A Neural Network Degradation Model for Computing and Updating Residual Life Distributions

Nagi Z. Gebraeel and Mark A. Lawley, *Member, IEEE*

***Abstract*—The ability to accurately estimate the residual life of partially degraded components is arguably the most challenging problem in prognostic condition monitoring. This paper focuses on the development of a neural network-based degradation model that utilizes condition-based sensory signals to compute and continuously update residual life distributions of partially degraded components. Initial predicted failure times are estimated through trained neural networks using real-time sensory signals. These estimates are used to derive a prior failure time distribution for the component that is being monitored. Subsequent failure time estimates are then utilized to update the prior distributions using a Bayesian approach. The proposed methodology is tested using real world vibration-based degradation signals from rolling contact thrust bearings. The proposed methodology performed favorably when compared to other reliability-based and statistical-based benchmarks.**

***Note to Practitioners*—We propose a neural-network-based degradation model that estimates utilizes real-time sensory signals to estimate the failure time of partially degraded components. The proposed model has been tested and validated on rolling element bearings by using real-time vibration signatures to estimate their failure times. In order to implement, one must first identify the sensory information that is correlated with the underlying degradation process. Next, a sample of components, similar to the one being monitored is tested. Degradation-based sensory information are acquired and stored along with the corresponding operating times of each acquisition. A group of neural networks are trained using supervisory training protocols. Each neural network is trained to identify the degradation pattern of one component in the sample. This is achieved by training the network to identify the operating time corresponding to each sensory signature that is input to the network. Sensory signals from similar components operating in the field are then input to the network model and used to predict the failure time based on the latest degradation state of the component being monitored. Steps 1–4 outline the details of the implementation. The sensory-updating methodology is used to continuously update the failure time predictions as subsequent real-time signals are acquired.**

***Index Terms*—Degradation modeling, neural network, reliability, vibrations.**

Manuscript received October 16, 2005; revised March 2, 2006 and August 25, 2006. This paper was recommended for publication by Associate Editor L. Monostori and Editor P. Ferreira upon evaluation of the reviewers' comments.

N. Z. Gebraeel is with the School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: nagi@isye.gatech.edu).

M. A. Lawley is with the School of Biomedical Engineering, Purdue University, West Lafayette, IN 47907 USA (e-mail: malawley@ecn.purdue.edu).

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/TASE.2007.910302

I. Introduction

Maintenance costs constitute a large portion of the operating and overhead expenses in many industries [1]. The U.S. industry alone spends \$200 billion each year on plant maintenance [2]. An estimated 30% of this cost results from inefficient maintenance operations such as, unscheduled downtimes due to unexpected failures, poor spare parts logistics, and replacement policies [2]. In the manufacturing sector, for example, unexpected systems interruptions have become prohibitively expensive since they immediately result in lost production, failed shipping schedules, and poor customer satisfaction. In order to cope with these problems, companies need to maintain highly reliable and robust manufacturing systems that ensure maximum equipment availability, elimination of unnecessary maintenance activities, and coordinated replacement and spare parts inventory decisions [3], [4].

Condition monitoring and reliability models have been widely used to make inferences about equipment lifetime. In the reliability area, lifetime testing is usually used to assess reliability characteristics of a population of identical components. These characteristics are important for developing reliability-based preventative maintenance policies. However, due to short product life cycles and the widespread use of highly reliable components, it has become extremely difficult to make lifetime inferences using traditional reliability testing [5].

In contrast, condition monitoring uses component-specific sensory information to assess a component's health. Such information is important for condition-based maintenance [2]. In most condition monitoring applications, a component's degradation state is evaluated based on deviations from the normal running conditions [6]. Surprisingly, these deviations do not necessarily equate to failure. Often, slight deviations result from changes in operating conditions or acceptable degradation levels. Consequently, it is necessary to establish a linkage between the reliability characteristics of a component's population and condition-based sensory signals unique to the individual component in an effort to establish a comprehensive assessment of degradation processes.

Many components exhibit characteristic patterns in the sensory signals acquired through condition monitoring techniques. These signal patterns are known as degradation signals and evolve with respect to the component's state of degradation [7]. Degradation signals capture the underlying physical transitions associated with degradation. If properly modeled, degradation signals will provide accurate inferences of the remaining lives of partially degraded components. In this paper, we develop a modular neural network-based degradation model that utilizes degradation signals to compute residual life distributions of a component. These distributions are then updated in real-time using subsequent sensory information. In essence, the residual life distributions are updated based on the latest physical transitions associated with the degradation process.

The remainder of this paper is organized as follows. Section II provides an overview of bearing condition monitoring and degradation modeling literature. Section III discusses our experimental setup. Section IV discusses how we estimate failure time using a network of neural nets, while Section V illustrates how an empirical Bayes technique that uses a sequence of failure time estimations to compute a distribution on residual life. Section VI then discusses our age replacement model and applies the total approach to our bearing data. Finally, Section VII concludes with a discussion of future research.

II. Literature Review

Condition monitoring techniques are used to develop degradation signals that are used to model the evolution of different types of degradation processes. Yang *et al.* [8] studied the effect of tool wear on surface roughness for a single point machining process. The degradation signal was defined as the roughness value of a machined part and was used to model the flank wear of the cutting tool. Shao *et al.* [9] used the RMS value of bearing vibration to develop a degradation signal to assess bearing health. Tseng *et al.* [10] studied the degradation of fluorescent lamp using the degree of luminosity as a measure of degradation.

Several modeling approaches have focused on using statistical techniques to model the evolution of degradation signals. Most statistical methods focused the development of random coefficients models to characterize the path followed by degradation signals in order to estimate a component's life [11], [12]. Lu and Meeker [12] used the random coefficients models to evaluate residual life distributions. Gebraeel *et al.* [24] developed a stochastic degradation model to compute and continuously update remaining life distributions. Other researchers have utilized reliability formalisms to assess component lifetime and reliability. Jardine *et al.* [14] used proportional hazard models to develop a vibration-based decision support system for condition based maintenance. Cempel *et al.* [30] presented a new method for deriving average symptom life curves and symptom reliability for condition monitoring of machines.

Artificial intelligence techniques such as neural networks and fuzzy logic have also been used to model degradation signals. The benefit of using neural networks lies in their ability to model the evolution of complex multidimensional degradation signals. Lee [15] developed a neural network model based on cerebellar model articulation controllers (CMACs) in order to discriminate and quantify machine degradation. Chinnam *et al.* [16] used neural networks to model the degradation of drill bits and evaluate their reliability characteristics. In this work, we focus on modeling bearing degradation. Bearing are widespread in all rotational equipment. Consequently, a large percentage of maintenance activities require inspection and replacement of bearings. In numerous applications, bearing failure can be catastrophic. For example, failure of the primary bearing of a helicopter rotor can have fatal consequences.

Vibration monitoring has been widely used to monitor the condition of bearings. Alguindigue *et al.* [17] developed a neural network model that identifies the type of bearing defect and classifies the bearing's condition using vibration signal. Shao and Nezu [9] presented two neural network approaches to classify the location of a bearing's defect. In [18], the authors developed a fuzzy expert system to detect roller bearing defects using high-frequency vibration information. Gebraeel *et al.* [23] presented several neural network models for predicting bearing failure time.

Statistical and analytical approaches have also been used to classify bearing health [19], [20]. Some analytical techniques model crack initiation and propagation for predicting bearing life [21]. Li *et al.* [21] developed an adaptive least squares algorithm to model the rate of defect growth in bearings. The authors established a relationship between the instantaneous defect area based and the amplitude of the associated vibration and acoustic emission signals.

The methodology presented in this paper differs from previous approaches in many aspects. Whereas some previous work relies on artificially induced defects [21], the proposed model captures the natural evolution of the degradation process without artificially induced defects. Furthermore, we focus on estimating failure time distributions and deriving remaining life distributions [9], [16], [23]. Unlike previous work that relies on fixed failure time distributions, we utilize online signals to continuously update the remaining life distributions of partially degraded components in a Bayesian manner. In addition, the benefit of using Neural Networks in this paper is their ability to model complex signals. For example, Gebraeel *et al.* [24] extract the bearing specific frequency amplitudes and aggregate them to form a degradation signal. In other words, each point on the degradation signal corresponds to a vibration spectrum. The degradation signal is then modeled as a stochastic process. However, this work does not perform such spectral aggregation, thus, allowing the utilization of the rich content available in the individual frequencies.

III. Degradation Modeling

The ability to accurately predict a component's remaining life is arguably the most challenging component of degradation modeling. This is primarily due to the randomness inherent in most degradation processes. To account for this dispersion, we study the degradation process of a sample of identical components, specifically, thrust ball bearings. Bearings can be monitored using vibration signals and are a good source for real-world degradation-based sensory information. In addition, the low-cost of test bearings enables high-volume destructive testing, thus, ensuring fidelity of our validation experiments.

A. Bearing Degradation and Failure

Bearings typically fail due to fatigue stresses. As the bearing progresses in its service life, minute subsurface cracks begin to form in the bearing's raceways. These cracks propagate to the surface and dislodge a piece of the raceway material resulting in surface spalls [22]. The repetitive passage of the rolling elements over these spalls results in the excitation of distinctive defective vibration frequencies. The defective frequencies are a function of rotational speed, spall location, and bearing geometry. The amplitude of these frequencies increases as the bearing degrades. Advanced stages of bearing degradation are characterized by the appearance of harmonic frequencies (integer multiples of the defective frequency) in the vibration spectrum.

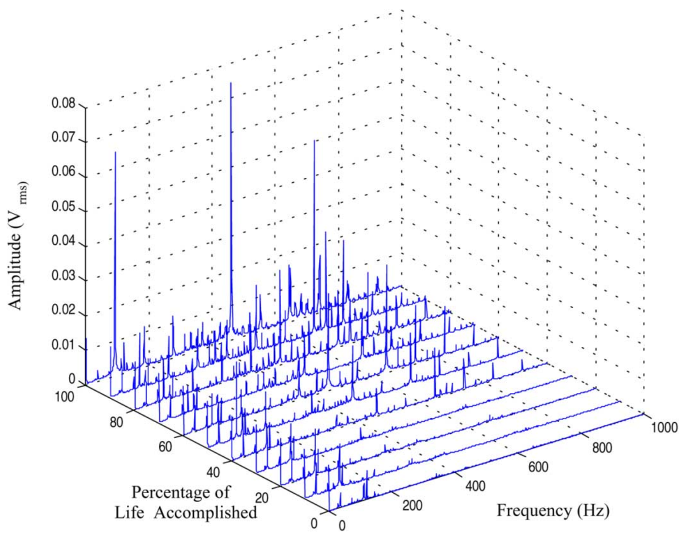

Fig. 1. Evolution of the vibration spectra of a degrading bearing over its service life.

Fig. 1 illustrates how bearing vibration spectra evolve with the degradation process. The characteristic increase in frequency amplitudes that accompany degradation can be used to develop bearing-specific degradation signals.

The degradation signal developed in this work consists of the average amplitude of the defective frequency and its first six harmonics (Fig. 2). It is similar to the degradation signal developed in [23]. Each point on the degradation signal is associated with a vibration spectrum from which the average amplitude is evaluated. The degradation signal is composed of two distinct parts corresponding to normal and defective bearing operation. The flat part of the degradation signal corresponds to normal bearing operation. It extends from the point of bearing installation until the first time a spall is formed. The formation of a spall is accompanied by a spike in vibration amplitude of the degradation signal. The time corresponding to this spike is defined as the defect time (Fig. 2). The second part begins at the defect time and is characterized by a randomly fluctuating signal with an increasing trend. It corresponds to an observable, yet acceptable, level of bearing degradation. This phase extends until the signal reaches a predetermined failure threshold, which is defined based acceptable industrial vibration standards for the application in question.

Since bearing degradation is a time consuming process, investigating the natural failure of a sample of bearings presents a challenging task. To overcome this problem, an experimental setup is designed to perform accelerated bearing life tests. In accelerated testing a component is subjected to loading conditions beyond its design specifications in order to accelerate failure and reduce test durations [7].

Degradation signals from different components evolve differently. In some applications it may be appropriate to model a given degradation measure as a linear function of time. For example, Christer and Wang [25] modeled the wear of brake pads using a linear model where the thickness decreased linearly as a function of time. In other applications where cumulative damage accelerates the degradation rate, the exponential functional form has been widely used to model degradation processes. For example, Yang and Jeang [8] used an exponential functional form to study the flank wear of cutting tools. Shao and Nezu [9] used an exponential functional form to model the RMS vibration level of bearings. Gebraeel *et al.* [23], [24] also used the exponential functional form to model bearing degradation. Based on experimental data, the exponential functional was observed to have a relatively accurate representation of bearing wear (Fig. 2).

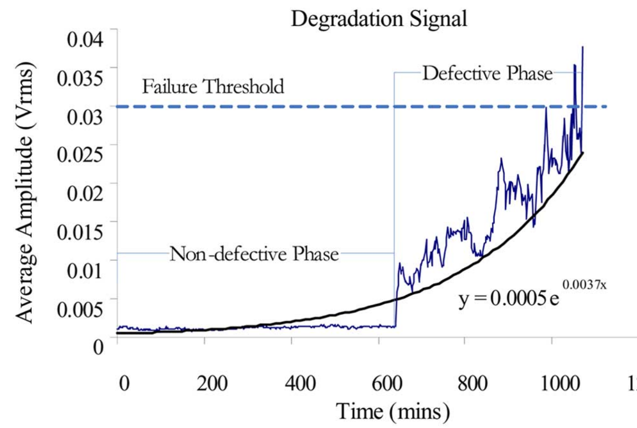

Fig. 2. Vibration based degradation signal.

B. Experimental Setup

Bearing degradation is time-consuming. Consequently, investigating the natural failure of several test bearings presents a challenging task. In the past, several researchers have resorted to introducing artificial defects as one way to accelerate the failure process [21], [29]. Unfortunately, artificially induced defects may interfere with the natural process that takes place during bearing degradation. To overcome this problem, an experimental setup is designed to perform accelerated bearing life tests on bearings. An accelerated testing scheme is designed to reduce the duration of failure tests by facilitating bearing failure [7].

Fig. 3 is a schematic diagram of the setup used for accelerated degradation testing of thrust ball bearings. In accelerated testing, components are subjected to loading conditions that are beyond their design specifications in order to accelerate their failure, thus reducing the duration of the tests [7]. A test bearing is placed in the testing chamber where its lower race is fixed to a stationary housing and its upper race is fastened to a rotating shaft. The setup is equipped with two closed-loop control systems, one for controlling the pressure of the cylinder used to apply the load on the test bearings and another for controlling the rotational speed. However, the degradation data used in this study are obtained from bearings that were tested under constant operating conditions, a load of 200 lbs and a rotational speed of 2200 rpm. The testing chamber is connected to a lubricating system. Each test bearing is immersed in an oil bath to provide continuous lubrication and cooling.

C. Accelerated Degradation Testing

To perform accelerated testing, each bearing is loaded beyond its designated rating in order to increase the stress at the point

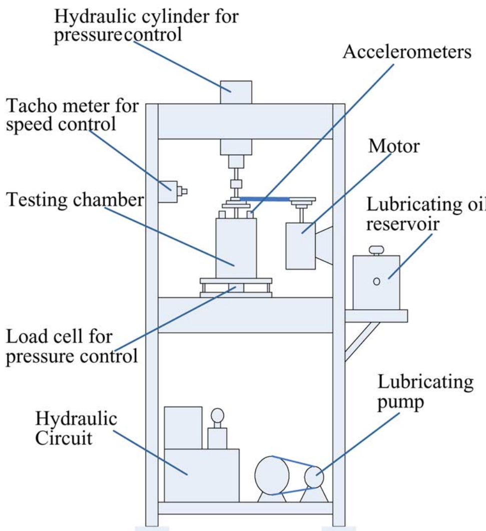

Fig. 3. Schematic diagram of the bearing testing setup.

of contact between the ball and the race. Increasing the contact stress can be accomplished either by increasing the axial load, or by reducing the number of rolling elements. The test bearings used for the purpose of this research each consists of ten rolling elements in the form of steel balls. Before a test bearing is installed in the setup, every other ball is removed from the cage. Consequently, the load on the bearing will be distributed evenly over the five remaining balls. This can be used to increase the contact stress at the point of contact between the ball and the surface of the raceway.

The maximum contact stress between the ball and the raceway is given by the following equations [22], [23]:

$$\text{Max } \sigma_c = 0.198 \sqrt[3]{\frac{P}{k_D^2 \times C_E^2}} \tag{1}$$

where $\text{Max } \sigma_c$ is the maximum contact stress between ball and bearing raceway in kips, P is the load per ball (lbs), K_D is the ball $\text{diameter} = 0.156''$, and C_E is given by

$$C_E = \frac{1-\gamma_1^2}{E_1} + \frac{1-\gamma_2^2}{E_2} \tag{2}$$

where $\gamma_1 \cdot \gamma_2$ are the Poisson's ratios of the ball (rolling element) material and the raceway material. Both materials are the same and have a Poisson $\text{ratio} = 0.3$. E_1 and E_2 are the moduli of elasticity of the ball and raceway materials. The twp materials are also the same and the modulus of $\text{elasticity} = 30{,}000$ psi. Using these values, we get $C_E = 6.07E - 08$.

Next, we compare the maximum allowable contact stress between the bearing's rolling elements and the raceway, and the contact stress induced by the bearing testing experimental setup.

1) Case I: Maximum Dynamic Rated Stress: The test bearings have a manufacturer rating of 365 lbs. Using the manufacturer's rating, the maximum load per ball for a bearing with ten balls will be $P = 36.5$ lbs. Substituting the value of P, C_E, and K_D in (1), we get $\text{Max } \sigma_c = 680$ kips.

2) Case II: Accelerated Stress: In the accelerated test, the bearing is configured with five balls and an axial load of 200 lbs. The accelerated load per ball is $P = 200/5 = 40$ lbs

By substituting the value of P, C_E, and K_D in (1), we get $\text{Max } \sigma_c = 702$ kips.

The use of five balls with 200 lbs increases the maximum contact stress beyond the manufacturer's load rating, thus, accelerating bearing degradation and failure.

During each bearing test, bearing vibration spectra are continuously acquired (every 2 min) using an accelerometer and a data acquisition system designed in Labview. Sensory signals are first passed through analog filters to remove any signal frequency components beyond the range of the ADC (analog to digital converter). A digital anti-aliasing filter removes frequency components greater than half the programmed sampling rate. The data is displayed first as a time-based waveform and converted to the frequency domain. Vibration spectra and degradation signals acquired from a set of bearings are used to build a degradation database, which is then used to train our neural network model.

Bearing defective frequencies can be computed using deterministic equations found in [22]. For the special case involving thrust bearings, the equations can be expressed as follows.

The fundamental defective frequency (corresponding to a raceway defect)

$$\text{BPF} = \frac{1}{2} \times z \times \frac{\text{RPM}}{60} = 92 \text{ Hz}.$$

Cage rotating frequency

$$\text{FTF} = \frac{1}{2} \times \frac{\text{RPM}}{60} = 18 \text{ Hz}.$$

Rotational frequency of the rolling elements

$$\text{BSF} = \frac{1}{2} \times \frac{\text{RPM}}{60} \times \frac{d_c}{d_r} = 96 \text{ Hz}$$

where d_r = diameter of the rolling elements, d_c = diameter of the cage = $(d_{\text{out}} + d_{\text{in}})/2$, and z = number of rolling elements.

According to industrial standards for machine vibration, ISO 2372 [26], an overall root mean square (rms) vibration acceleration level ranging between 2.0 and 2.2 Gs is considered a "danger level" for applications involving "medium-sized general-purpose" machinery. Each degradation test is terminated once the root mean square of the overall vibration acceleration reaches a threshold of 2.2 Gs. The corresponding amplitude of the degradation signal was found to range between 0.03 and 0.035 V_{rms}. We define the degradation signal amplitude of 0.03 V_{rms} as our failure threshold. Note that in high precision machinery, the formation of a spall may be considered as failure in which case this methodology will not be suitable.

IV. Neural Network Modeling

Neural networks are data processing systems that consist of a number of interconnected processing elements called neurons [27]. There are several types of neural networks. The neural network models developed in this work belongs to the class of neural networks, where the neurons are organized in a sequence of layers namely a single input layer, one or more hidden layers, and an output layer. Neural networks can be used to establish a complex regression function between a set of network inputs and outputs. This is achieved through a network training

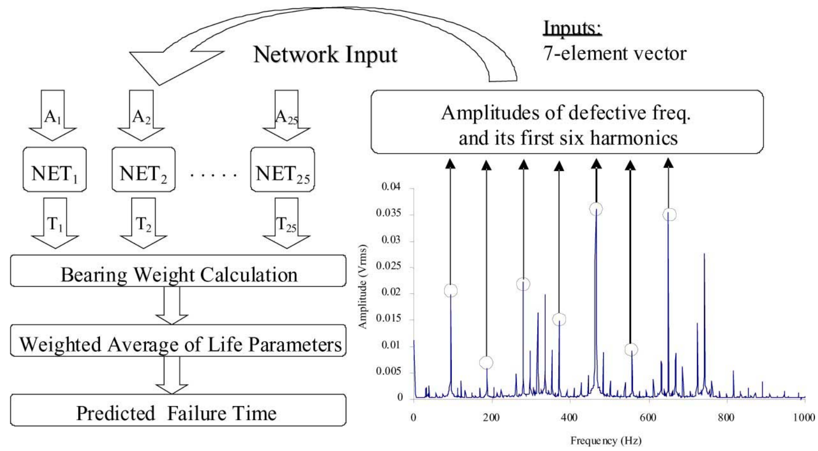

Fig. 4. Configuration of the neural network model used to predict bearing failure times.

procedure. There are two main types of training methodologies; 1) supervised training where the network is trained using a specified sequence of inputs and outputs and (2) unsupervised training where the primary function of the network is to classify network inputs.

The neural network model developed in this work consists of a group of feed-forward back-propagation neural networks that undergo supervised training. Each network is used to model the degradation process of a single bearing using the history of its vibration signals. The neural networks are trained to establish a relationship between the time a bearing has been in service and the corresponding vibration spectrum. The number of neural networks M is set to equal the number of bearings used to develop the degradation database ($M = 25$). The training process is limited to the second phase of the degradation process that extends from the defect time until the failure time.

In other words, the neural network models will be used to predict bearing failure times after a spall has occurred.

Each neural network FN_i is trained to identify the evolution of the vibration spectrum of an individual bearing B_i. Input to the neural network consists of a seven-element vector of frequency amplitudes $A_{\rm in} = [a_{i1}^n, a_{i2}^n, a_{i3}^n, \ldots, a_{i7}^n]$. The first element of the input vector corresponds to the amplitude of the defective frequency ($\mathrm{BPF} = 92$ Hz) and the remaining six elements correspond to amplitudes of the first six harmonics of the defective frequency. These frequencies and their corresponding amplitudes are extracted from each vibration spectrum acquired during each sampling epoch. We define the sampling epoch n as the time at which a data acquisition is performed, where $n = 1, \ldots, N_i$, and N_i is the last sampling epoch before bearing failure. The output of each neural network is a single-element vector $[t_1^n]$ that represents how long bearing B_i has been operating after the first defect had occurred (after the defect time). During training, the following sequence $\{[A_{i1}] \ldots [A_{\rm in}] \ldots [A_{iN_1}]\}$ is input to each FN_i. The corresponding output sequence is given by $\left\{[t_i^1] \ldots [t_i^n] \ldots \left[t_i^{N_i}\right]\right\}$, where $t_i^{N_i}$ is B_i's failure time. The outcome of the training process is a set of 25 trained neural networks, each network trained on the degradation characteristics of an individual bearing.

Next, we consider a bearing B_k operating in the field with an amplitude vector $[A_{kn}]$ at a given sampling epoch n. The amplitude vector is input to each of the 25 trained neural networks, as illustrated in Fig. 4. Since each network FN_i is trained using the degradation information of different bearings, the operating times output by each neural networks are different. The differences between the estimated operating times and the actual bearing operating time (assumed to be known) are used to compute a sequence of squared error values. These error values are used to derive a corresponding sequence of normalized weights, which will be utilized to predict bearing failure times.

A. Computing Network Weights

The operating times $\{[t_1^n], \ldots, [t_{25}^n]\}$ estimated by the neural networks are used to compute squared error values using the following expression:

$$e_{ik}^n = (t_i^n - p_k^n)^2 \tag{3}$$

where e_{ik}^n is the squared error between the operating time t_i^n (output from network, i, at epoch n) and the actual operating time p_k^n of bearing k.

Each neural network is associated with a squared error value from the vector of 25 squared errors $[e_{1k}^n, \ldots, e_{25k}^n]$. The error values are then used to compute corresponding weights associated with each neural network. The weights are computed using the following expression:

$$W_i^n = K\,(e_{ik}^n)^{-1} \tag{4}$$

where, W_i^n is the weight associated with FN_i at epoch n, and $K = \sum_{i=1}^{25} 1/(e_{ik})^2$.

Note that expression (2) implies that higher weights correspond to smaller error values. For example, if we have two error values $e_{1k}^n < e_{2k}^n$, this translates to the fact that bearing operating time estimated by FN_1 is closer to the actual operating time p_k^n compared with operating time estimated by FN_2. The Section V describes how these weights are sued to predict failure times.

V. Computing Failure Time Distributions

Bearing degradation characteristics are captured by the degradation signal. The amplitude of the degradation signal rises until it reaches a predetermined failure threshold at which point the bearing is said to have failed. Based on our experimental observations, the degradation signal is modeled using an exponential functional form (Fig. 2). Exponential models have often been used to model fatigue and wear processes, especially those associated with bearings (see [9, 23]). The following outlines the necessary steps required to compute failure time distributions.

Step 1) Each degradation signal of the training bearings is fitted with an exponential trend of the form $\alpha e^{\beta t}$. This results in two vectors of exponential parameters $\langle \alpha_1, \ldots, \alpha_{25} \rangle$ and $\langle \beta_1, \ldots, \beta_{25} \rangle$ that capture the degradation characteristics associated with the training bearings.

Step 2) The vector of weights derived using expression (2) is then used to compute a weighted average of the exponential parameters $\langle \alpha_1, \ldots, \alpha_{25} \rangle$ and $\langle \beta_1, \ldots, \beta_{25} \rangle$, as illustrated by expressions (5) and (6)

$$\alpha_k^n = \sum_{i=1}^{25} W_i^n \alpha_i \tag{5}$$

$$\beta_k^n = \sum_{i=1}^{25} W_i^n \beta_i \tag{6}$$

where W_i^n is the weight calculated using the output of neural network FN_i, and α_i and β_i are the exponential parameters of the corresponding degradation model.

Step 3) The exponential parameters α_k^n and β_k^n are used to predict B_k's failure time using expression (7)

$$t_k^{Nk} = \left(\frac{1}{\beta_k^n}\right) \times \ln\left(\frac{D}{\alpha_k^n}\right) \tag{7}$$

where $D \sim$ is the failure threshold.

The underlying assumption of this methodology is that bearings with similar vibration characteristics at given operating times are likely to have similar failure times.

Step 4) Repeat steps 2 and 3 at successive sampling epochs until an initial sample of predicted failure times is collected. The size of the sample may differ according to the application. The sample of predicted failure times is used to compute a prior failure time distribution for the bearing being monitored.

Next, we present a sensory updating methodology where the prior distribution of the component's failure time is updated using subsequent condition-based sensory information.

A. Sensory Updating Methodology

This section describes the development of a sensory updating methodology that utilizes the failure times predicted by the neural network model to compute and update residual life distributions for partially degraded components. These distributions are important because they capture the uncertainty associated with component degradation.

The updating procedure is performed using Bayesian techniques. Bayesian inference combines prior knowledge of a distribution of a random parameter θ with observed sample data x in order to derive a posterior distribution, which reflects the updated beliefs given a sample observation. The posterior distribution $\pi(\theta|x)$ is the distribution of θ given the sample observation x and is given as $\pi(\theta|x) = h(x,\theta)/m(x)$, and $h(x,\theta) = \pi(\theta) f(x|\theta)$ is the joint density of x and θ. In this work, an empirical Bayes approach is used to compute and update the residual life distributions using real-time vibration signals. Empirical Bayes is a Bayesian updating technique, where the prior knowledge is derived from observed sample data.

Consider a bearing B_k, we assume that the failure times predicted by the neural network model at each sampling epoch n are random picks from independent normal distributions $N\left(\theta_n^k, \sigma_B^2\right)$, where σ_B^2 is assumed to be known and can be estimated using the failure times of the 25 training bearings. If an initial sample of predicted failure times was observed, it could be argued that the average value of this sample can be used to estimate the true failure time. However, this can be a risky assumption since the bearing failure time varies with progressive degradation. On the other hand, we can assume that the true values of the bearing's failure time at each sampling epoch are from a common prior distribution $N\left(\mu_\theta^k, \sigma_\theta^2\right)$, where the hyperparameters μ_θ^k and σ_θ^2 can be estimated from the initial sample of predicted bearing failure times using expressions (8) and (9) [28]

$$\mu_\theta^k = \overline{x} \tag{8}$$

$$\sigma_\theta^2 = \max\left\{0, \frac{1}{p}s^2 - \sigma_B^2\right\} \tag{9}$$

where p corresponds to the number of observed failure times used to compute the prior distribution and $s^2 = \sum_{n=1}^{10}(x_n - \overline{x})^2$.

We are interested in finding the posterior (updated) distribution of θ_n^k given a new failure time observation x_i. The posterior distribution is expressed as $N\left(\mu_n^{EB}(x_n), V_n^{EB}\right)$, where μ_n^{EB} and V_n^{EB} are evaluated at each sampling epoch i using expressions (10) and (11), respectively [28]

$$\mu_n^{EB}(x_n) = x_n - \hat{B}\left(x_n - \mu_\theta^k\right) \tag{10}$$

$$V_n^{EB}(x_n) = \sigma_B^2\left(1 - \frac{(p-1)}{p}\hat{B}\right) + \frac{2}{(p-3)}\hat{B}^2\left(x_n - \mu_\theta^k\right)^2 \tag{11}$$

where $\hat{B} = (p - 3/p - 1)\sigma_B^2/\left(\sigma_B^2 + \sigma_\theta^2\right)$.

Given the actual cumulative service time t_c, we can compute the posterior residual life distribution of the bearing that is being monitored using expression (12)

$$N\left(\mu_n^{EB}(x_n) - t_c, V_n^{EB}\right). \tag{12}$$

The posterior residual life distribution also follows a normal distribution, thus it is necessary to preclude negative values of residual life. As a result, the distribution of the residual life is truncated at zero. The cdf of the residual life T_R can be expressed as

$$p\left\{T_R \leq t|\mu_\theta^k, \quad \sigma_\theta^2, T_R \geq 0\right\} = \frac{\Phi(g(t)) - \Phi(g(0))}{1 - \Phi(g(0))} \tag{13}$$

where $g(t) = \left(\mu_n^{EB}(x_n) - t_c - t/\sqrt{V_n^{EB}}\right)$ and $\Phi(g)$ is the cdf of a standard normal.

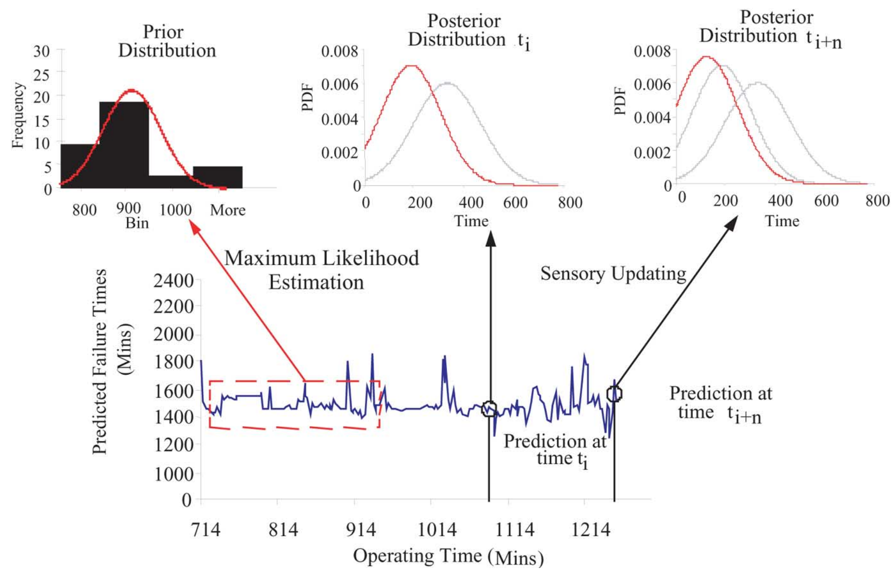

Fig. 5. Empirical Bayes methodology for sensory updating methodology.

Fig. 5 illustrates the methodology for computing and updating residual life distributions. First, an initial sequence of consecutive vibration spectra is input to the neural network model and used to estimate a corresponding sequence of predicted bearing failure times. The initial sequence of predicted failure times is used to estimate the hyperparameters μ_θ^k and σ_θ^2, and derive a prior failure time distribution. A subsequent vibration spectrum acquired, for example, at time t_i is then input to the neural network model. The resulting predicted failure time estimated by the neural network model is then used to update the prior distribution (Fig. 5). Successive updating of the residual life distributions can be calculated in a similar fashion using newly observed failure times (e.g., t_{i+n}).

Since the observed failure time values are computed using real-time vibration signals, they are correlated with the degradation state of the bearing being monitored. Thus, the updated residual life distributions capture the actual degradation state of the bearing.

VI. Model Implementation Using Bearing Degradation Information

The neural network-based degradation model developed in this paper is validated using a different set of 25 validation bearings (bearings 26–50). We model the degradation phase of each of the validation bearings. Recall the degradation phase begins once the first spall has occurred and ends when the bearing fails. During each test, vibration spectra are continuously acquired and used to compute and update the residual life distribution of the validation bearing that is being monitored. Once the bearing fails, an average prediction error is computed at several degradation percentiles. The degradation percentiles are computed by dividing the degradation phase (identified once the bearing fails) into ten equal time stamps such that the defect time corresponds to the zeroth degradation percentile, whereas the actual failure time corresponds to 100th degradation percentile.

In validation bearing #32, the first spall was observed after 516 min (in accelerated time) of service. The onset of spalling is referred to as the first sampling epoch, which represents the first valid bearing failure time prediction. Recall that the neural network model is designed specifically to predict bearing failure times for partially degraded bearings. The third column of [1] consists of the predicted failure times. In this work, we choose the first ten predicted failure time values to estimate the hyperparameters μ_θ^k and σ_θ^2 of the prior distribution of θ_n^k using expressions (6) and (7). The parameters $\sigma_B^2 = 66497$ is estimated using the actual failure times of the training bearings used to train the neural networks. Subsequent vibration spectra are used to update the prior distributions using the empirical Bayes method discussed earlier. The posterior means μ_n^{EB} and variances V_n^{EB} are presented in columns 4 and 5.

Fig. 6 presents the probability density functions (pdf) of the posterior residual life distributions corresponding to five different degradation percentiles (10%, 30%, 50%, 70%, and 90%). Assuming continuous monitoring, we adopt a conservative prediction policy by retaining the lowest value of μ_n^{EB} and its corresponding V_n^{EB}. These values are used to compute the residual life distribution of a bearing at any sampling epoch. Consequently, successive residual life distributions tend to shift towards the zero as a result of the conservative policy.

Prediction errors that measure the percentage difference between the actual and expected bearing failure times at each sampling epoch are computed using (14)

$$D_k^n = \frac{t^{N_k} - (p_k^n + E_n[\mathrm{RLD}_k])}{t^{N_k}} \tag{14}$$

where D_k^n is the prediction error associated with bearing B_k, computed at the sampling epoch n using the degradation model t^{N_k} is the actual failure time of the bearing k, p_n^k is the current total operating time of the bearing k at sampling epoch n, and $E_n[\mathrm{RLD}_k]$ is the expected value of the residual life distribution of B_k computed at the nth sampling epoch.

A. Results

The performance of the proposed degradation methodology is evaluated by computed prediction errors (expression 14) using implementation for the 25 validation bearings. Reference

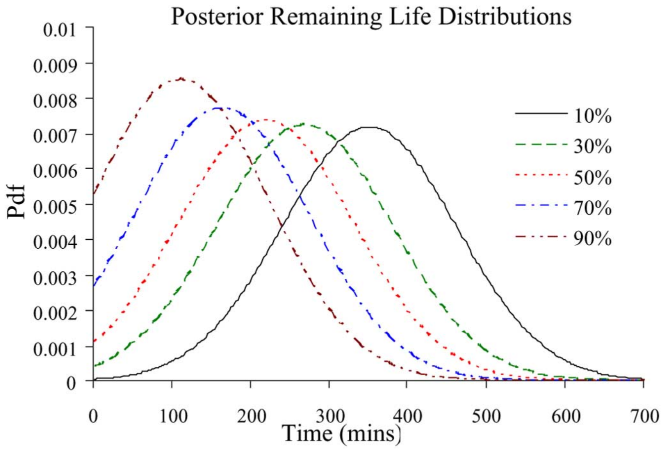

Fig. 6. Residual life distributions at different bearing service life percentiles.

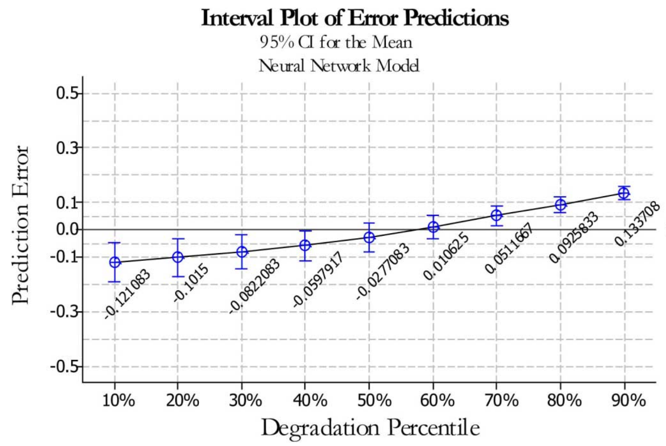

Fig. 7. Prediction errors using the neural network-based degradation model.

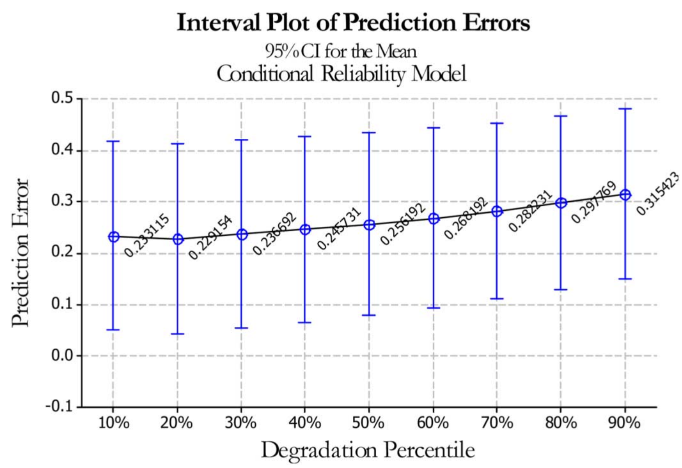

Fig. 8. Prediction errors using the conditional reliability distribution.

[1] presents a graph of mean and confidence interval of the average prediction errors evaluated at the following bearing degradation percentiles, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, and 90%. Initially, the degradation model overestimates the expected value of bearing remaining life resulting in negative average errors. As bearings progresses through their life, the model tends to underestimate the remaining life due to the conservative manner by which μ_i^{EB} is chosen. The mean of the absolute value of prediction error is 7.56% with a standard deviation of 4.1%.

B. Benchmarks

This section describes two benchmarks with which we compare the results of our neural network-based degradation model.

1) Benchmark 1: This policy relies solely on reliability information rather than the degradation characteristics. This benchmark assumes that the failure times follow a Weibull distribution. We chose the Weibull distribution because it had the best fit for the bearing failure time data. The distribution parameters were estimated using available failure times of the 25 training bearings. The scale parameter was estimated as 2.93, while the shape parameter was 746.

At any given time t_i, the remaining life time of a bearing is estimated using the conditional reliability distribution given that the bearing has survived up to time t_i. The prediction error is the difference between the estimated failure time and the actual failure time and is calculated as follows:

$$CR_k^n = \frac{t^{N_k} - (\mathrm{MTTF}(t_i))}{t^{N_k}} \times 100 \tag{15}$$

where CR_k^n is the prediction error associated with bearing B_k computed at time t_i using the conditional reliability policy, t^{N_k} is the actual failure time of the bearing k, $\mathrm{MTTF}(t_i)$ is the residual mean time to failure at time t_i and is defined as

$$\mathrm{MTTF}(t_i) = \frac{1}{R(t_i)} \int_{t_i}^{\infty} R(t+t_i) d(t+t_i). \tag{16}$$

The benchmark was applied to the degradation data associated with the 25 validation bearings. An interval plot of the prediction error (expression 15) is presented in Fig. 8. The average prediction error of using this benchmark is approximately 26.3% with a standard deviation of 3.04%.

C. Benchmark 2

This benchmark utilizes a degradation methodology proposed by Lu and Meeker [12] in which the authors propose parametric models (linear, exponential, etc.) to describe the path of the degradation signal. The sample path of unit i at time t_j is given by

$$y_{ij} = \eta(t_j, \Phi, \Theta) + \varepsilon_{ij} \quad i = 1, 2, \ldots, n \tag{17}$$

where t_j is the time of the jth measurement (sampling epoch), $\varepsilon_{ij} \approx N\left(0, \sigma_\varepsilon^2\right)$ is a measurement error with a constant variance σ_ε^2, η_{ij} is the actual degradation path of the ith unit at time t_j, Φ is a vector of parameters that represents the fixed effects associated with degradation, Θ_i is a vector of parameters representing the random effects associated with the degradation process of the ith unit. The authors assume that Θ_i $(i = 1, 2, \ldots, n)$ follow a multivariate distribution $G_\theta(\cdot)$.

This benchmark utilizes the fact that the evolution of bearing degradation signals can be reasonably represented using an exponential functional form. Consequently, the degradation path is

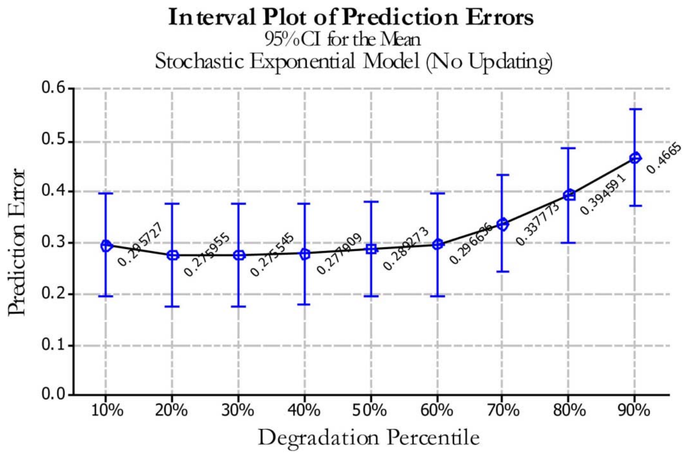

Fig. 9. Failure time prediction errors using a parametric exponential degradation model.

defined as $\eta(t) = \phi + \theta e^{\beta t}$. For convenience, we work with the natural log of the exponential model and arrive at transformed linear degradation model

$$L_{ij} = \ln(y_{ij}) = \ln\phi + \ln\theta + \beta t_{ij} \tag{18}$$

where L_{ij} is the logged signal, $\ln\phi$ is the constant effect parameter, $\ln\theta$ and β are random effects parameters.

In this benchmark, $\ln\phi$ represents the amplitude of the degradation signal associated with normal bearing operation. This phase of the degradation signal is characterized by low amplitude and is generally constant and can be neglected for this application. The random parameters $\theta' = \ln\theta$ and β follow normal distributions with the following parameters $\theta' \sim N\left(\mu_\theta, \sigma_\theta^2\right)$ and $\beta \sim N\left(\mu_\beta, \sigma_\beta^2\right)$, where $P\{\theta' < 0\}$ and $P\{\beta < 0\}$ are negligible. The distribution of the degradation signal is expressed as follows:

$$\psi(t) = \ln(\eta(t)) \sim N\left(\mu_{\theta'} + \mu_\beta t, \sigma_{\theta'}^2 + \sigma_\beta^2 t^2\right). \tag{19}$$

The parameters of this distribution are estimated using sample values of the exponential parameters associated with the training bearings.

The proportion of components failing by time t is expressed as $F_T(t) = P\{T \geq t\}$. This is equivalent to saying that the distribution of the failure time T is equal to the probability of the degradation signal exceeding a failure threshold D. This probability can be approximated by the following expression:

$$F_T(t) \approx \Phi\left(\frac{\frac{t - [D - \mu_{\theta'}]}{\mu_\beta}}{\sqrt{\frac{\left[\sigma_{\theta'}^2 + \sigma_\beta^2 t^2\right]}{\mu_\beta^2}}}\right). \tag{20}$$

The distribution of the residual life is evaluated at any given time t_k, by conditioning on the survival up to time t_k. Fig. 9 presents an interval plot of the prediction error between the failure times estimated using the Lu and Meeker's degradation model and the actual failure times of the validation bearings. The average prediction error across all the degradation percentiles is approximately 32.3% with a standard deviation of 6.6%.

TABLE I
PRESENTS RESULTS OF SENSORY UPDATING OF ONE OF THE VALIDATION BEARINGS, $\mu_\theta^k = \overline{x} = 916.3$, $s^2 = 8085.5$, AND $\sigma_\theta^2 = 0$ (ACTUAL BEARING $\mathrm{Failure\ Time} = 842\ \mathrm{min}$)

Sampling Epoch	Service Time	Predicted Failure	${}^{EB}_{i}\ x$	$V_i^{EB}\ x$	Prediction Error (%)
1	516	954			-16.1
2	518	956			-16.3
3	520	955			-16.2
4	522	935			-13.7
5	524	895			-8.9
6	526	896			-8.9
7	528	884			-7.6
8	530	898			-9.2
9	532	907			-10.3
10	534	886			-7.8
11	536	899	912.79	12362.9	-9.4
12	538	917	916.72	12363.8	-11.6
13	540	810	892.99	12366.6	1.4
...	...	...	...	...	...
146	764	778	903.41	12974.7	-4.3
147	766	780	903.42	12973.4	-4.3
148	768	782	909.32	12549.7	-7.5
149	770	787	906.36	12731.7	-5.9
150	772	794	906.59	12715.8	-6.0
151	774	802	909.84	12524.2	-7.8
152	776	811	910.39	12499.0	-8.1
153	778	821	900.64	13258.1	-2.8
154	780	829	892.81	14349.1	1.5

Unlike the neural network-based degradation model proposed in this paper, the two benchmarks do not incorporate any sensory updating. The results indicate that both benchmark policies overestimate the remaining life of the bearings. We claim that this bias is possibly a result of the large variance in the failure times, which in turn results in large values of $\sigma_{\theta'}^2$ and σ_β^2.

VII. CONCLUSION

This work develops a neural network-based degradation model for computing and updating residual life distributions of partially degraded components. The novelty of this methodology lies in the ability to update a component's remaining life distribution using *in situ* condition-based sensory signals. The real-time sensory signals capture the latest degradation state of the component and the resulting updated distributions are directly linked to the physical degradation state of the component.

The model is tested and validated using thrust ball bearings as the testbed component. Vibration signals resulting from bearing degradation are used to estimate prior bearing failure time distributions using a neural network model. Subsequent signals are then used to update the prior distribution using a Bayesian approach and compute posterior residual life distributions. The performance of the model was evaluated by comparing the expected value of the predicted failure time and the actual failure time of a set of validation bearings. Results of the prediction errors were compared with results from two benchmark policies. The first benchmark utilized conditional reliability, while the second benchmark utilized a degradation model developed by Lu and Meeker [12]. The proposed methodology had a mean prediction error of 7.56% compared with the first and second benchmarks that had prediction errors 26.3% and 32.3%, respectively.

Although the model was tested rolling element thrust bearings, it is not limited to this type of component. Further research related to the scalability of this methodology is currently underway. In general, the methodology requires: 1) identifying physical phenomena that best capture the physical transitions associated with degradation processes and the operating conditions; 2) investigating appropriate condition monitoring technology to observe the evolution of these phenomena; 3) identifying characteristic patterns in the sensory condition monitoring signals; 4) developing degradation signals that evolve—increase or decrease—with the component's degradation; 5) training the neural network to model; and 6) development of optimal stopping rules to identify an optimal/economic time to stop updating.

This work establishes a linkage low-level condition monitoring data and residual life distributions of bearings. The bearings were tested under constant load and speed. Further research is required to study the effect of changing the operating conditions and incorporating additional input parameters that capture operating conditions, such as load and speed. Further research is also needed to establish and optimal frequency of updating and when to stop updating.

REFERENCES

[1] R. Smith and B. Hawkins, *Lean Maintenance: Reduce Costs, Improve Quality, and Increase Market Share*. Burlington, MA: Elsevier Butterworth-Heinemann, 2004.

[2] R. K. Mobley, *An Introduction to Predictive Maintenance*. Burlington, MA: Elsevier Butterworth-Heinemann, 2002.

[3] R. Dekker, "Applications of maintenance optimization models: A review and analysis," *Reliability Eng. Syst. Safety*, vol. 51, pp. 229–240, 1996.

[4] P. A. Scarf, "On the application of mathematical models in maintenance," *Eur. J. Oper. Res.*, vol. 99, no. 3, pp. 493–506, 1997.

[5] W. Q. Meeker, L. A. Escobar, and J. C. Lu, "Accelerated degradation tests: Modeling and analysis," *Technometrics*, vol. 40, no. 2, pp. 89–99, 1998.

[6] C. Booth and J. R. McDonald, "The use of artificial neural networks for condition monitoring of electrical power transformers," *Neurocomputing*, vol. 23, pp. 97–109, 1998.

[7] W. Nelson, *Accelerated Testing Statistical Models, Test Plans, and Data Analysis*. New York: Wiley, 1990.

[8] K. Yang and A. Jeang, "Statistical surface roughness checking procedure based on a cutting tool wear model," *J. Manuf. Syst.*, vol. 13, no. 1, pp. 1–8, 1994.

[9] Y. Shao and K. Nezu, "Prognosis of remaining bearing life using neural networks," *Proc. Inst. Mech. Eng., Part I, J. Syst. Control Eng.*, vol. 214, no. 3, pp. 217–230, 2000.

[10] S. Tseng, M. Hamada, and C. Chiao, "Using degradation data to improve fluorescent lamp reliability," *J. Quality Technol.*, vol. 27, no. 4, pp. 363–369, 1995.

[11] K. Yang and G. Yang, "Degradation reliability assessment using severe critical values," *Int. J. Reliability, Quality, Safety Eng.*, vol. 5, no. 1, pp. 85–95, 1998.

[12] C. Lu and W. Meeker, "Using degradation measures to estimate a time-to-failure distribution," *Technometrics*, vol. 35, no. 2, pp. 161–174, 1993.

[13] D. C. Swanson, "A general prognostic tracking algorithm for predictive maintenance," in *Proc. IEEE Aerosp. Conf.*, 2001, vol. 6, pp. 62971–62977.

[14] A. K. S. Jardine and T. Joseph, "Optimizing condition-based maintenance decisions for equipment subject to vibration monitoring," *J. Quality Maintenance Eng.*, vol. 5, no. 3, pp. 192–202, 1999.

[15] J. Lee, "Measurement of machine performance degradation using a neural network model," *Computers in Industry*, vol. 30, no. 3, pp. 193–209, 1996.

[16] R. B. Chinnam and P. Mohan, "Online reliability estimation of physical systems using neural networks and wavelets," *Int. J. Smart Eng. Syst. Des.*, vol. 4, no. 4, pp. 253–264, 2002.

[17] I. E. Alguindigue, A. Loskiewicz-Buczak, and R. E. Uhrig, "Monitoring and diagnosis of rolling element bearings using artificial neural networks," *IEEE Trans. Ind. Electron.*, vol. 40, pp. 209–217, 1993.

[18] H. H. Huang and H. P. B. Wang, "Integrated monitoring and diagnostic system for roller bearings," *Int. J. Adv. Manuf. Technol.*, vol. 12, no. 1, pp. 37–46, 1996.

[19] J. Shiroishi, Y. Li, T. Kurfess, and S. Danyluk, "Bearing condition diagnostics via vibration and acoustic emission measurements," *Mech. Syst. Signal Process.*, vol. 11, no. 5, pp. 693–705, 1997.

[20] Y. T. Su and Y. T. Sheen, "Signature analysis of roller bearing vibrations: Lubrication effects," *Proc. Instit. Mech. Eng.*, vol. 206, pp. 193–202, 1992.

[21] Y. Li, S. Billington, and C. Zhang, "Dynamic prognostic prediction of defect propagation on rolling element bearings," *Lubrication Eng.*, vol. 42, no. 2, pp. 385–392, 1999.

[22] T. A. Harris, *Rolling Bearing Analysis*. New York: Wiley, 2001.

[23] N. Gebraeel, M. Lawley, R. Liu, and V. Parmeshwaran, "Life distributions from component degradation signals: A neural net approach," *IEEE Trans. Ind. Electron.*, vol. 51, pp. 694–700, 2004.

[24] N. Gebraeel, M. Lawley, R. Li, and J. K. Ryan, "Life distributions from component degradation signals: A Bayesian approach," *IIE Trans. Reliability*, vol. 37, no. 6, pp. 543–557, 2005.

[25] A. H. Christer and W. Wang, "A model of condition monitoring of a production plant," *Int. J. Prod. Res.*, vol. 30, no. 9, pp. 2199–2211, 1992.

[26] M. P. Blake and W. S. Mitchel, *Vibration & Acoustic Measurement Handbook*. New York: Spartan Books, 1972.

[27] L. H. Tsoukalas and R. E. Uhrig, *Fuzzy and Neural Approaches in Engineering*. New York: Wiley, 1997.

[28] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*. New York: Springer-Verlag, 1985.

[29] Y. F. Wang and P. J. Kootsookos, "Modeling of low shaft speed bearing fault for condition monitoring," *Mech. Syst. Signal Process.*, vol. 12, no. 3, pp. 415–426, 1998.

[30] C. Cempel and A. Ziolkowski, "A new method of symptom based processing for machinery condition recognition and forecasting," *Mech. Syst. Signal Process.*, vol. 9, no. 2, pp. 129–137, 1995.

[31] C. Cempel, H. G. Natke, and J. T. P. Yao, "Symptom reliability and hazard for systems condition monitoring," *Mech. Syst. Signal Process.*, vol. 14, no. 3, pp. 495–505, 2000.

Nagi Z. Gebraeel received the B.S. degree in industrial engineering from the University of Alexandria, Alexandria, Egypt, in 1995, and the M.S. and Ph.D. degrees in industrial engineering from Purdue University, West Lafayette, IN, in 1998 and 2003, respectively.

Currently, he is an Assistant Professor of Industrial Engineering at the Department of Mechanical and Industrial Engineering, University of Iowa. His research interests are in the areas of degradation modeling, reliability, and manufacturing systems with focus on prognostic health management.

Dr. Gebraeel is a member of the Institute of Industrial Engineers (IIE) and the Institute for Operations Research and the Management Sciences (INFORMS), and the Quality Statistics and Reliability Society.

Mark A. Lawley (M'01) received the Ph.D. degree in mechanical engineering from the University of Illinois at Urbana–Champaign, Urbana, in 1995.

He is an Associate Professor of Biomedical Engineering at Purdue University. He served for two years as an Assistant Professor of Industrial Engineering at the University of Alabama and then for nine years as an Assistant Professor and then Associate Professor in Purdue's School of Industrial Engineering. He has held engineering positions with Westinghouse Electric Corporation and Emerson Electric Company. In these positions, he worked on projects involving analysis, design, and control of large-scale systems, particularly in automated manufacturing. As a researcher in academics, he has authored numerous technical papers in systems analysis and control. Since January 2005, he served as Regenstrief Faculty Scholar for Purdue's Regenstrief Center for Healthcare Engineering. In this position, he has applied his systems perspective and engineering tools to numerous problems in healthcare delivery. His research has been supported by the National Science Foundation, Union Pacific, Consilium Software, General Motors, St. Vincent Ministries, Ascension Health, and the Indiana State Department of Health.

Dr. Lawley has won two Best Research Paper Awards.