in an identical p-value v_p for their respective series $\{\xi_k^l\}$ and $\{\xi_i^u\}$. In such a circumstance, the order that gives the least violation against the normality assumption of AR model residuals (according to Eq. (4) and as shown in Fig. 1) will be selected. Calculation of the model coefficients of the selected p_{optimum} will be introduced in Section 2.2.

$$\sum_{k=1}^{N_l} \xi_k^l + \sum_{i=1}^{N_u} \xi_i^u \tag{4}$$

The practical implementation of the proposed system, as shown in Fig. 2, features the use of a Wilcoxon rank-sum test. Theoretically, the generated KSS series ξ_i, $i = 1,\ldots,N$ follows an extreme-value distribution, since ξ_i takes the largest (vertical) distance between an empirical distribution function from our data and the fitted distribution function. Thus, the conventional t-test that works best for Gaussian distributed data does not coincide with this scenario. Therefore, a non-parametric hypothesis test, like the Wilcoxon rank-sum test, that does not have a strong restriction on the distribution property of the examined data is more appropriate.

The following subsections will outline related theoretical details of the proposed vibration monitoring system.

2.2. Time-varying AR model with the aid of EKF (EKF-AR)

With the following notations

$$s_t = \text{vec}([\phi_1(t), \phi_2(t), \ldots, \phi_p(t)]^{\text{T}}) \tag{5}$$

$$Z_t = (z_t, z_{t-1}, \ldots, z_{t-p+1}) \tag{6}$$

an appropriate state space representation of Eq. (1) with stochastic coefficients can be given by

$$s_{t+1} = f(s_t) + \gamma_t \tag{7}$$

$$z_t = Z_{t-1} s_t + \alpha_t \tag{8}$$

where s_t is the $p \times 1$ state vector, γ_t is a $p \times 1$ sequence of zero-mean white Gaussian state noise, uncorrelated with s_1 and α_t, α_t is the same as in Eq. (1) and uncorrelated with s_1 and γ_t, and Eq. (8) has an adaptive time-varying coefficient Z_{t-1} of dimension $d \times p$. Also, $\hat{s}_0 (= E(s_0))$, the Gaussian $p \times 1$ initial state vector with covariance matrix $P_{0|0} (= \text{Cov}(s_0))$, and the noise covariance matrices are defined as

$$E\{\gamma_k \gamma_i^{\text{T}}\} = Q_k \delta_{k-i} \tag{9}$$

$$E\{\alpha_k \alpha_i^{\mathrm{T}}\} = R_k \delta_{k-i} \tag{10}$$

where T denotes transposition, δ denotes the Kronecker delta sequence, Q_k denotes the covariance matrix of state noise γ_t, and R_k denotes the covariance matrix of measurement noise α_t. As discussed in Section 1, suppose that the evolution law of the state vector s_t is a first-order autoregressive process with unknown coefficient matrix M_t, which results in a state-space representation

$$s_t = M_t s_{t-1} + \gamma_t \tag{11}$$

$$z_t = Z_{t-1} s_t + \alpha_t \tag{12}$$

that replaces the representation of Eqs. (7) and (8). For simplicity, M_t is assumed to be an unknown constant diagonal coefficient matrix, $M_i = \mathrm{diag}[m_1, \ldots, m_n]$. 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 s_t and identify θ which must be treated as a random vector such as

$$\theta_{t+1} = \theta_t + \zeta_t \tag{13}$$

where ζ_i is any zero-mean white Gaussian noise sequence uncorrelated with α_t and with pre-assigned positive definite covariance $\mathrm{Cov}(\zeta_t) = W_t$. In application, we may choose $W_t = W$ for all t. Now the system Eqs. (11) and (12) together with Eq. (13) can be reformulated as the nonlinear model

$$\begin{bmatrix} s_t \\ \theta_t \end{bmatrix} = \begin{bmatrix} M_{t-1}(\theta_{t-1}) s_{t-1} \\ \theta_{t-1} \end{bmatrix} + \begin{bmatrix} \gamma_{t-1} \\ \zeta_{t-1} \end{bmatrix} \tag{14}$$

$$z_t = [Z_{t-1} \;\; 0] \begin{bmatrix} s_t \\ \theta_t \end{bmatrix} + \alpha_t \tag{15}$$

where the parameters are treated as additional states and form an augmented state vector. The EKF procedure which requires a real-time linear Taylor approximation can then be applied to estimate the state vector which contains θ_t as its components, i.e.,θ_t is estimated optimally in an adaptive manner. Thus, the following EKF algorithm can be derived:

Set

$$\begin{bmatrix} \hat{s}_0 \\ \hat{\theta}_0 \end{bmatrix} = \begin{bmatrix} E(s_0) \\ \hat{\theta}_0 \end{bmatrix} \tag{16}$$

and

$$P_0 = \begin{bmatrix} \mathrm{Cov}(s_0) & 0 \\ 0 & W_0 \end{bmatrix} \tag{17}$$

The recursive prediction equations are

$$P_{t|t-1} = \begin{bmatrix} M_{t-1}(\hat{\theta}_{t-1}) & \dfrac{\partial}{\partial \theta}[M_{t-1}(\hat{\theta}_{t-1}) \hat{s}_{t-1}] \\ 0 & I \end{bmatrix} P_{t-1|t-1} \begin{bmatrix} M_{t-1}(\hat{\theta}_{t-1}) & \dfrac{\partial}{\partial \theta}[M_{t-1}(\hat{\theta}_{t-1}) \hat{s}_{t-1}] \\ 0 & I \end{bmatrix}^{\mathrm{T}} + \begin{bmatrix} Q_{t-1} & 0 \\ 0 & W_{t-1} \end{bmatrix} \tag{18}$$

$$\begin{bmatrix} \hat{s}_{t|t-1} \\ \hat{\theta}_{t|t-1} \end{bmatrix} = \begin{bmatrix} M_{t-1}(\hat{\theta}_{t-1}) \hat{s}_{t-1} \\ \hat{\theta}_{t-1} \end{bmatrix} \tag{19}$$

and the updating equations are

$$G_t = P_{t|t-1}[Z_{t-1} \;\; 0]^{\mathrm{T}} [[Z_{t-1} \;\; 0] P_{t|t-1} [Z_{t-1} \;\; 0]^{\mathrm{T}} + \hat{R}_t]^{-1} \tag{20}$$

$$P_{t|t} = [I - G_t[Z_{t-1} \;\; 0]] P_{t|t-1} \tag{21}$$

$$\begin{bmatrix} \hat{s}_t \\ \hat{\theta}_t \end{bmatrix} = \begin{bmatrix} \hat{s}_{t|t-1} \\ \hat{\theta}_{t|t-1} \end{bmatrix} + G_t(z_t - Z_{t-1} \hat{s}_{t|t-1}) \tag{22}$$

where Q_t is assumed to be known *a priori* and $\hat{R}_t = \mathrm{Cov}(\alpha_t)$. From the incoming measurement information z_t and the optimal state prediction $\hat{s}_{t|t-1}$ obtained in the previous step, the estimation sequence is defined to be

$$e_t = z_t - Z_{t-1} \hat{s}_{t|t-1} \tag{23}$$

which leads to the estimate of R_t, using the t most recent residuals, given by

$$\hat{R}_t = \frac{1}{t-1}\sum_{k=1}^{t}(e_k - \bar{e})(e_k - \bar{e})^{\mathrm{T}} \tag{24}$$

where

$$\bar{e} = \frac{1}{t}\sum_{k=1}^{t} e_k \tag{25}$$

is the mean of the estimations up to the tth time instant.

The extended Kalman filtering technique typically requires complete specifications of both dynamical and statistical model parameters of the system. However, in a number of practical situations, these models may contain parameters which may deviate from their nominal values by unknown random bias. This unknown random bias may degrade the performance of the filter or cause a divergence of the filter [23]. The two-stage extended Kalman filter (TEKF) and adaptive fading extended Kalman filter (AFEKF) which consider this problem in nonlinear systems, have been developed recently [24]. The TEKF assumes that the information of a random bias is known. The AFEKF appears to work with incomplete information of the random bias. An adaptive two-stage extended Kalman filter (ATEKF) has also been proposed [24]. The ATEKF has been shown to be able to estimate unknown random bias by using the AFEKF. The work in this paper does not use the ATEKF, the AFEKF or the TEKF because the input data we are considering is well behaved. The results suggest that the bias that may result is insignificant. Future work will be carried out to quantify the actual bias in the EKF-based method.

2.3. Selection of optimum AR model order w.r.t. variable load conditions

Assume that there is a total of N_l and N_u healthy-condition GMR signals corresponding to the lower and upper limit load conditions, respectively, used for modeling and each GMR signal is of length n which covers a complete revolution of the target gear. For each of the, say N_l, GMR signals of the lower limit load condition, the adaptive estimation procedure of the above proposed EKF-AR model generates a model coefficient matrix s_i expressed by

$$s_i = (s_1^i \;\; s_2^i \;\; \cdots \;\; s_n^i) \tag{26}$$

where s_t^i, $t = 1,\ldots,n$ represents the estimate of state vector s_t at the tth instant of time of the ith GMR signal in N_l, where $i = 1,\ldots,N_l$ and s_t is of dimension $p \times 1$. Since the phase information of rotating motion may not be available in future implementations, an averaged model coefficient vector is calculated for each of the N_l GMR signals by using

$$\bar{s}_i = \sum_{t=1}^{n} s_t^i \tag{27}$$

The same procedure can be applied to the N_u GMR signals of the upper limit load condition, which results in

$$\overrightarrow{s}_j = \sum_{t=1}^{n} s_t^j \tag{28}$$

where $j = 1,\ldots,N_u$. To reduce the load-dependent effects, a further averaging operation

$$s_p := \bar{s} = \frac{\sum_{i=1}^{N_l}\bar{s}_i + \sum_{j=1}^{N_u}\bar{s}_j}{N_l + N_u} \tag{29}$$

is carried out.

Therefore, for the present AR model order p, an averaged state vector s_p is obtained. Substituting s_p to the present AR(p) and filtering the N_l and N_u GMR signals, respectively, an estimation sequence defined by Eq. (23) can be obtained for each of the N_l and N_u GMR signals. By applying the one-sample K–S test to each estimation sequence, a *numerical value condition indicator* (or KSS) is obtained. In turn, a total of N_l and N_u KSSs, expressed by $\{\xi_k^l\}$ and $\{\xi_i^u\}$, where $k = 1,\ldots,N_l$ and $i = 1,\ldots,N_u$ as shown in Fig. 1, can be yielded for the lower and upper limit load conditions, respectively. To evaluate the statistical equivalence of these two condition indicator series $\{\xi_k^l\}$ and $\{\xi_i^u\}$, a non-paired two-sample Satterthwaite's t'-test can be applied at the significant level of 5 percent, which yields a p-value, v_p, for the present AR model order p. Following the diagram as shown in Fig. 1 and the description in Section 2.1, a *curve* (or *series*) of p-values, v_ps, can be obtained for the examined order range $[p_s{:}p_e]$. For visual convention, $(1-v_p)$s will be used in this study. Thus, a smaller $1-v_p$ corresponds to a more desirable order. When multiple orders possess the same smallest $1-v_p$, the order with the least violation against the normality assumption of AR model residuals (the smallest $\sum_{k=1}^{N_l}\xi_k^l + \sum_{i=1}^{N_u}\xi_i^u$) will be chosen to be the optimum.

The use of non-paired two-sample Satterthwaite's t'-test stems from a practical consideration that the two condition indicator series $\{\xi_k^l\}$ and $\{\xi_i^u\}$ may not have identical dimension and equal variance. Therefore, a conventional t-test is not applicable. The normality assumption required by performing Satterthwaite's t'-test is set aside, since our experimental

analysis indicates that this is a rather weak assumption, especially when a healthy-condition indicator series of the target gear is analyzed.

2.4. Automatic identification of alerts for incipient fault using Wilcoxon rank-sum test

Visual inspection emphasizes the role of human supervision and thus degrades the level of automation. In terms of practical visual inspection by which alerts for incipient gear faults are generated, we propose a novel method based on the Wilcoxon rank-sum test, which is fully automatic in practice. The proposed condition indicator KSS differs from others in that the physical condition of a gear is defined as 'healthy' when the KSS is lower than its critical value (which is 0.8950). Therefore, only the numerical values of KSS that exceed 0.8950 are of interest. This is a unique advantage possessed by KSS as opposed to other condition indicators, since during the entire lifetime of a gearbox only the numerical values of KSS that exceed 0.8950 will be examined based on the definition of 'healthy condition'. All values of other condition indicators must be examined because they do not possess inherent numerical intervals that cover the healthy condition of the target gear. Therefore, different alert generating procedures for incipient gear faults are specifically designed for KSS. Figs. 3 and 4 illustrate the automatic identification of alerts for incipient gear faults for KSS and kurtosis, respectively.

As shown in Fig. 3, there are five possible situations relating to the combination of (i,j):

(a1) $\{(m,n)|m=1,2;\ n=1\}$;
(a2) $\{(m,n)|m=2,3;\ n=2\}$;
(a3) $\{(m,n)|m\geqslant 3;\ n=1\}$;
(a4) $\{(m,n)|m\geqslant 4;\ n=2\}$;
(a5) $\{(m,n)|m\geqslant 3;\ n\geqslant 3\}$.

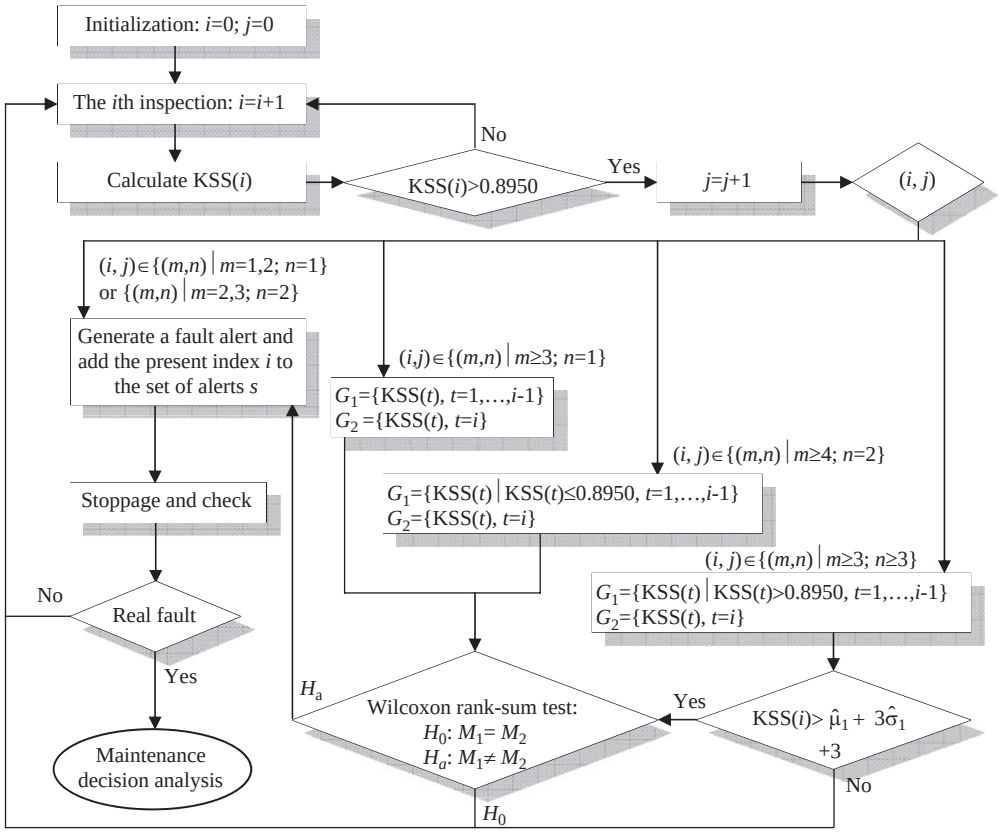

Fig. 3. Automatic identification of alerts for incipient gear faults for one-sample K–S test statistic (KSS), where the index i stands for the sequence number of vibration inspection; the index j stands for sequence number of violation against the normality assumption of AR model residuals; the calculation of KSS(i) is carried out by applying the one-sample K–S test to the AR(p_{optimum})'s estimates sequence of the ith GMR signal; 0.8950 is the critical value of one-sample K–S test; M_1 and M_2 used for Wilcoxon rank-sum test stand for the medians of two series G_1 and G_2, respectively; $\hat{\mu}_1$ and $\hat{\sigma}_1$ are the sample estimates of mean and standard deviation of the series G_2; and the definition of the set of alerts s is given in Fig. 2.

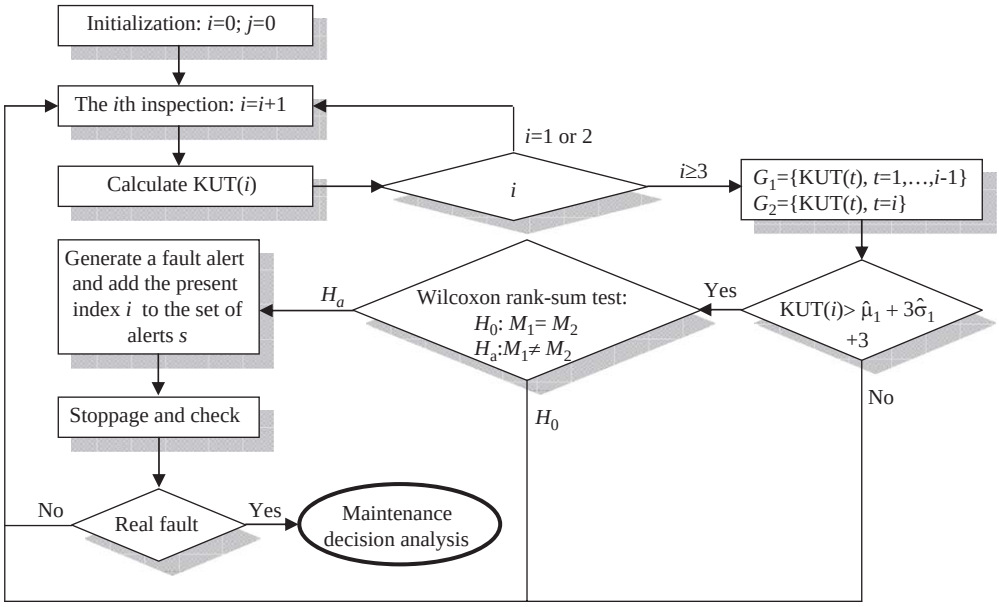

Fig. 4. Automatic identification of alerts for incipient gear faults for kurtosis or condition indicators without a critical interval to measure the normality assumption, where KUT represents the kurtosis statistic. Definitions of other notations are given in Fig. 3.

The first two possible situations (a1) and (a2) generate a fault alert and then a stoppage for troubleshooting. This is mainly due to the fact that for situation (a1), a violation of the normality assumption of AR model residuals (defined by $\mathrm{KSS}(i)>0.8950$) is present at either of the first two inspections, $i=1$ and 2. The comparison using the Wilcoxon rank-sum test between G_1 and G_2, if constructed, is invalid or unlikely to be carried out because either of these two groups has at most one data element. Therefore, when the first violation of the normality assumption of AR model residuals takes place within the first two inspections (situation (a1)) we take the conservative action that a stoppage will always be executed at the inspection where this first violation is encountered.

For situation (a2) where two violations ($n=2$) of the normality assumption of AR model residuals take place within the first two inspections ($m=2$) or the first three inspections ($m=3$), the responsive action is the same as that in situation (a1), since it is unable to carry out a further analysis when available data are sparse. Therefore, we take the same conservative action and schedule a stoppage under either of these two possibilities, $m=2$ or 3.

It should be noted that the stoppage scheduled in situations (a1) and (a2) actually implies a unique advantage possessed by KSS in comparison to condition indicators without a critical interval to help make a prompt decision on whether or not to conduct a stoppage. Therefore, for condition indicators as shown in Fig. 4, the first two inspections, $i=1$ or 2, must be skipped without further analysis even if the first two inspections present abnormally large condition indicator values.

For situation (a3), when the first violation ($j=1$) takes place at or later than the third inspection ($m\geqslant 3$), the Wilcoxon rank-sum test becomes applicable, since two or more healthy values of KSS that satisfy $\mathrm{KSS}(i)\leqslant 0.8950$ are available before the present abnormal value. In this case, a different approach is used. Two data groups G_1 and G_2 are constructed. If the result of subsequent Wilcoxon rank-sum test is H_a, a fault alert is generated and a stoppage is enforced. If the test result is H_0, the routine vibration monitoring carries on at the next inspection as described by $i=i+1$.

For situation (a4), when the second violation ($j=2$) is present at or later than the fourth inspection ($m\geqslant 4$), the Wilcoxon rank-sum test is again applicable, since two or more healthy values of a condition indicator that satisfy $\mathrm{KSS}(t)\leqslant 0.8950$ are available before the present abnormal one. Note that, by the restriction of $\mathrm{KSS}(t)\leqslant 0.8950$ in G_1, the value of KSS at the first violation ($j=1$) is ruled out from the Wilcoxon rank-sum test. The reason is that when the first violation took place, which is actually a false alert, corresponding analysis has already been conducted in situations (a1) and (a2).

For situation (a5), when three or more violations of the normality assumption of AR model residuals are present, our focus of analysis is exclusively on the values of the condition indicator that satisfy $\mathrm{KSS}(t)>0.8950$. In these circumstances all the violations of the normality assumption before the present one are false alerts. Therefore, the objective at this stage is to determine whether the present violation is statistically larger than the historical recordings of false alerts. In this case, two groups of G_1 and G_2 using only values of false alerts that satisfy $\mathrm{KSS}(t)>0.8950$ are constructed respectively for the subsequent test of $\mathrm{KSS}(i)>\hat{\mu}_1+3\hat{\sigma}_1$ and the Wilcoxon rank-sum test. By doing so, a fault alert is triggered when an abnormally large value (defined by $\mathrm{KSS}(i)>\hat{\mu}_1+3\hat{\sigma}_1$) of KSS is statistically different (defined by the alternative hypothesis H_a in Wilcoxon rank-sum test) from the historical recordings of false alerts. However, for kurtosis or condition indicators without a critical interval, it is not justifiable to focus exclusively on either its historical recordings which caused false alerts or those which did not cause false alerts, since there is no critical value available to discriminate these two types of

condition indicator values. Both of these two types of values may represent the healthy condition of the target gear even if some of them caused false alerts under the current analytical policy as illustrated in Fig. 4. Therefore, the constructed group G_1 in Fig. 4 incorporates all the historical recordings of a condition indicator.

It is noteworthy that the check of $\mathrm{KSS}(i) > \hat{\mu}_1 + 3\hat{\sigma}_1$ stems from a natural consideration in practical condition monitoring that one is only interested in the values of a condition indicator that are abnormally larger than the normal historical recordings. Here, the normal historical recordings of a condition indicator are represented by the statistical control limit expressed by $\hat{\mu}_1 + 3\hat{\sigma}_1$. The control limit applied here is a time-varying parameter, which, for KSS, varies with the new false alert values. For kurtosis or other similar condition indicators, it varies with each new value available at each new inspection.

3. Experimental set-up

3.1. General description

The data used for validation in this study were obtained from the Applied Research Laboratory (ARL) at the Pennsylvania State University (PSU). Entire lifetime vibration data sets of two single-reduction gearboxes, TR#14 and TR#12, were analyzed, where the abbreviation TR stands for test-run. ARL researchers at PSU used various combinations of motors, load equipment, and sensors on its mechanical diagnostic test bed (MDTB) and ran the MDTB to failure, collecting the resulting data and carefully documenting their observations [25]. The underlying goal of the experiments was to determine appropriate fault indicators and relate them to the rate of damage for many failure states. It is noted in Ref. [26] that data of the single axis shear piezoelectric accelerometer A03 used for axial direction measurements presents the best quality data. Therefore, data collected by A03 will be used for condition detection in this study. At every *inspection* (or *data file*) the original raw MDTB signal is a 200,000-point time-series corresponding to a 10 s interval sampled at 20 kHz. The TSA signal was calculated at every inspection and the corresponding GMR signal was extracted from the TSA signal. Fig. 5 shows an overview of the MDTB.

3.2. Scheme of comparative investigation

Two previous studies conducted analysis of the same MDTB data sets and proposed two condition indicators termed the fault growth parameter (FGP) [26] and the modified fault growth parameter (FGP1) [27] for detection of gearbox deterioration. A comparative analysis incorporates these two condition indicators as well as the proposed KSS in this study and kurtosis, where the kurtosis is applied to the same AR model residuals at each inspection as the KSS. In addition, two other condition indicators (the one-sample K–S test statistic and kurtosis applied directly to GMR signals) are included. The direct application of kurtosis to GMR signals is the conventional gear motion residual kurtosis.

The operations applied to these four condition indicators are outlined as follows:

(b1) For KSS applied to AR model residuals:
  (i) selection of optimum AR model order as shown in Fig. 1;
  (ii) automatic identification of alerts as shown in Fig. 2, using the procedure proposed in Fig. 3.

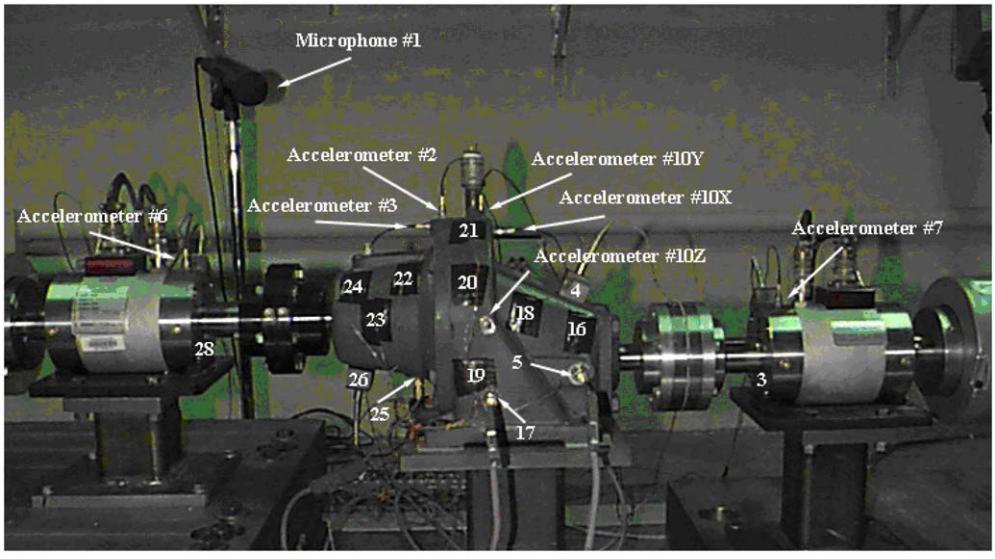

Fig. 5. An overview of the MDTB.

(b2) For kurtosis applied to AR model residuals:
  (i) selection of optimum AR model order as shown in Fig. 1, where the one-sample K–S test statistic in Fig. 1 is simply replaced with kurtosis;
  (ii) automatic identification of alerts as shown in Fig. 2, using the procedure proposed in Fig. 4;
(b3) For KSS applied to GMR signals: Automatic identification of alerts as shown in Fig. 2, using the procedure proposed in Fig. 3.
(b4) For kurtosis applied to GMR signals: Automatic identification of alerts as shown in Fig. 2, using the procedure proposed in Fig. 4.
(b5) For FGP and FGP1: Automatic identification of alerts as shown in Fig. 2, using the procedure proposed in Fig. 4.

For convenience, we define the KSS in (b1) as KSS1, kurtosis in (b2) as KUT1, KSS in (b3) as KSS2, and kurtosis in (b4) as KUT2. In the following analysis, these abbreviations together with FGP and FGP1 will be used.

The MDTB test-runs developed failures in a natural progression without stoppage and artificially created faults during their lifetime. It is thus difficult to locate the exact onset of incipient fault in the time domain. However, we can estimate the approximate time of incipient fault occurrence based on the complete trace of an already obtained condition indicator or on a specific experimental analysis. In this study, such an estimate is made on the traces of FGP and FGP1 for TR#14, since these two condition indicators were recently reported in literature [26,27], while TR#12 presents a special case that will be described in Section 5 where the approximate onset of incipient fault condition was estimated by a specific experimental analysis that is not presented in this study.

4. Constant and sinusoidal load condition

4.1. General description of TR#14

A total of 323 data files were collected by accelerometer A03 during the lifetime of TR#14. The gearbox used in TR#14 has a 70:21 tooth ratio and was driven at a constant 100 percent nominal output torque level of 555 in lbs for 95 h during which time 172 data files were collected. Subsequently, a sinusoidal load condition varying between 50 percent nominal output torque level (277 in lbs) and 300 percent nominal output torque level (1665 in lbs) which lasted 19.3 h was applied and 151 data files were collected. Fig. 6(a) shows the load condition of TR#14. The input speed only varies slightly between 1750 and 1752 rev/min throughout the entire test-run. The gear was driven until there were seven broken teeth on the drive gear, as shown in Fig. 6(b). The TSA signal, obtained at each inspection after a time-synchronous averaging operation over 87 ensembles, had a 2284 data point length. The extracted GMR signal was generated in this test-run by removing 12 sidebands on either side of every meshing frequency from the FFT spectrum of each TSA signal.

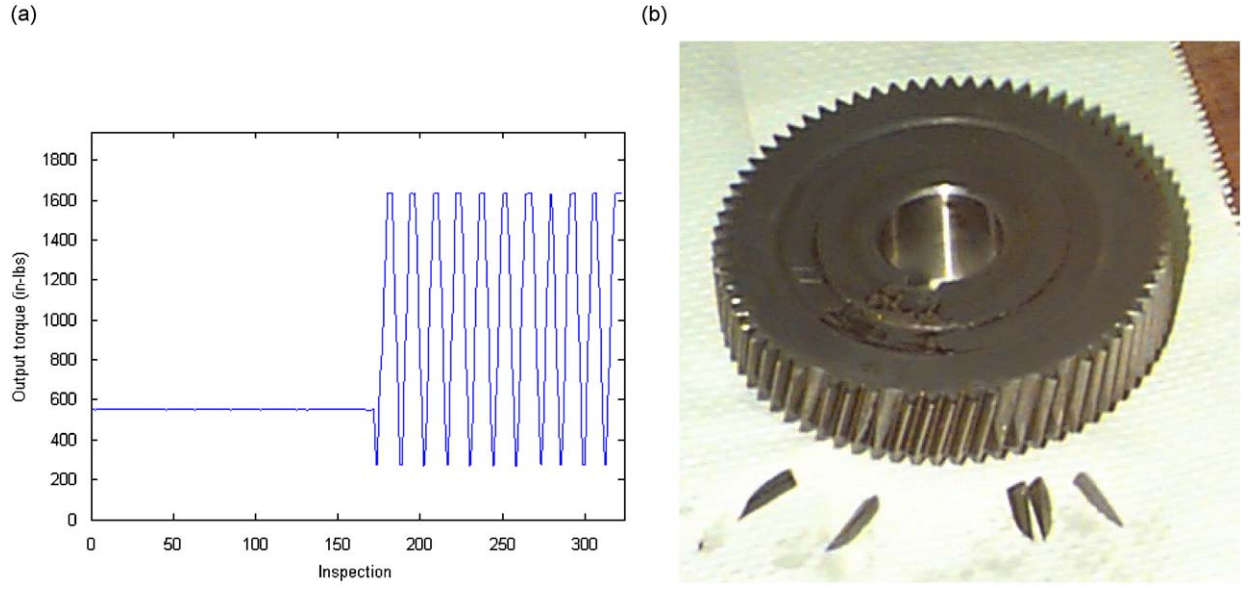

Fig. 6. Load condition and the drive gear of TR#14. (a) The output torque level switches from a constant 555 in lbs at the 172nd inspection to a sinusoidal load condition at the 173rd inspection; (b) Drive gear with seven broken teeth.

4.2. Order selection

From the total of 323 GMR signals in TR#14, 12 healthy-condition GMR signals (173, 174, 188, 189, 202, 203, 216, 217, 230, 231, 244, 245) under the 50 percent nominal output torque level, and 15 healthy-condition GMR signals (180, 181, 182, 183, 194, 195, 197, 208, 209, 210, 211, 222, 223, 224, 225) under the 300 percent nominal output torque level, are selected to establish the AR model under variable load conditions. Fig. 7 shows the AR model order selection for KSS1 and KUT1. The considered range of AR model order is (1:100). The selected optimum orders are 7 for KSS1 and 96 for KUT1.

As shown in Fig. 7, the p-value, 0.8616, of the selected $p_{\text{optimum}} = 7$ for KSS1 is larger than 0.3987 of the selected $p_{\text{optimum}} = 96$ for KUT1. Such a considerable difference in p-value implies that the proposed KSS1 as a one-sample K–S test-based condition indicator is able to achieve much better statistical equivalence under variable load conditions than kurtosis.

4.3. Detection

Fig. 8 shows the gear condition detection using KSS1, KUT1, KSS2, KUT2, FGP and FGP1, respectively. Table 1 summarizes the overall performance evaluation of these six condition indicators for TR#14 using the three properties of most concern in practical condition monitoring, namely load-independence, quantity of false alerts, and the first effective alert for an incipient gear fault.

By examining the traces of FGP and FGP1, as shown in Figs. 8(e) and (f), the approximate occurrence of incipient gear fault is around the 268th inspection, where a slight amplitude increase is visible in comparison with the previous amplitude levels. As shown in Fig. 8 and summarized in Table 1, the proposed automatic alert identification method for incipient fault identified the first effective alert at the 267th inspection for both FGP and FGP1, which is one inspection earlier than the 268th inspection of KUT1, KSS2, and KUT2 and two inspections earlier than the 269th inspection of KSS1. By checking the original data CD of TR#14 [25], it was found that the difference between the 268th and 269th inspections is trivial in terms of time of inspection (only 1.5 min). However, there is a 30 min time difference between the 267th and 268th inspections.

To a certain degree, such an early alert at the 267th inspection reflects a highly desirable property of the proposed automatic alert identification method for identifying incipient faults for condition indicators without a healthy-condition critical interval. As shown in Fig. 8(e), the amplitude of FGP(267) is similar in amplitude to the previous inspections that triggered false alerts as summarized in Table 1. The proposed automatic alert identification method for incipient fault using Wilcoxon rank-sum test successfully identified the statistical difference between FGP(267) and the previous values of FGP. However, such an advantage possessed by FGP and FGP1 comes at an expense of generating a considerable number of false alerts as shown in Figs. 8(e) and (f) and summarized in Table 1. The FGP and FGP1 both suffer from large numbers of false

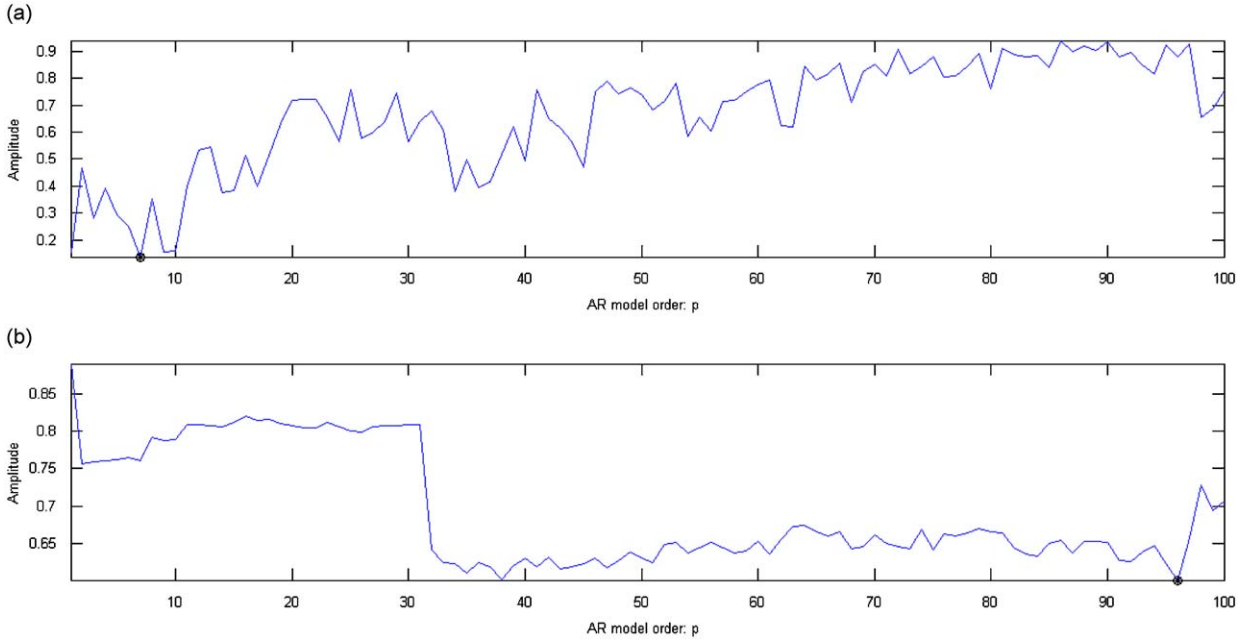

Fig. 7. AR model order selection of TR#14, where the ordinate represents the amplitude of $1-v_p$, and the order marked by a dot in each subplot is the selected order p_{optimum}. (a) $p_{\text{optimum}} = 7$ for KSS1, where $1-v_p$ is 1.3845E−01 which corresponds to a p-value of 0.8616; (b) $p_{\text{optimum}} = 96$ for KUT1, where $1-v_p$ is 6.0133E−01 which corresponds to a p-value of 0.3987.

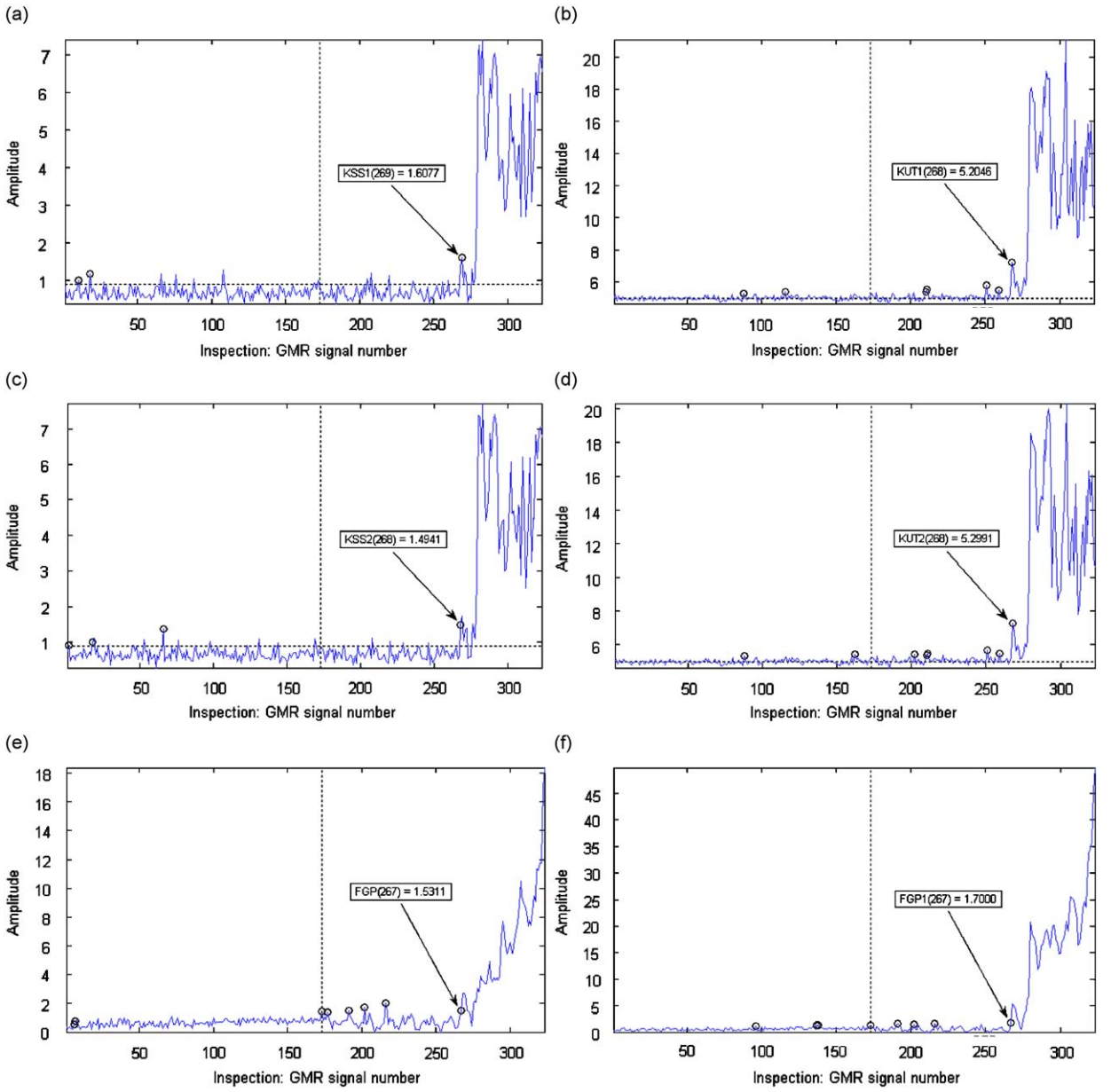

Fig. 8. Gear condition detection of TR#14. The vertical dotted line in each subplot denotes the GMR signal #173, i.e. the first inspection carried out under the sinusoidal load condition. The horizontal dotted line in subplots (a) and (c) denotes 0.8950, the critical value of one-sample K–S test. The horizontal dotted line in subplots (b) and (d) denotes 3, the value of normality based on kurtosis. The marked inspection with numerical annotation in each subplot corresponds to the true alert for incipient fault identified by the method proposed in Section 2.4. The other marked inspections without annotation in each subplot are false alerts. (a) KSS1 using AR(7); (b) KUT1 using AR(96); (c) KSS2; (d) KUT2; (e) FGP; (f) FGP1.

Table 1
Performance evaluation of condition indicators for TR#14.

Condition indicator	Load-independence	Inspections generating false alert	First alert for true incipient fault
KSS1	Yes	[10,18]	269
KUT1	Yes	[88,116,210,211,251,259]	268
KSS2	Yes	[2,18,66]	268
KUT2	Yes	[88,162,202,210,211,251,259]	268
FGP	No	[6,7,173,177,191,202,216]	267
FGP1	No	[96,137,138,173,191,202,216]	267

alerts when compared with KSS1 and KSS2. The combination of such an advantage and disadvantage is due to the theoretical basis of both FGP and FGP1, which are exclusively concerned with the amount and severity of outliers exceeding a predetermined critical region [26,27]. It is noteworthy that most false alerts of FGP and FGP1 are generated under the sinusoidal load condition starting from the 173rd inspection. Such a fact implies that both FGP and FGP1 are not robust with respect to variable load conditions.

On the other hand, the kurtosis-based KUT1 and KUT2 produced six and seven false alerts respectively, which are also significantly larger than the two and three false alerts of KSS1 and KSS2 respectively, as shown in Fig. 8 and summarized in Table 1. The larger sets of false alert of KUT1 and KUT2 are attributed to the extraordinary sensitiveness of kurtosis to the tails of a distribution, which is to some extent similar to those of FGP and FGP1.

Comparing to KUT, KUT1, FGP, and FGP1, the proposed KSS1 features the smallest set of false alerts and the best load-independence. As shown in Fig. 8(a) and summarized in Table 1, there are only two false alerts generated at the 10th and 18th inspections and no load-dependent behavior is present. However, it is not appropriate at this time to comment on the practical significance of the approximately 31.5 min time difference between the two alerts for incipient gear fault, namely the 267th inspection of FGP and FGP1 and the 269th inspection of KSS1, since it depends on each different case in actual engineering condition monitoring.

As an alternative, the p-values summarized in Table 2 furnish a highly desirable quantitative means to evaluate the load-independence of each condition indicator and the effectiveness of the EKF-AR model established on the original healthy-condition GMR signals. It should be noted that there is only one p-value available for KSS2 and KUT2, whereas there can be multiple p-values available for KSS1 and KUT1, since the established EKF-AR model can be of a variety of orders. However, only the p-value of the selected optimum AR model order p_{optimum} is considered here.

The Satterthwaite's t'-test p-values of KSS1 and KUT1 are given in Table 2; also refer to Fig. 8. The p-value of KSS2 in Table 2 is obtained by calculating one-sample K–S test statistic of the same two groups of original GMR signals that were used for modeling in Section 4.2 (12 healthy-condition GMR signals under the 50 percent nominal output torque level and the fifteen healthy-condition GMR signals under the 300 percent nominal output torque level) and subsequently carrying out Satterthwaite's t'-test on the resulting two groups of one-sample K–S test statistic. The identical process is applied to KUT2 to obtain its p-value as shown in Table 2. The p-values of FGP and FGP1 in Table 2 are generated by a direct application of Satterthwaite's t'-test on their values, which are obtained from Refs. [26,27], corresponding to the same two inspection groups, respectively.

As shown in Table 2, the p-values of KSS1 and KSS2, 0.8616 and 0.8655, have little difference. As described in Section 3.2, the proposed KSS1 is an application of one-sample K–S test to model residuals generated by the EKF-AR model established on original GMR signals, whereas KSS2 is a direct application of one-sample K–S test to original GMR signals. Therefore, the little difference in their p-values implies that in this case the established EKF-AR model does not result in a significant improvement in load-independence from KSS2 to KSS1. However, the p-values of KSS1 and KSS2 significantly exceed those of KUT1, KUT2, FGP, and FGP1. Such a difference indicates that the load-independent property can be best achieved by KSS1 and KSS2.

A favorable situation with evident improvement in p-value happens to the kurtosis-based KUT1 and KUT2. As shown in Table 2, the p-value of KUT1 under the selected $p_{\text{optimum}} = 96$ is 0.3987, while the p-value of KUT2 is 0.2140. The former p-value is significantly larger than the latter one. The false alert set of KUT1 is of one element less than that of KUT2 as summarized in Table 1. Thus, the variable load induced variation is considerably attenuated in KUT1. Alternatively, the statistical equivalence of kurtosis under variable load conditions is significantly improved, while the number of false alerts is slightly decreased. Such an increase of p-value clearly justifies the necessity and effectiveness of establishing an EKF-AR model on the original healthy-condition GMR signals to improve load-independence.

In contrast, FGP and FGP1 present considerably smaller p-values, 5.6283E−04 and 2.0649E−05, as shown in Table 2, which confirm that these two condition indicators are more strongly related to load condition. Such a disadvantage can also be detected by their traces within the inspection interval (173:270) in Figs. 8(e) and (f), respectively, which clearly vary with the sinusoidal load condition.

Based on the three equivalently weighted major properties as summarized in Table 1, the proposed KSS1 outperforms KUT1, KSS2, FGP, and FGP1 in two of them (the least false alerts and the best load-independence). It should also be noted that the proposed KSS2 presents a similar overall performance to that of KSS1 in this case study, where its set of false alerts

Table 2
Satterthwaite's t'-test p-values of condition indicators for TR#14.

Condition indicator	p-Value
KSS1	0.8616
KUT1	0.3987
KSS2	0.8655
KUT2	0.2140
FGP	5.6283E−04
FGP1	2.0649E−05

has merely one more than that of KSS1, and the time delay of 1.5 min between their first effective alerts for incipient gear fault (the 268th and 269th inspections), is marginal.

5. Constant load condition with one abrupt jump

5.1. General description of TR#12

There was a total of 155 data files collected during the lifetime of TR#12. The gearbox used in TR#12 had a 70:21 tooth ratio and was driven at the 100 percent nominal output torque (555 in lbs) for 96 h during which 124 data files were collected. Subsequently the gearbox was run at 300 percent nominal output torque (1665 in lbs) for 10.77 h during which 31 data files were collected. Fig. 9(a) shows the load condition of TR#12. The input speed was 1788 rev/min throughout the entire test-run. The gear was driven until there were two broken teeth on the drive gear, as shown in Fig. 9(b). The TSA signal obtained at each inspection is 2237 data points long after an averaging operation over 89 ensembles. The GMR signals generated in this test-run were obtained by removing 17 sidebands on either side of the meshing frequency from the FFT spectrum of each TSA signal.

TR#12 presents a fairly special gear vibration monitoring case which features two types of fault condition simultaneously. The first is a deviation from prescribed nominal operating conditions as noted by some significant frequency components appearing in the frequency domain in the middle of this test run. However, the signal strength of the significant spectral peaks in the frequency domain is evenly distributed in the time domain. This is rather different from a conventional time domain impact type fault signal, which is represented by a broad band distribution in the frequency domain. The source that introduces these significant spectral peaks in the frequency domain to this test run is unclear, since the test run was carried out under constant operating conditions from brand new to breakdown without a stoppage. Therefore, it is more appropriate to call these significant frequency components ghost components. A gearbox that is subject to strong ghost components may not necessarily suffer physical damage, but it will always cause deviation from prescribed nominal operating conditions and consequently exacerbate the stability of a manufacturing process or the dimensional accuracy and quality of manufactured products.

Detection of strong ghost components applied to a gearing system is a difficult job for condition indicators which are mostly concerned with the impulsive features of a time domain signal rather than variations in its statistical distribution. The spectral peaks caused by strong ghost components of a signal do not contribute impulsive features to its time domain representation. Therefore, such a dynamic change becomes invisible to time domain impulse related condition indicators. The strong ghost components in TR#12 are found to emerge around the 73rd inspection, and were confirmed by a series of subsequent detection tests which are not presented here. The location of these ghost components is between the second and the third meshing harmonics in the frequency domain. The second fault condition in TR#12 is the conventional time-domain impact caused by tooth breakage which is found around the 140th inspection based on the analysis of the traces of FGP and FGP1.

5.2. Order selection

From the total of 155 GMR signals in TR#12, the first 20 GMR signals (1:20) under the 100 percent nominal output torque level, and the first six GMR signals (125:130) under the 300 percent nominal output torque level, were selected to

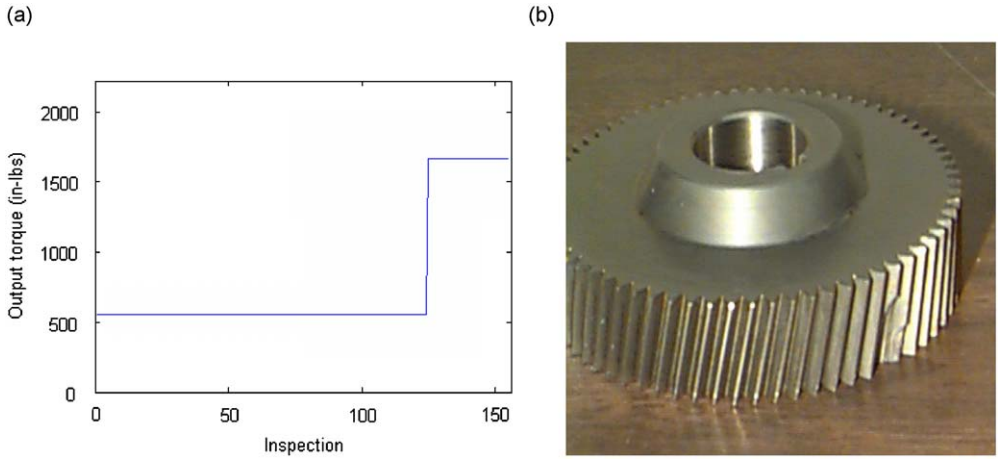

Fig. 9. Load condition and the drive gear of TR#12. (a) The output torque level switches from 555 in lbs at the 124th inspection to 1665 in lbs at the 125th inspection; (b) drive gear with two broken teeth.

establish the AR model under variable load conditions. Fig. 10 shows the AR model order selection for KSS1 and KUT1. The considered range of AR model order was still (1:100) here. The selected optimum orders were 8 for KSS1 and 100 for KUT1.

As shown in Fig. 10, the p-value, 0.9947, of the selected $p_{\text{optimum}} = 8$ for KSS1 is larger than 0.8829 of the selected $p_{\text{optimum}} = 100$ for KUT1. This confirms again the advantage of the proposed KSS1 over the kurtosis-based condition indicator KUT1 in achieving better statistical equivalence under variable load conditions. It needs to be pointed out that the selected $p_{\text{optimum}} = 100$ for KUT1 is the upper limit of the considered AR model order range in this study. The trace of $1-\nu_p$ of KUT1 in Fig. 10(b) shows that lower values of $1-\nu_p$ can possibly be obtained by exploring more orders beyond 100. However, our experimental analysis has shown that such an attempt will result in a tremendous increase in computational load. For the sake of real time applications, high orders are not recommended. Most importantly, it was found in our extensive tests that the resulting condition detection using KUT1 under high orders beyond 100 does not bring noteworthy improvement. Therefore, we simply take the selected $p_{\text{optimum}} = 100$ for KUT1 in this study.

In addition, the trace of $1-\nu_p$ of KUT1 in Fig. 10(b) demonstrates a staircase descending trend from 1 to 100 with a period of approximately 31. In contrast to KUT1, the trace of $1-\nu_p$ of KSS1 in Fig. 10(a) demonstrates a staircase ascending trend also in an approximate period of 31 but with more severe variability. In fact a similar presence can also be found in the traces of $1-\nu_p$ of KSS1 and KUT1 for TR#14 as shown in Fig. 7. The reason for this structure is uncertain at this stage and subject to further investigation.

5.3. Detection

Fig. 11 shows the gear condition detection using KSS1, KUT1, KSS2, KUT2, FGP and FGP1, respectively. Table 3 summarizes the overall performance evaluation of these six condition indicators for TR#12 in terms of the three dominant properties.

As introduced in Section 5.1, the first incipient fault condition occurs around the 73rd inspection, while the second incipient fault condition is present around the 140th inspection. It was found in our analysis that the ghost components interrupting the gearing system from the intermediate stage of this test run cause the original GMR signals within the (73:124) range to become distributed with a flat-topped feature with kurtosis value smaller than 3. We thus choose a modified kurtosis statistic by applying the operation of |KUT1-3| and |KUT2-3| to measure the deviation from the normality assumption of AR model residual or healthy-condition GMR signals, where the operation |•| denotes the absolute value. Subsequent identification of an alert for incipient fault will be based on the modified kurtosis values.

As shown in Fig. 11 and summarized in Table 3, the proposed automatic alert identification method for incipient faults identified the first effective alert for the first fault condition at the 73rd inspection for KSS1, KUT1, and KSS2, whereas the

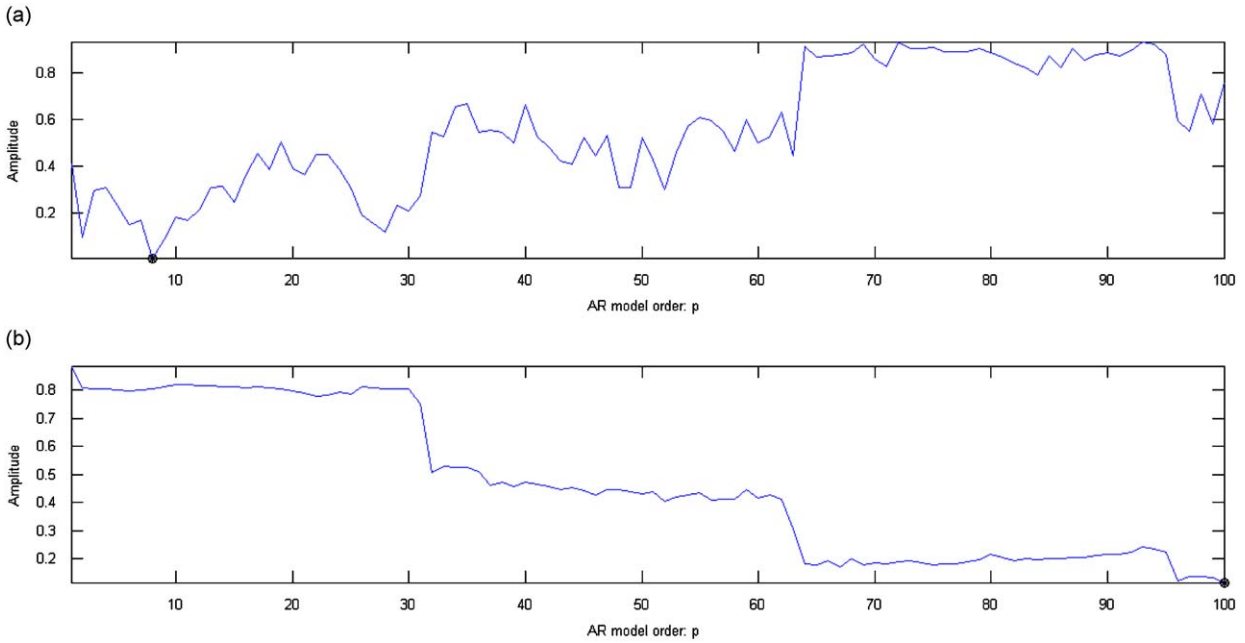

Fig. 10. AR model order selection of TR#12, where the ordinate represents the amplitude of $1-\nu_p$, and the order marked by a dot in each subplot is the selected order p_{optimum}. (a) $p_{\text{optimum}} = 8$ for KSS1, where $1-\nu_p$ is 5.3526E−03 which corresponds to a p-value of 0.9947; (b) $p_{\text{optimum}} = 100$ for KUT1, where $1-\nu_p$ is 1.1715E−01 which corresponds to a p-value of 0.8829.

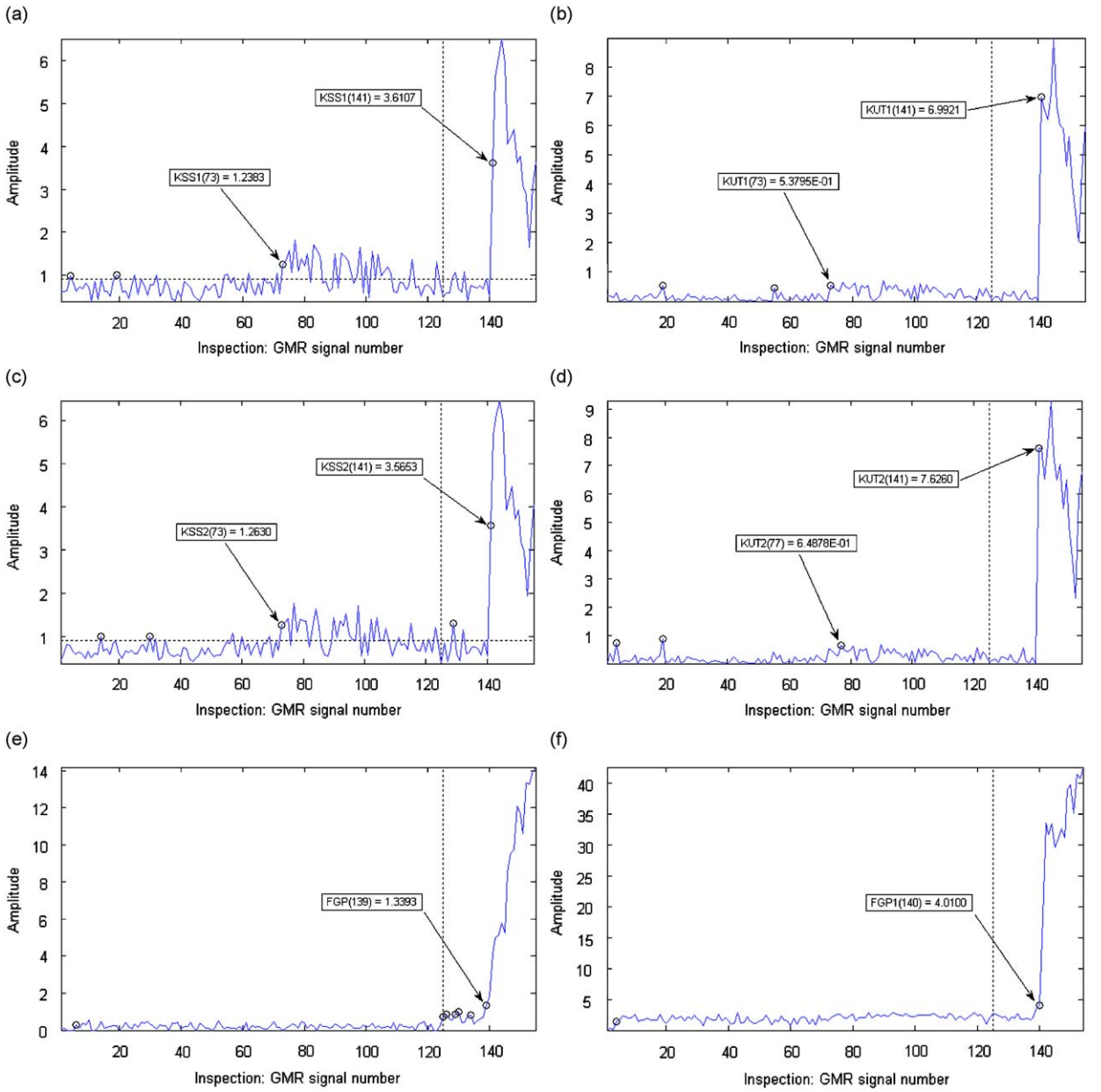

Fig. 11. Gear condition detection of TR#12. The vertical dotted line in each subplot denotes the GMR signal #125, i.e. the first inspection carried out under the second constant load condition. The horizontal dotted line in subplots (a) and (c) denotes 0.8950, the critical value of one-sample K–S test. The marked inspections with a numerical annotation in each subplot correspond to the true alerts for incipient faults identified by the method proposed in Section 2.4. The other marked inspections without annotation in each subplot are false alerts. (a) KSS1 using AR(8); (b) KUT1 using AR(100); (c) KSS2; (d) KUT2; (e) FGP; (f) FGP1.

first effective alert for KUT2 is at the 77th inspection. By checking the original data of TR#12 [25], it is found that the difference between the 73rd and 77th inspections is 2 h. All of these four condition indicators generate the same number of false alerts (4) before the presence of the first fault condition. In spite of such a difference, load-independence is achieved by all of these four condition indicators when comparing their traces within the (1:72) inspection interval and those within the (125:138) inspection interval as shown in Figs. 11(a)–(d). However, FGP and FGP1 fail to detect the presence of the first fault condition, although they generate only one false alert before the occurrence of the first fault condition at the 6th and 4th inspections, respectively.

Furthermore, FGP demonstrates apparent load-dependent behavior by comparing its performance on both sides of the vertical dotted line in Fig. 11(e) which denotes the switch between the two constant load conditions. The trace of FGP1 as shown in Fig. 11(f) does not present obvious load-dependent behavior. However, a further analysis using the *p*-value of

Table 3
Performance evaluation of condition indicators for TR#12.

Condition indicator	Load-independence	Fault #1		Fault #2	
		Inspections generating false alert	First alert	Inspections generating false alert	First alert
KSS1	Yes	[4,19]	73	None	141
KUT1	Yes	[19,55]	73	None	141
KSS2	Yes	[14,30]	73	[129]	141
KUT2	Yes	[4,19]	77	None	141
FGP	No	[6]	None	[125,126,129,130,134]	139
FGP1	No	[4]	None	None	140

Table 4
Satterthwaite's t'-test p-values of condition indicators for TR#12.

Condition indicator	p-Value
KSS1	0.9947
KUT1	0.8829
KSS2	0.7402
KUT2	0.2188
FGP	7.4289E−05
FGP1	1.2747E−02

Satterthwaite's t'-test indicates that a statistically significant difference exists in the performance of FGP1 under different load conditions which will be discussed in the following text. In addition, it is found in Refs. [21,27] that the seemingly flat transition of FGP1 from the 100 percent nominal output torque level to the 300 percent nominal output torque level around the vertical dotted line in Fig. 11(f) is not an inherent property of FGP1.

The presence of the second fault condition is identified by KSS1, KUT1, KSS2, and KUT2 at the 141st inspection. No false alert is generated by KSS1, KUT1, and KUT2. KSS2 generates a false alert at the 129th inspection. FGP features the earliest alert for incipient fault at the 139th inspection, which as denoted by the original data [25] and provides an effective alert by 1 h in comparison with the 141st inspection of KSS1, KUT1, KSS2, and KUT2. However, such an advantage is seriously degraded by a significant number of false alerts as shown in Table 3. Evidently, these false alerts of FGP are caused by the jump in load condition as shown in Fig. 11(e). FGP1 generates the first alert at the 140th inspection and provides a 30 min advanced warning when compared to the 141st inspection of KSS1, KUT1, KSS2, and KUT2 and without false alerts in this test run. Extra tests applying FGP1 to other test runs that are not incorporated in this study have shown that FGP1 is generally of the same load-dependent severity as FGP [21,27].

An alternative check using Satterthwaite's t'-test p-values summarized in Table 4 confirms our previous analysis of load-independence for these six condition indicators. The p-values in Table 4 are obtained by carrying out the Satterthwaite's t'-test on the two groups of a condition indicator's values corresponding to the same 20 GMR signals (1:20) under the 100 percent nominal output torque level, and the six GMR signals (125:130) under the 300 percent nominal output torque level, as used in Section 5.2 for optimum AR model order selection.

As shown in Table 4, the p-value of KSS1, 0.9947, is significantly larger than those of the other five condition indicators. In this test, the increase from the p-value of KSS2, 0.7402, to that of KSS1, 0.9947, signifies a significant improvement in the highly desirable load-independence. This justifies the effectiveness of the established EKF-AR model on the two group healthy-condition GMR signals used in Section 5.2. A favorable and much more evident situation takes place with the p-values of the kurtosis-based KUT1 and KUT2 (0.8829 and 0.2188, respectively) where the former is significantly larger than the latter. A further cross comparison between the p-values of KSS1 and KUT1 or the p-values of KSS2 and KUT2 reveals that the one-sample K–S test statistic is able to achieve much more robust performance under variable load conditions.

In contrast, the extremely low p-value of FGP, 7.4289E−05, implies that FGP is strongly dependent on load condition. This is consistent with our visual examination of the trace of FGP as shown in Fig. 11(e). On the other hand, the low p-value of FGP1, 1.2747E−02, is mainly due to its unstable performance within the initial inspection interval (1:5) rather than the load switch induced variability, as shown in Fig. 11(f).

In general, the proposed KSS1 is superior to the other five condition indicators in that it possesses the best load-independence and successfully detects the presence of the first fault condition with only a small set of false alerts. Although FGP and FGP1 outperform KSS1 by an earlier alert for the second fault condition, they fail to detect the presence of the first fault condition. In particular, the earliest alert yielded by FGP comes with a significantly large set of false alerts, while the seemingly stable performance of FGP1 has been found to be an occasional phenomenon.

6. Conclusions

This study investigated a fully automatic vibration monitoring system for gearbox condition detection. The proposed technique establishes a time-varying AR model with the aid of an extended Kalman filter-based on the healthy-condition gear mesh residual (GMR) signals. The order of the established AR model is selected using an automatic selection criterion based on a non-paired two-sample Satterthwaite's t'-test which enables the model to be robust with respect to variable load conditions. This specific AR model order selection technique is based upon the fact that time-series model order has been found to be the decisive parameter which determines whether or not an established time-series model is able to be independent of variable load conditions. By independence we mean that the resulting condition indicator is capable of providing a constant performance value under variable load conditions applied to a gearbox.

In this study, the proposed condition indicator is the one-sample K–S goodness-of-fit test statistic which is applied to the residuals of the established AR model. The major advantage of employing the one-sample K–S goodness-of-fit test statistic as a condition indicator is that it possesses a critical interval for evaluating the normality of AR model residuals. The normality of the vibration signal is a well-accepted assumption for a rotating machine running in a healthy condition. Fluctuation of the one-sample K–S goodness-of-fit test statistic within the critical interval still indicates a healthy condition of the monitored gearbox. This is an appealing advantage over the conventional kurtosis parameter.

Two specific automatic alert identification methods for incipient gear fault detection are proposed. The first is for the one-sample K–S goodness-of-fit test statistic. The second is for a kurtosis indicator or similar condition parameter which does not have a critical interval. Experimental results showed that the proposed automatic alert identification technique was able to generate the earliest alerts for incipient gear faults, particularly faults that result in impacts in the time-domain signal. However, these earliest alerts were obtained at an expense of a drastic increase of false alarms. The numerous false alarms were attributed to different reasons for different condition indicators. The variability of the kurtosis parameter and the load-dependent nature of FGP and FGP1 are two examples. Comparing Table 1 and Table 3 shows that KSS1 and KSS2 generally have fewer false alarms in comparison with other candidate condition indicators.

Although the earliest alerts of incipient faults can be obtained by FGP and FGP1 with the aid of the proposed automatic alert identification technique, they are not visually significant. Since FGP and FGP1 are subject to load conditions, the load-dependent behavior casts a serious doubt on the validity of the generated alerts. For instance, as shown in Figs. 8(e) and (f), the alerts of FGP and FGP1 at the 267th inspection are located at a place where there exists a strong possibility that maintenance personnel may take this alert as a normal reaction to the sinusoidal load condition before scheduling a shutdown for troubleshooting.

In addition, it is noteworthy that the somewhat similar technique proposed in Ref. [7] assumes the healthy-condition TSA signals to be stationary and attempts to establish a stationary AR model with constant model coefficients on the basis of gear vibration signals collected under a specific load condition. Implementation of the established AR model must also be under the identical load condition. However, in this study, healthy-condition GMR signals are used instead of TSA signals, while the GMR signals used for modeling are assumed to be non-stationary in light of imperfections and irregularities of gear tooth profile. Most importantly, the resulting AR model is established on the basis of multiple GMR signals collected under distinct load conditions rather than a specific TSA signal.

Therefore, the appealing performance of the proposed KSS1 parameter is based on the statistical stability of one-sample K–S test statistic and the critical region which covers the healthy condition of a target gear and greatly reduces the number of false alarms. Additionally, the proposed KSS1 is a statistical distribution statistic which is not affected by signal strength provided that different load conditions do not cause specific and exclusive variations in the signal. Thus, load-independence is much easier to achieve using KSS1 rather than FGP and FGP1. Therefore, when savings on maintenance cost, robust performance under variable load conditions, and conformance with prescribed nominal operating conditions are of most concern, the proposed KSS1 provides the most desirable overall quality and can be directly employed by a CBM program as a condition covariate.

Acknowledgments

The authors are grateful for the financial support provided by the Natural Sciences and Engineering Research Council (NSERC) of Canada and Syncrude Canada Limited, the National Natural Science Foundation of China under Contact no. 50675232. The authors' gratitude also goes to the Applied Research Laboratory at the Pennsylvania State University for providing the MDTB data.

References

[1] Y. Ohue, A. Yoshida, M. Seki, Application of the wavelet transform to health monitoring and evaluation of dynamic characteristics in gear sets, *Proceedings of the I MECH E Part J Journal of Engineering Tribology* 218 (2004) 1–11.

[2] G. Meltzer, N.P. Dien, Fault diagnosis in gears operating under non-stationary rotational speed using polar wavelet amplitude maps, *Mechanical Systems and Signal Processing* 18 (2004) 985–992.

[3] G.Y. Luo, D. Osypiw, M. Irle, On-line vibration analysis with fast continuous wavelet algorithm for condition monitoring of bearing, *Journal of Vibration and Control* 9 (2003) 931–947.

[4] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.C. Yen, C.C. Tung, H.H. Liu, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proceeding of the Royal Society of London Series A* 454 (1998) 903–995.
[5] Z.K. Peng, P.W. Tse, F.L. Chu, A comparison study of improved Hilbert–Huang transform and wavelet transform: application to fault diagnosis for rolling bearing, *Mechanical Systems and Signal Processing* 19 (2005) 974–988.
[6] S.T. Quek, P.S. Tua, Q. Wang, Detecting anomalies in beams and plate based on the Hilbert–Huang transform of real signals, *Smart Materials and Structures* 12 (2003) 447–460.
[7] W. Wang, A.K. Wong, Autoregressive model-based gear fault diagnosis, *Journal of Vibration and Acoustics* 124 (2002) 172–179.
[8] F.L. Wang, C.K. Mechefske, Adaptive modelling of transient vibration signals, *Mechanical Systems and Signal Processing* 20 (2006) 825–842.
[9] 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.
[10] M.P. Tarvainen, S.D. Georgiadis, P.O. Ranta-aho, P.A. Karjalainen, Time-varying analysis of heart rate variability signals with a Kalman smoother algorithm, *Physiological Measurement* 27 (2006) 225–239.
[11] C. Sodsri, Time-varying Autoregressive Modelling for Nonstationary Acoustic Signal and its Frequency Analysis, PhD Thesis, The Pennsylvania State University, 2003.
[12] M. Basseville, I.V. Nikiforov, *Detection of Abrupt Changes: Theory and Application*, Prentice Hall, Englewood Cliffs, NJ, 1993.
[13] M. Basseville, M. Abdelghani, A. Benveniste, Subspace-based fault detection algorithms for vibration monitoring, *Automatica* 36 (2000) 101–109.
[14] Y.M. Zhan, V. Makis, A robust diagnostic model for gearboxes subject to vibration monitoring, *Journal of Sound and Vibration* 290 (2006) 928–955.
[15] C.K. Mechefske, Z. Wang, Using fuzzy linguistics to select optimum maintenance and condition monitoring strategies, *Mechanical Systems and Signal Processing* 15 (2001) 1129–1140.
[16] C.K. Mechefske, Objective machinery fault diagnosis using fuzzy logic, *Mechanical Systems and Signal Processing* 12 (1998) 855–862.
[17] P.J. Vlok, J.L. Coetzee, D. Banjevic, A.K.S. Jardine, V. Makis, Optimal component replacement decisions using vibration monitoring and the proportional-hazards model, *Journal of the Operational Research Society* 53 (2002) 193–202.
[18] P.D. McFadden, Detecting fatigue cracks in gears by amplitude and phase modulations of the meshing vibration, *ASME Transactions, Journal of Vibration, Acoustics, Stress and Reliability in Design* 108 (1986) 165–170.
[19] G. Dalpiaz, A. Rivola, R. Rubini, Effectiveness and sensitivity of vibration processing techniques for local fault detection in gears, *Mechanical Systems and Signal Processing* 14 (2000) 387–412.
[20] W. Wang, Early detection of gear tooth cracking using the resonance demodulation technique, *Mechanical Systems and Signal Processing* 15 (2001) 887–903.
[21] Y.M. Zhan, V. Makis, A.K.S. Jardine, Adaptive state detection of gearboxes under varying load conditions based on parametric modelling, *Mechanical Systems and Signal Processing* 20 (2006) 188–221.
[22] B. Huang, Detection of abrupt changes of total least squares models and application in fault detection, *IEEE Transactions on Control Systems Technology* 9 (2001) 357–367.
[23] M.J. Moorman, Mathematical analysis of bias in the extended Kalman filter, *Proceedings of the 30th Conference on Decision and Control*, Brighton, UK, December 1991.
[24] K.H. Kim, J.G. Lee, Adaptive two-stage EKF for INS-GPS loosely coupled system with unknown fault bias, *Journal of Global Positioning* 5 (2006) 62–69.
[25] MDTB test-run data CDs: TR#12 and TR#14, *Condition-based Maintenance Department*, Applied Research Laboratory, The Pennsylvania State University.
[26] A.J. Miller, A New Wavelet Basis for the Decomposition of Gear Motion Error Signals and its Application to Gearbox Diagnostics, MS Thesis, The Graduate School of The Pennsylvania State University, 1999.
[27] D. Lin, M. Wiseman, D. Banjevic, A.K.S. Jardine, An approach to signal processing and condition-based maintenance for gearboxes subject to tooth failure, *Mechanical Systems and Signal Processing* 18 (2004) 993–1007.