A Probabilistic Fault Detection Approach: Application to Bearing Fault Detection

Bin Zhang, *Senior Member, IEEE*, Chris Sconyers, Carl Byington, Romano Patrick, Marcos E. Orchard, *Member, IEEE*, and George Vachtsevanos, *Life Senior Member, IEEE*

***Abstract*—This paper introduces a method to detect a fault associated with critical components/subsystems of an engineered system. It is required, in this case, to detect the fault condition as early as possible, with specified degree of confidence and a prescribed false alarm rate. Innovative features of the enabling technologies include a Bayesian estimation algorithm called particle filtering, which employs features or condition indicators derived from sensor data in combination with simple models of the system's degrading state to detect a deviation or discrepancy between a baseline (no-fault) distribution and its current counterpart. The scheme requires a fault progression model describing the degrading state of the system in the operation. A generic model based on fatigue analysis is provided and its parameters adaptation is discussed in detail. The scheme provides the probability of abnormal condition and the presence of a fault is confirmed for a given confidence level. The efficacy of the proposed approach is illustrated with data acquired from bearings typically found on aircraft and monitored via a properly instrumented test rig.**

***Index Terms*—Fault detection, fault progression modeling, feature extraction, particle filtering, rolling element bearing, signal enhancement.**

I. Introduction

A CONDITION-BASED MAINTENANCE (CBM) program calls for the transitioning from time-based part replacement decisions in operational systems to condition-based maintenance. For the U.S. Army's CBM+ plan, there is interest, for example, in implementing diagnostic and prognostic algorithms that eliminate the need to use "time before overhaul" definitions, which currently drive maintenance and retirement schedules of certain mechanical components in vehicles [1]. Motivation for the present work stems from an interest to support such efforts, so that this paper addresses the task of enabling early detection of faults through the use of a fault (also known as anomaly) detector, which, in turn, allows for the prediction of the evolution of the faults and the estimation of the remaining useful life (RUL) of a failing component.

A variety of techniques have been proposed based on estimation theory, failure sensitive filters, multiple hypothesis filter detection, generalized likelihood ratio tests, model-based approach, among others [2]–[11]. Past research in anomaly or fault detection for engineered systems focused also on techniques to detect changes in the data by employing statistical analysis techniques, reasoning tools from the soft computing arena, spectral methods, and approaches borrowed from information theory, among others [12]–[15]. The application domain is primarily critical aircraft systems. In the absence of complex dynamic plant models, though, that are required to represent accurately all operating modes, we seek alternate approaches that can take advantage of available measurements, simple system fatigue models, and statistical analysis tools to recognize the presence of a deviation from the no-fault or baseline condition. Several authors have analyzed the advantages of particle-filtering-based fault detection and identification (FDI) algorithms, for this type of scenarios, with excellent results when compared to classical FDI approaches [16], [17]. Particle-filtering-based method, however, differs from other approaches since it does not require Gaussian additive noise assumptions to establish a closed-form expression for the evaluation of the likelihood or for purposes of residual statistical analysis. Furthermore, the proposed method provides both statistical confidence levels and an estimate for the type II detection error, two of the most important customer specifications desired in a FDI routine.

Early detection will allow for prediction of the fault evolution and estimation of the remaining useful life of the failing component before the fault progresses to a state that endangers the system's operational integrity. Early detection is usually accompanied by an unacceptable number of false alarms. A reliable and accurate fault detector, therefore, must be capable of declaring a fault only when a specified degree of confidence is reached with a prescribed false alarm rate.

In [18], we introduced a fault detector based on the estimation of features or condition indicators from sensor data that are characteristics of the abnormal behavior of the system. The key attribute of the features is that they characterize a large number of targeted faults. Their selection and extraction are critical processes that affect unequivocally the success or failure of the detector. Another critical component is the fault progression model describing the degrading state of the system. In practice, this model is often unknown and severely hinders the application of model-based detection method. With the absence

Manuscript received October 12, 2009; revised January 31, 2010; accepted May 22, 2010. Date of publication July 15, 2010; date of current version April 13, 2011. This work was supported in part by the Army Research Laboratories (ARL) under Contract W911NF- 07-2-0075.

B. Zhang, C. Byington, and R. Patrick are with the Impact Technologies, LLC, Rochester, NY 14623 USA (e-mail: bin.zhang@gatech.edu).

C. Sconyers is with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0250 USA.

M. E. Orchard is with the Departamento de Ingenieriacute;a Eléctrica, Universidad de Chile, 2007 Santiago, Chile.

G. Vachtsevanos is with the Impact Technologies, LLC, Rochester, NY 14623 USA and also with Georgia Institute of Technology, Atlanta, GA 30332 USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIE.2010.2058072

of data, it is impossible to build a data-driven model. This paper thus develops a model based on Paris' fatigue law and modified it to fit into fault mode of interest and integrate it into the fault detection architecture. A parameter adaptation algorithm is then introduced. Features and fault progression models are then set in a particle filter (PF) framework where the current feature distribution is compared with its baseline counterpart to detect a deviation or discrepancy between the two.

The paper is organized into two major sections: The first one details the fault detector architecture while the second uses an illustrative example with seeded fault to demonstrate the efficacy of the proposed approach.

II. Fault Detection Architecture

A. Fault Detection Architecture

A fault detector is a module of the Prognostics and Health Management (PHM) system intended to recognize, as early as possible, fault conditions or anomalies in the operation of a monitored system. In most real applications, fault detector is required to perform this task while minimizing both the probability of false alarms and the detection time (time between the initiation of a fault and its detection), given a fixed threshold for false positives. The approach introduced in this paper involves a comparison between the current condition of the system and its expected operational behavior. In that sense, the availability of historical data is always assumed for purposes of defining an appropriate baseline.

To properly monitor the process, a set of features characterizing the most critical aspects of the system's fault behavior is used. The features could be any signature or characteristic derived from sensor data to reveal the evolution and progression of a fault in the system. Although some of these features may be used for the detection of a very specific fault mode, it is desirable that most of them undergo significant perturbations in the presence of several different fault scenarios.

Fig. 1 depicts the major modules of the proposed architecture for a fault detector [18]. In this architecture, real-time measurements (from sensors conveniently located and designed to respond to a large class of fault modes) and information about the current operational mode are provided online. Data are preprocessed and enhanced before computing the features that will assist to efficiently monitor the behavior of the plant. Features, as shown in Fig. 1, can be extracted from the time domain, the frequency domain, and the envelope analysis. In the example in Section III, the feature is extracted from the envelope signal in the frequency domain. Then, statistical analysis applied to this set of features, which compare their evolution in time with respect to baseline data, is performed to simultaneously arrive at the probability of abnormal conditions, the probability of false alarms, and a simple on/off indicator. This indicator exhibits the exact time instant when the presence of a fault can be confirmed at a given confidence level (for example, 95%). If time and computational resources allow for further analysis, feature information can be used to complete the tasks of fault isolation, identification, and failure prognosis. The modules of the architecture have been tested in experiment with actual

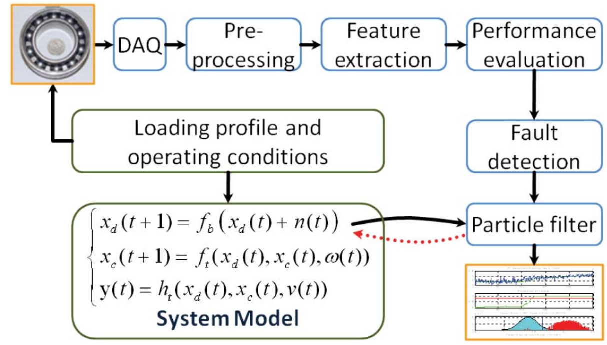

Fig. 1. Proposed architecture for a fault detector.

vibration data and they have shown to perform efficiently with low and affordable computational burden.

The architecture proposed in Fig. 1 is perfectly suitable for a particle-filter-based framework for fault detection [18], [19]. Particle Filtering is an emerging and powerful methodology for sequential signal processing based on the concepts of Bayesian theory and sequential importance sampling. Particle Filtering can be used for a wide range of applications in science and engineering, and is very suitable for nonlinear systems or in the presence of non-Gaussian process and observation noise. It has been successfully applied to the diagnosis and prognosis of gearbox and battery [19], [20] fault detection and prediction of the RUL. This approach offers a convenient compromise between data-driven and model-based techniques, but also the means to discuss its performance in terms of statistical indices.

$$\begin{aligned}
&\begin{bmatrix} x_{d,1}(t+1) \\ x_{d,2}(t+1) \end{bmatrix} = f_b\left(\begin{bmatrix} x_{d,1}(t) \\ x_{d,2}(t) \end{bmatrix} + n(t)\right)\\
&x_c(t+1) = x_c(t) + \beta\left(x_c(t)\right) \cdot x_{d,2}(t) + \omega(t)\\
&y(t) = x_c(t) + v(t)\\
&f_b(x) = \begin{cases} [1 \quad 0]^T, & \text{if } \left\| x - [1 \quad 0]^T \right\| \leq \left\| x - [0 \quad 1]^T \right\| \\ [0 \quad 1]^T, & \text{else} \end{cases}\\
&[x_{d,1}(0) \quad x_{d,2}(0) \quad x_c(0)] = [1 \quad 0 \quad 0]. \qquad (1)
\end{aligned}$$

The fault detection procedure fuses and utilizes the information present in a feature vector (observations) with the objective of determining the operational condition (state) of a system and the causes for deviations from desired behavioral patterns. From a nonlinear state estimation standpoint, this may be accomplished by the use of a particle filter-based module built upon the nonlinear dynamic state model (1), where f_b is a nonlinear mapping, $x_{d,1}$ and $x_{d,2}$ are Boolean states that indicate normal and faulty conditions, respectively, x_c is the continuous-valued state that represents the fault dimension, $y(t)$ is the fault dimension, $w(t)$ and $v(t)$ are noise signals, $n(t)$ is i.i.d. uniform white noise, and β is a time-varying model parameter that describes the progression of the fault dimension under a fatigue stress. It depends on the loading profile that is being applied to the component (bearing, for example) under test and is given in the form of the advancement da of fault dimension per cycle dR

$$\beta = \frac{da}{dR} = C(\Delta K)^m \qquad (2)$$

where a is the fault dimension, C and m are material constants, and ΔK is the stress intensity factor that is determined by fault dimension, load, etc. Models (1) and (2), however, are often unknown in practice and this severely limits the implementation of fault detection. To solve this problem, the models are modified in a form suitable for fault detection and an adaptive online training algorithm is applied to parameter tuning.

In the model (1), two separate Boolean variables $x_{d,1}$ and $x_{d,2}$ are used because there are a number of particles and each particle indicates its individual normal or faulty condition. The condition of different particles could be different. For particles that $x_{d,2} = 1$, a faulty condition is indicated and the fault dimension progresses according to $x_c(t+1) = x_c(t) + \beta(x_c(t))$; while when $x_{d,2} = 0$, a healthy bearing is indicated and the fault dimension does not progress.

A particle-filter-based fault detection routine using model (1) allows for a statistical characterization of both Boolean and continuous-valued states, as new feature data (measurements) are received. As a result, at any given instant of time, this framework will provide an estimate of the probability masses associated with each fault mode, as well as a probability density function (PDF) estimate for meaningful physical variables in the system. Once this information is available within the fault detection module, it is conveniently processed to generate proper fault alarms and to inform about the statistical confidence of the detection routine.

The outputs of the detection module may be defined as the expectations of the Boolean states in model (1). This approach provides a recursively updated estimate of the probability for each fault condition considered in the analysis. These expectations may activate alarm indicators or prognostic modules if they exceed appropriate thresholds for the probability of detection (typically, 90% or 95%). This is a particularly useful approach when the normal operation of the system is defined through a dynamic state-space model. In addition, it is also possible to define the output of the detection module as the statistical confidence needed to declare the fault via hypothesis testing. This test is performed employing the PDF estimate of one of the continuous valued states in model (1) and another PDF defining the desired condition (baseline). This approach allows for the inclusion of variables with a physical meaning into decision making. Additionally, it is particularly useful when diagnosing deviations from a specified setpoint.

It is also important to mention that the proposed fault diagnosis framework allows for the use of the PDF estimates for the system continuous-valued states (computed at the moment of fault detection) as initial conditions in failure prognostic routines, giving a suitable insight into the inherent uncertainty in the prediction problem.

It must be emphasized that sensor data acquired through an onboard health and usage monitoring system must be processed appropriately before it becomes an input to the fault detection module (see Fig. 1). Raw vibration data must be preprocessed via filtering and enhancement routines and features or condition indicators extracted to reduce the data dimensionality while maintaining the targeted (fault) information content for further utility by the fault detector. These essential steps are described in detail in the sequel. The illustrative example of the next section is used to highlight the development and application of the sequence of processes shown in the architecture of Fig. 1 leading to our ultimate objective-fault detection.

B. Fatigue Fault Progression Modeling

As mentioned earlier, system model (1) is often not available, a solution is to build a general model and estimate the model parameters through on-line adaptation and learning.

The fatigue crack progression model is built on the basis of Paris' law, as shown in (2). It expresses the fault growth rate in terms of length per running cycle as an exponential function of stress intensity factor range ΔK. In our case of a bearing, the defect dimension is represented by spall area D and therefore, model (2) is modified to

$$\frac{dD}{dt} = C(D)^m. \tag{3}$$

Equation (3) states that the rate of defect growth is related to the instantaneous defect area D under a steady operating condition. Again, C and m are material-related constants. Therefore, a defect area growth model, in discrete time form with unit step size can be written as:

$$D(t+1) = D(t) + C\left(D(t)\right)^m. \tag{4}$$

It is noticed that the progression of defect area under tight controlled conditions could show significantly different behavior. Therefore, the deterministic model (4) must be modified to take into consideration this situation. Theoretically, the uncertainty is due to the stochastic characteristics of Paris' Law and, therefore, it is reasonable to add a random variable. In practice, adding a random variable into Paris' Law is the same as adding a random variable into defect area growth model (4). Therefore, model (4) can be further developed by introducing a random variable $N(t)$ to represent the uncertainties. In practice, due to the change of operating conditions and environment, as well as the fact that the true values of parameters C and m are not available. They are treated as time-varying unknown parameters $C(t)$ and $m(t)$ that need to be adjusted online. Taking into consideration of these factors, model (4) is modified as

$$D(t+1) = D(t) + C(t)\left(D(t)\right)^{m(t)} + N(t). \tag{5}$$

To determine the parameters, a recursive least square algorithm with a forgetting factor is employed as follows.

Step 1) define a cost function as

$$J(\theta) = \frac{1}{2}\sum_{t=1}^{T} \lambda^{T-t}\left[D(t) - D\left(\hat{\theta}(t-1)\right)\right]^2 \tag{6}$$

where λ is forgetting factor, which is usually given in the range of $0 < \lambda \leq 1$, and $\hat{\theta}(t) = [C(t)\ m(t)]^{\mathrm{T}}$ is parameters need to be determined.

Step 2) Calculate the derivatives with respect to parameters θ:

$$\phi(t) = \left.\frac{dD(t,\theta)}{d\theta}\right|_{\theta=\hat{\theta}(t-1)}. \tag{7}$$

Step 3) Parameters update is given by:

$$\hat{\theta}(t) = \hat{\theta}(t-1) + P(t)\phi(t)\left[D(t) - D\left(\hat{\theta}(t-1)\right)\right] \quad (8)$$

and $P(t)$ is updated as

$$P(t) = \frac{P(t-1)}{\lambda}\left[1 - \frac{\phi(t)\phi^T(t)P(t-1)}{\lambda + \phi^T(t)P(t-1)\phi(t)}\right]. \quad (9)$$

The recursive least square with a forgetting factor applies an exponential weighting to the past data. In the cost function (6), the influence of past data reduces gradually as new data come in. This algorithm can be easily applied online.

III. Illustrative Example

A. Target System

We consider in this case study, faults occurring in a typical mechanical component-a rolling element bearing, found in many aerospace and industrial applications. The bearing, as a critical component in rotating equipment, has been studied extensively over the past years with respect to its fault or degradation modes, models describing its behavior and the development of diagnostic and prognostic algorithms to detect faults and predict their remaining useful life. Bearing failure may severely affect the health of the assembly. Of interest are assemblies of a helicopter's drive system, such as the oil cooler. The components most prone to failure are bearings supporting the cooling fan of the oil cooler assembly. Fault modes include grease breakdown, corrosion, and spalling.

In the absence of actual on-platform baseline and fault data, suitable for fault diagnosis and failure prognosis, Impact Technologies conducted a series of seeded fault tests using commercially available bearings that have similar configurations as those encountered on the helicopter's oil cooler. Proof-of-concept of fault detection in this case, is the primary objective of the current study. As more appropriate test data become available, these findings will be further improved, tuned, and validated.

One intended task in this paper is to propose techniques to improve existing U.S. Army detection capabilities for UH-60 oil cooler fan bearing faults, such as those described in [2]. In an initial effort to demonstrate the basic capabilities of the fault detection methodology discussed in this paper, the team proceeded to test enhanced fault detection techniques on an industrial bearing with a small, naturally occurring fault. It should be noted that this bearing is similar but not exactly the same type found in the UH-60 oil cooling system. Future work will build upon these initial developments leading to the implementation of fault detection of fatigue-related faults in UH-60 oil cooler fan bearings.

The bearing tested was a superprecision steel bearing subjected to steady operating conditions (3000 r/min) in the bearing test stand shown on Fig. 2, but with about 1.2 times that of the rated maximum load. A naturally occurring spall was observed in the bearing after about 8-h running time.

Vibration data (sampling frequency of 204 800 Hz) were acquired at different times throughout the test after the system

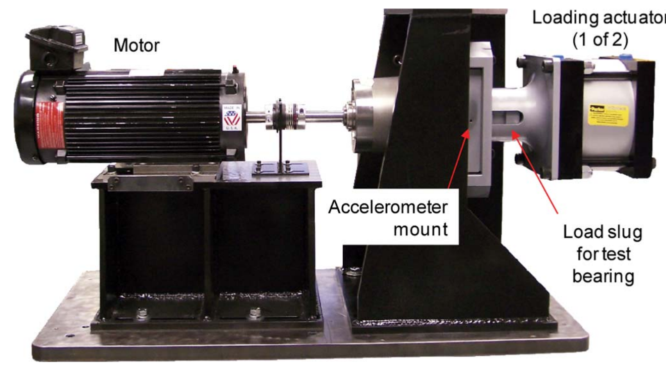

Fig. 2. Bearing test stand.

TABLE I
VIBRATION DATA SEGMENTS ANALYZED TO DEMONSTRATE THE FAULT DETECTION METHODOLOGY

Data segment No.	DN1	DN2	DN3	DN4
Data length	8 s	8 s	8 s	8 s
Bearing run time	1.5 hrs	3.5 hrs	13 hrs	15.5 hrs
Estimated Bearing condition	Healthy	No obvious spall	Spalled	Spalled

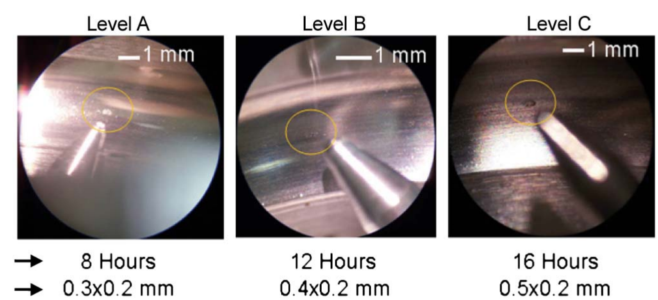

Fig. 3. Inspections of a bearing with a naturally occurring spall.

was allowed to become thermally stable. To demonstrate the fault detection methodology, several data segments listed in Table I are utilized. The second row describe the duration of time. The third row indicates the service time of bearing. The fourth row gives the estimated bearing condition, which is not measured fault dimension but only a rough classification through eye check. Some measured fault dimensions are shown in Fig. 3.

The components that are part of the proposed architecture are now described. First, the data are collected from the system of interest. It is preprocessed by adaptive line enhancement technique. Then, envelope analysis is applied to the signal and a feature is extracted from envelope signal's frequency spectra. Feature, together with a fault progression model, is fed into a particle-filtering-based detection algorithm. The fault detection generates some results in different forms and sends the results to customer for decision making.

B. Signal Enhancement

In practice, the collected vibration signals are often corrupted by a wideband noise, which will degrade the quality of features and will eventually increase the false alarm rate while decreasing the accuracy and precision of failure prognosis. Signal

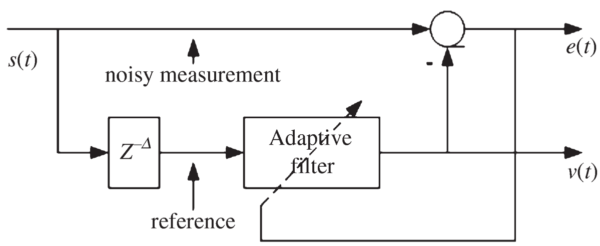

Fig. 4. Adaptive line enhancement scheme.

enhancement techniques are, therefore, essential for early and accurate fault (anomaly) detection. Improvements in terms of increased signal-to-noise ratio, reduced signal variability, and optimum selection and extraction of features will contribute to higher detection confidence and smaller false alarms rates. Such signal enhancement tools may be applied to the time domain signal directly or its frequency spectrum. Signal enhancement has shown to be effective in improving signal quality [21]. Blind deconvolution, spectral subtraction, and adaptive line enhancement techniques have been applied to signals with a reasonable degree of success.

The vibration can be described as $s(t) = v_b(t) + d(t)$, where $s(t)$ is the measured signal, $v_b(t)$ is the true vibration signal which can be considered as the sum of narrowband signals spaced by a fault characteristic frequency, and $d(t)$ is broadband noise. It is known that when a fault, for example, a spall on the raceway surface, is initiated in a bearing, an impulse of vibration is generated when the rolling elements pass through the spall [14]. This impulse is normally repetitive because the rolling elements contact the defect on the bearing surface in a periodic manner. The impulse generated by the defect is much shorter than the bearing rotating period. Therefore, the energy of the impulse will be distributed across a very wide frequency range, which will excite various resonances of the bearing [23]. The frequency of the impulses is usually regarded as the bearing characteristic defect frequency and is critical for fault detection purposes.

Since adaptive line enhancement technique [22] is widely used to separate narrow band signals from broadband noise, this algorithm is applied to bearing vibration data in the experiment. The adaptive line enhancement scheme is shown in Fig. 4. In the scheme, the delayed measurement $s(t - \Delta)$ is used as the reference signal and is fed into an adaptive filter. The output of the filter is compared with the measurements$s(t)$ to derive the error signal $e(t) = s(t) - F(s(t - \Delta))$, where $F(\cdot)$ is the adaptive filtering process. The error signal is then used to tune the filter parameters via a least mean square method [24].

As shown in Fig. 4, in this adaptive line enhancement scheme, the correlation of the broadband noise $d(t)$ in the delayed reference signal $s(t - \Delta)$ and the measurement $s(t)$ is removed by the delay Δ. However, the correlation of the true vibration signal $v_b(t)$ in these two signals is still high. By minimizing the error, the filter tends to keep the correlated narrowband true vibration signal while it cancels the uncorrelated broadband noise. The adaptive filter will eventually converge to a bandpass filter that has a pass band centered by the frequencies of the narrowband true vibration signal.

C. Feature Extraction

Features or Condition Indicators are characteristic signatures of the fault signal that reduce the data dimensionality without sacrificing the signal's information. They form the foundation for "good" fault detection algorithms. Their selection from a large candidate set and their efficient extraction from raw data are important functions in the PHM design.

Performance metrics are defined and fusion techniques are pursued in order to arrive at an "optimum" feature vector [21]. One of the most prominent vibration signal processing techniques for detection and diagnosis of rolling element bearing fault is envelope analysis [23]. The success of this technique is due to its ability to separate the vibration generated by a defective bearing from the one generated by other machine elements. For bearings with a localized defect caused by corrosion in service, the contact between rolling elements and raceway surface(s) at the defect zone under load generates an impulse train of vibration signals. In this impulse train, each impulse force has a very short time duration compared to the interval between impulses. In the frequency domain, this short time impulse will cause an energy distribution across a wide-frequency range and, therefore, excite various resonances of the bearing and the surrounding structure. Due to the periodic characteristics of the impulse train, the excitation of resonance is repetitive. Envelope analysis, by demodulating the vibration signals at the resonances that the impulse train is excited, provides a mechanism for extracting the periodic excitation of the resonance from vibration signals. The frequency of the extracted signal is the frequency of the impulse train, i.e., the characteristic bearing defect frequency.

Envelope analysis is realized through the application of a bandpass filter followed by a Hilbert transform and the construction of the analytic signal. The envelope signal is simply the absolute value of the analytic signal. Finally, via Fourier transform, the frequency spectrum of the envelope signal is derived for feature extraction.

The feature extraction scheme is illustrated in Fig. 5. In this example, the weighted energy in a frequency band centered on a frequency of interest is calculated. First, a weighting window is defined centered around the frequency of interest as in subfigure (a). The windows have the shape of a Gaussian or Hamming kernel. The length of the window defines the frequency band required to calculate the feature. The spectrum of the envelope signal is shown in subfigure (b). This envelope spectrum is multiplied by the weighting factor window to produce the weighted frequency spectrum shown in subfigure (c). In this feature extraction, the frequency band for envelop analysis is [0.2 0.4] of the Nyquist frequency and Gaussian shape weighting window are used.

The feature of interest is the sum of all frequency components in this weighted frequency band. The feature values extracted from different data sets are shown in Fig. 6. It can be seen that the feature values from the original data are very noisy. On the other hand, the feature values from the enhanced data show that signal variability is reduced substantially while the features themselves are almost monotonic, as the fault dimension increases with the operation of the bearing. Some performance metrics to evaluate the quality of feature are

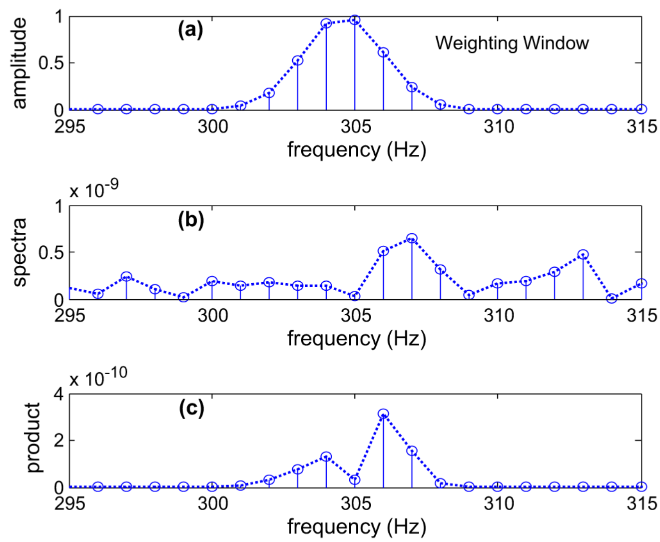

Fig. 5. Feature extraction.

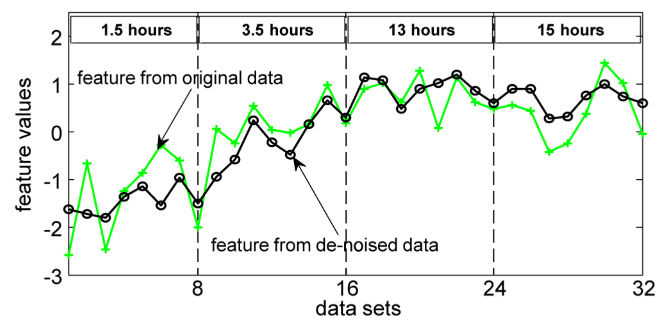

Fig. 6. Features from vibration data before and after signal enhancement.

discussed and given below. Therefore, the classification of baseline and faulty conditions through enhanced data is much easier and can reduce the false alarm rate significantly.

D. Feature Performance

The performance of the extracted feature must be evaluated to determine if it is a good candidate. Typical metrics include a precision measurement, termed the Percent Mean Deviation (PMD), measuring the noise content of a feature. It evaluates the mean deviation between the raw feature vector x and its smoothed counterpart x_s (usually a butterworth low-pass filter is used to get x_s from x), and is given by:

$$\mathrm{PMD}(x) = \left(\sum_{i=1}^{n} (x_i - x_{s,i})/x_{s,i} \right) /n \times 100. \qquad (10)$$

Feature with a small PMD value is preferred. Note that this feature is the weighted energy and, therefore, feature values are always larger than zero.

For fault detection, we consider two classes, Normal (N) and Abnormal (A). Suppose that the features from these two classes are from two Gaussian distribution and therefore can be used to construct two Gaussian PDFs. Then, the distance between these two PDFs can be used as performance metrics. The distance is

$$Distance(x) = |\bar{X}_n - \bar{X}_a| / \sqrt{s^2_{X_n}/n_n + s^2_{X_a}/n_a} \qquad (11)$$

where $\overline{X}_n$, $s^2_{X_n}$ are the mean and variance of feature values that belong to N and $\overline{X}_a$and $s^2_{X_a}$ are the mean and variance of feature values that belong to A. Obviously, a large value for distance is desirable for fault detection.

Performance metrics for the features extracted from the vibration data before and after enhancement on PMD are 4.02% and 2.18%, and on distance are 1.73 and 3.05, respectively. These features from enhanced signals are ready to be used in diagnosis.

E. The Fault Detector

Two main operating conditions are distinguished for the oil cooler bearing, the normal or baseline conditon, reflecting the fact that the bearing is new and does not exhibit any signs of spalling, and the faulty conditon, indicating an abrupt change in our indicators due to the presence of the fault (anomaly).

In this case, a particle-filtering-based detection module is implemented using model (1) to describe the expected rate of growth in the fault. Parameters associated to β are given as $\theta(0) = [C(0), m(0)] = [0.3, 0.5]$, $\omega \sim N(0.5, 0.05^2)$, and $v \sim N(0.5, 4)$, respectively. For parameter adaptation, $P(0) = [1000; 0\ 100]$ and forgetting factor $\lambda = 0.99$.

Due to the online adaptation method developed previously, values of C and $m$are online-tuned every time when a new measurement comes in. Theoretically, the initial values can be arbitrarily given. However, in practice, a good initial point not only leads to fast convergence, but also provides good performance. In our case, only very limited data points are available, it is difficult for adaptation algorithm to reach optimal value with little training. Therefore, good initial values are important for our problem. The initial values are set according to our understandings to the problem.

Besides detecting the faulty condition, it is desired to obtain some measures of the statistical confidence of the alarm signal. For this reason, two outputs are extracted from the detector module: The first one is the expectation of the Boolean state $x_{d,2}$, which constitutes an estimate of the probability of a fault, while the second is the statistical confidence needed to declare the fault via hypothesis testing (H_0: "The fault is not present" versus H_1: "The fault is present"). An online indicator of statistical confidence for the detection procedure is the probability of detection.

Fig. 7 shows the result of fault detection at different time instant (data sets are collected in time sequence). In this figure, the false alarm rate (type I error) is given as 5% and the confidence to claim the fault is 90%. That is, no fault is claimed until the probability of detection is higher than a confidence threshold of 90%. The false alarm rate and confidence of detection are the performance indices associated with the problem and therefore, are often from customer. This figure contains three subfigures that are simultaneously computed online. The first subfigure, depicted as a function of time, shows the feature or a mapping of fault/anomaly progression. In this subfigure, blue line is the feature value extracted from raw vibration data while green line is the filtered feature from particle filtering. The second subfigure depicts the probability of the fault at each time instant (green line) as it approaches a specified confidence threshold

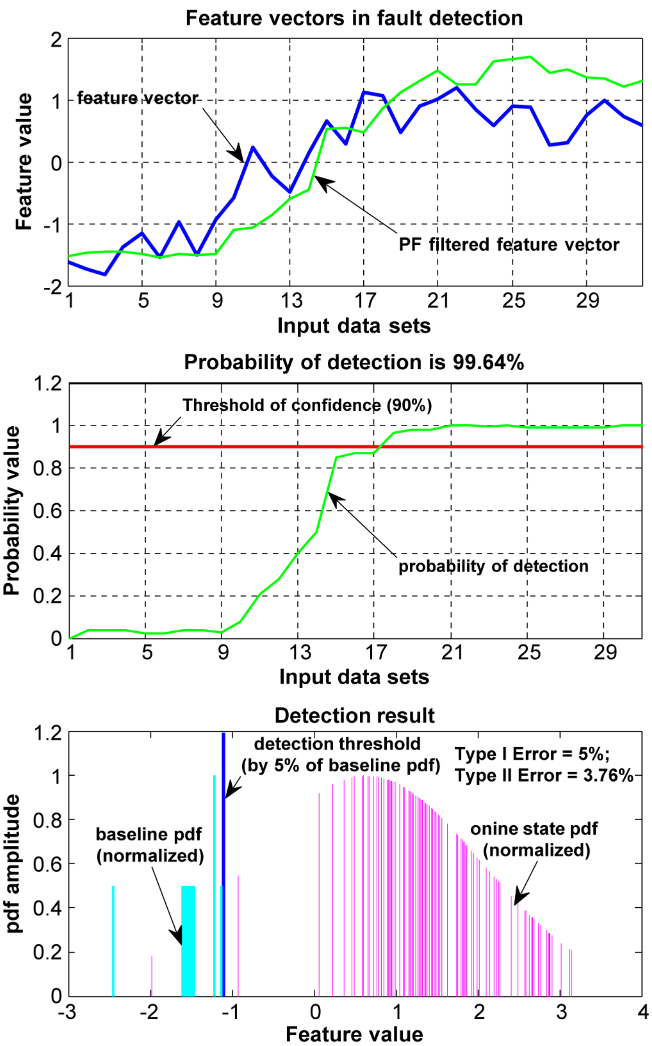

Fig. 7. Results of fault detection at different measurements.

(red line) and is based on the estimate of the Boolean state in model (1). The probability of detection is estimated as the expectation of the Boolean state $x_{d,2}$. A fault alarm may be triggered whenever this indicator reaches the confidence threshold level. The third subfigure is a depiction of the two PDFs-the baseline one (cyan) and its real-time counterpart (magenta) computed online from particle filtering as new data and features become available. The vertical line that discriminates between the two PDFs in this last figure is fixed by the desired type I detection error (probability of false positives) and baseline PDF. The type II detection error (with 1-type II error being the detection confidence) is also shown on the figure. Note that estimates for "probability of detection" and "type II error" are calculated as a new measurement comes in.

The results shown in this experiment is for system under constant operating condition. To address varying conditions, the leading factors (such as load) of fault can be divided into groups (such as light, medium, and heavy) and get baseline data for each group. Then, online fault detection can be run in parallel for each group. Another solution is to set a nominal operating condition and map the features and fault progression model under other operating conditions to this nominal condition. The fault detection under other operating conditions can be conducted under the mapped nominal condition.

IV. Conclusion

Fault detection is an essential module of the PHM system, particularly when health and usage monitoring hardware/software are operating online in real-time platform. Conflicting performance requirements-accuracy of detection, false alarm rates, and time to detection after the initiation of a fault (anomaly) must be resolved and balanced if the output of the detector will assist the system operator and lead to meaningful prognosis of the failing component's RUL. At the same time, the inherent system, measurement, and model uncertainties must be addressed and managed for optimum detection results. Bayesian estimation methods, combined with measurements and efficient data processing tools, promise to provide effective solutions to these challenging problems. Extensive research and development work and test results from actual system applications are needed to improve and validate these emerging technologies.

References

[1] D. T. Suggs and D. R. Wade, "Vibration based maintenance credits for the UH-60 oil cooler fan assembly," presented at the AHS Specialists' Meeting Condition Based Maintenance, Huntsville, AL, Feb. 12–13, 2008.

[2] J. Keller, D. Dunaway, R. Branhof, and P. Grabill, "Examples of condition based maintenance with the vibration management enhancement program," in *Proc. Amer. Helicopter Soc. 61st Annu. Forum*, Grapevine, TX, Jun. 1–3, 2005.

[3] E. Bechhoefer, "A method for generalized prognostics of a component using Paris law," in *Proc. Amer. Helicopter Soc. 64th Annu. Forum*, Montreal, CA, Apr. 29–May 1, 2008.

[4] B. Ayhan, M.-Y. Chow, and M.-H. Song, "Multiple discriminant analysis and neural network based monolith and partition fault detection schemes for broken rotor bar in induction motors," *IEEE Trans. Ind. Electron.*, vol. 53, no. 4, pp. 1298–1308, Jun. 2006.

[5] B. Ayhan, H. Trussell, M.-Y. Chow, and M.-H. Song, "On the use of a lower sampling rate for broken rotor bar detection with DTFT and AR-based spectrum methods," *IEEE Trans. Ind. Electron.*, vol. 55, no. 3, pp. 1421–1434, Mar. 2008.

[6] M. Khan and M. Rahman, "Development and implementation of a novel fault diagnostic and protection technique for IPM motor drivers," *IEEE Trans. Ind. Electron.*, vol. 56, no. 1, pp. 85–92, Jan. 2009.

[7] A. Lebaroud and G. Cleac, "Classification of induction machine faults by optimal time-frequency representations," *IEEE Trans. Ind. Electron.*, vol. 55, no. 12, pp. 4290–4298, Dec. 2008.

[8] R. Isermann and P. Balle, "Trends in the application of model-based fault detection and diagnosis of technical processes," *Control Eng. Pract.*, vol. 5, no. 5, pp. 709–719, May 1997.

[9] A. Y. Goharrizi and N. Sepehri, "A wavelet-based approach to internal seal damage diagnosis in hydraulic actuators," *IEEE Trans. Ind. Electron.*, vol. 57, no. 5, pp. 1755–1763, May 2010.

[10] L. Frosini and E. Bassi, "Stator current and motor efficiency as indicators for different types of bearing faults in induction motors," *IEEE Trans. Ind. Electron.*, vol. 57, no. 1, pp. 244–251, Jan. 2010.

[11] W. Yang, P. J. Tavner, C. J. Crabtree, and M. Wilkinson, "Cost-effective condition monitoring for wind turbines," *IEEE Trans. Ind. Electron.*, vol. 57, no. 1, pp. 263–271, Jan. 2010.

[12] D. Tolani, M. Yasar, S. Chin, and A. Ray, "Anomaly detection for health management of aircraft gas turbine engines," in *Proc. Amer. Control Conf.*, Portland, OR, Jun. 2005, pp. 459–464.

[13] W. Zhou, T. Habetler, and R. Harley, "Bearing fault detection via stator current noise cancellation and statistical control," *IEEE Trans. Ind. Electron.*, vol. 55, no. 12, pp. 4260–4269, Dec. 2008.

[14] A. Ibrahim, M. El Badaoui, F. Guillet, and F. Bonnardot, "A new bearing fault detection method in induction machine based on instantaneous power factor," *IEEE Trans. Ind. Electron.*, vol. 55, no. 12, pp. 4252–4259, Dec. 2008.

[15] N. Gebraeel and J. Pan, "Prognostic degradation models for computing and updating residual life distributions in a time-varying environment," *IEEE Trans. Rel.*, vol. 57, no. 4, pp. 539–550, Dec. 2008.

[16] N. de Freitas, "Rao-Blackwellised particle filtering for fault diagnosis," in *Proc. IEEE Aerosp. Conf.*, 2002, pp. 1767–1772.

[17] V. Kadirkamanathan, P. Li, M. H. Jaward, and S. G. Fabri, "Particle filtering-based fault detection in nonlinear stochastic systems," *Int. J. Syst. Sci.*, vol. 33, no. 4, pp. 259–265, Mar. 2002.

[18] B. Zhang, C. Sconyers, C. Byington, R. Patrick, M. Orchard, and G. Vachtsevanos, "Anomaly detection: A robust approach to detection of unanticipated faults," in *Proc. IEEE Int. Conf. Prognostics Health Manage.*, Denver, CO, Oct. 2008, pp. 1–8.

[19] M. Orchard and G. Vachtsevanos, "A particle filtering approach for on-line fault diagnosis and failure prognosis," *Trans. Inst. Meas. Control*, vol. 31, no. 3/4, pp. 221–246, Jun. 2009.

[20] M. Abbas, A. Ferri, M. Orchard, and G. Vachtsevanos, "An intelligent diagnostic/prognostic framework for automotive electrical systems," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2007, pp. 352–357.

[21] B. Zhang, T. Taimoor, R. Patrick, G. Vachtsevanos, M. Orchard, and A. Saxena, "Application of blind deconvolution de-noising in failure prognosis," *IEEE Trans. Instrum. Meas.*, vol. 58, no. 2, pp. 303–310, Feb. 2009.

[22] Y. Li, S. Billington, C. Zhang, T. Kurfess, S. Danyluck, and S. Liang, "Adaptive prognosis for rolling element bearing condition," *Mech. Syst. Signal Process.*, vol. 13, no. 1, pp. 103–113, Jan. 1999.

[23] I. Howard, "A review of rolling element bearing vibration: Detection, diagnosis and prognosis," Airframes Engines Div., Melbourne, Australia, DSTO-RR-0013, 1994.

[24] P. Zarchan and H. Musoff, *Fundamentals of Kalman Filtering: A Practical Approach*, vol. 208, *Progress in Astronautics and Aeronautics*. Reston, VA: AIAA, 2005.

Bin Zhang (S'03–M'07–SM'08) received the B.E. and M.S.E. degrees from Nanjing University of Science and Technology, Nanjing, China, in 1993 and 1999, respectively, and the Ph.D. degree from Nanyang Technological University, Singapore, in 2007.

He was a Research Fellow with School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, before he joined Impact Technologies, LLC, Rochester, NY. His interests are fault diagnosis and failure prognosis, systems and control, intelligent systems and their applications.

Dr. Zhang is an Associate Editor of IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS.

Chris Sconyers received the B.S. degree in EE and the B.S. degree in CS from Texas A&M University, College Station, in 2004. He received the M.S. degree in ECE from Georgia Institute of Technology, Atlanta, in 2006 where he is working toward the Ph.D. degree in the field of UAV-based surveillance and reconnaissance.

His interests are in data and image processing for the purposes of target tracking and subject localization.

Carl Byington received the B.S. degree in mechanical engineering from the University of Pennsylvania in 1988 and the M.S. degree in aeronautical engineering from George Washington University in 1990.

He is a Professional Engineer and the Director of Systems Engineering at Impact Technologies, Rochester, NY. He possesses over 22 years of design, analysis, and testing experience with fluid power, thermal, mechanical, and electrical systems. He has performed as the Principal Investigator (PI) for innovative PHM technologies on over 60 programs for the military, NASA, and other customers. He is also the lead instructor for the annual "PHM/CBM Design" short course offered by Impact to government and industry participants. He has produced over 100 publications related to advanced sensing techniques; signal processing; vibration monitoring; diagnostics and control; prognostics; data fusion; and equipment health management. He is also coauthor on three U.S. patents related to predictive monitoring technologies.

Dr. Byington is a member of ASME, AIAA, SAE, STLE and AHS. He is also the current Chairman of the Machinery Diagnostics & Prognostics Committee within the ASME Tribology Division. He is a past Program Chairman of the MFPT Society and the current Technical Program Chairman of the PHM Society 2010 meeting.

Romano Patrick received the MBA degree and the Ph.D. degree in electrical and computer engineering from Georgia Institute of Technology, Atlanta, in 2007, and the degrees from Univerusity of Texas, Arlington, and Universdad Panamericana, Mexico City, Mexico, respectively.

He is a Project Manager at Impact Technologies, Rochester, NY. His work and experience have involved electronic, mechanical and software design, state-of-the-art process automation, and machine health monitoring.

Marcos E. Orchard (M'06) received the B.Sc. degree and the Civil Industrial Engineering degree with electrical major from Catholic University of Chile, Santiago, Chile, in 1999 and 2001, respectively and the M.S. and Ph.D. degrees from Georgia Institute of Technology, Atlanta, in 2005 and 2007, respectively.

He is an Assistant Professor with the Department of Electrical Engineering, University of Chile, Santiago, Chile and he was part of the Intelligent Control Systems Laboratory at The Georgia Institute of Technology, Atlanta. His current research interest is the design, implementation, and testing of real-time frameworks for fault diagnosis and failure prognosis.

George Vachtsevanos (S'62–M'63–SM'89–LSM'04) received the B.E.E. degree and the Ph.D. degree in EE from the City University of New York, New York, in 1962 and 1970, respectively, the M.E.E. degree from New York University, New York, in 1963.

He is a Professor Emeritus of ECE at the Georgia Institute of Technology (Georgia Tech), Atlanta. He directs the Intelligent Control Systems laboratory at Georgia Tech, conducting research in intelligent control, fault diagnosis, and prognosis of large-scale dynamical systems and control technologies for unmanned aerial vehicles, and neurotechnology. He has published over 240 technical papers.