CrossMark

A joint particle filter and expectation maximization approach to machine condition prognosis

Jinjiang Wang[1] · Robert X. Gao[2] · Zhuang Yuan[1] · Zhaoyan Fan[3] ·
Laibin Zhang[1]

Abstract This paper presents a probabilistic model based approach for machinery condition prognosis based on particle filter by integrating physical knowledge with in-process measurements into a state space framework to account for uncertainty and nonlinearity in machinery degradation process. One limitation of conventional particle filter is that condition prognosis is performed based on the model with predetermined parameters obtained from simulation studies or lab-controlled tests. Due to the stochastic nature of machinery defect propagation under varying operating conditions, model parameters may vary in practice which causes prediction errors. To address it, an integrated state prediction and parameter estimation framework based on particle filter and expectation-maximization algorithm is formulated and investigated. The model parameters are adaptively estimated based on expectation-maximization algorithm utilizing hidden degradation state and available in-process measurements. Particle filter is then performed on the identified model with estimated parameters following Bayesian inference scheme to improve the robustness and accuracy of machinery condition prognosis. The effectiveness of the developed method is demonstrated through a simulation study and an experimental run-to-failure bearing test in a wind turbine.

✉ Robert X. Gao
Robert.Gao@case.edu

[1] Department of Mechanical and Transportation Engineering, China University of Petroleum, Beijing 102249, China

[2] Department of Mechanical and Aerospace Engineering, Case Western Reserve University, Cleveland, OH 44106, USA

[3] Department of Mechanical, Industrial, and Manufacturing Engineering, Oregon State University, Corvallis, OR 97331, USA

Introduction

Machinery condition prognosis refers to the estimation of time to failure of the machine, as well as the risk associated with existing or future failure modes (Heng et al. 2009). It is critical to establishing optimized maintenance plans to enhance production, minimize costly downtime, and avoid catastrophic damage. Different sensing techniques have been integrated into the manufacturing system for timely acquisition of its working status to improve the system reliability. According to the sensing parameters, these sensing techniques can be categorized into direct sensing approach and indirect sensing approach (Teti et al. 2010). In the direct sensing approach, the actual quantities, such as crack length, tool wear width, etc. are measured. Such direct sensing approach could directly indicate the machinery condition, however, it is usually performed offline, and usually interrupts the normal operation of the machine. On the contrary, the auxiliary in-process quantities such as force, vibration, acoustic emission, and motor current, are measured in the indirect sensing approach. The indirect sensing approach can online monitor the machinery, thus it is more suitable for practical applications. Since the indirect sensing parameters are the indirect indicators of machinery condition, the machinery condition is then deduced using different data processing techniques.

Numerous efforts have been made to develop a variety of methods for machinery condition prognosis. In Jurkovie et al. (2016), the performance of three different machine learning methods including support vector regression, polynomial regression and artificial neural network have been investigated for the prediction of independent output cutting
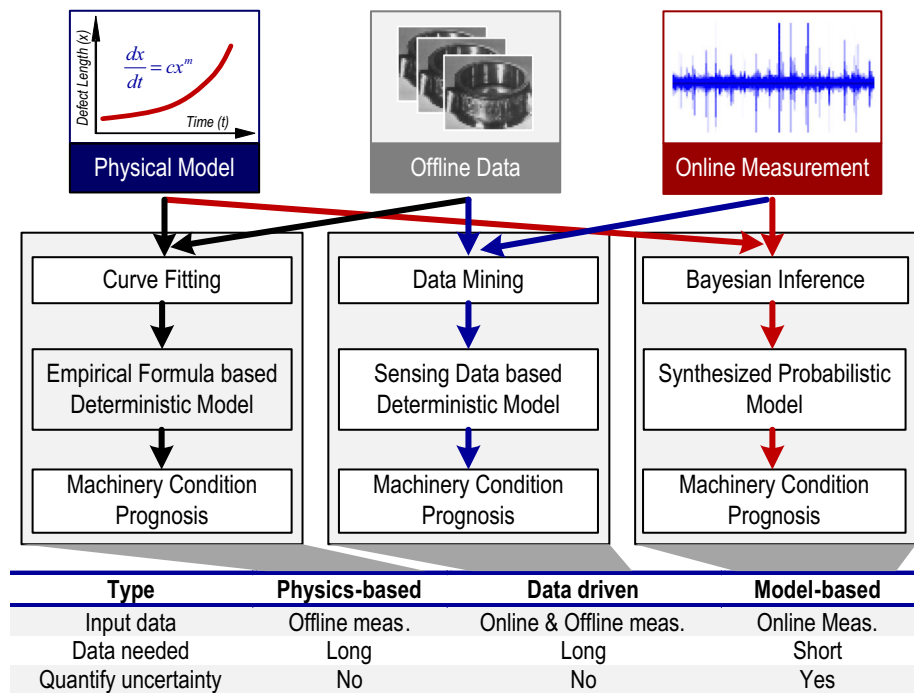
Fig. 1 Comparison of different approaches for machinery condition prognosis

Type	Physics-based	Data driven	Model-based
Input data	Offline meas.	Online & Offline meas.	Online Meas.
Data needed	Long	Long	Short
Quantify uncertainty	No	No	Yes

parameters in a high speed turning process. In Vogl et al. (2016), it reviews the challenges, needs, methods and best practice for PHM within manufacturing systems highlighted by diagnostics, prognostics, dependability analysis, data management, and business. An integrative method of logical analysis and non-parametric cumulative incidence functions is presented to address the multiple failure modes issue in prognosis and the results shows the competitive performance of presented method over neural network and support vector machine (Ragab et al. 2016). The prognostics-based decision support methods are summarized for condition based maintenance and a decision tree learning method is provided to guide the model selection (Bousdekis et al. 2016).

According to the utilization of sensing information, these prognosis techniques can be categorized into physics based approach, data driven approach, and model based approach (also known as physics-data integrative approach) Peng et al. (2010) as illustrated in Fig. 1. Physics based approach typically uses empirical models to describe system physics and these empirical models are usually expressed by a series of ordinary or partial differential equations (Peng et al. 2010). For instance, the bearing degradation could be described by Paris' formula (Paris et al. 1961) or empirical equation L10 to model the crack growth or predict spalling initiation during the bearing life test. Data driven approach derives model from online and offline historical measurements based on artificial intelligence techniques without physical knowledge needed (Heng et al. 2009; Lever et al. 1997). Different artificial intelligence techniques have been investigated for machinery condition prognosis, including artificial neural network (Zhang et al. 2013), support vector regression (Benkedjouh et al. 2015), ARMA/GARCH model (Pham and Yang 2010), hidden Markov model (Peng and Dong 2011), Naïve Bayes (Mehta et al. 2015), fuzzy logic (Gokulachandran and Mohandas 2015), and adaptive neuro-fuzzy inference system (Sarkeyli et al. 2015), etc.

As shown in Fig. 1, physics based approach usually utilizes curve fitting to identify parameters in an empirical equation using offline sensing measurements through extensive experiments. The machinery condition prognosis is then estimated based on deterministic model with determined parameters. In practice, physics based approach may not be the most practical solution since the parameters in the model are validated by large sets of data (Peng et al. 2010). However, it is usually difficult or impossible to obtain sufficient offline measurement data in real application. Compared with physics based approach, data driven approach may be more available in some practical cases in which it is relatively easier to gather data than to build accurate physics models (Heng et al. 2009). However, there are still some limitations. (1) A large amount of historical data (more than physics based techniques needed) is required to train the model in data driven approach and the performance of prediction highly relies on the quality of training data. But it is usually a costly and time consuming process to obtain the required run-to-failure experimental data. (2) The model is derived in a deterministic fashion and only applicable under specific operating conditions. On the other hand, physics based approach and data driven approach do not consider the uncertainty due to component variation and varying operating conditions in the

model construction, and are usually difficult to quantify the uncertainty in the mechanical degradation process.

In comparison, model based approach takes advantage of the merits in physics based approach and data driven approach, and addresses their limitations. Given that the physical knowledge governing defect propagation has been established in physics based approach, model based approach integrates physical knowledge and in-process measurements into a state space model. Since the machinery defect status is usually not directly accessible using in-process measurements, defect status needs to be estimated or predicted from in-process measurements, to which Bayesian inference provides a rigorous mathematic solution. Based on Bayesian inference, the present defect status is estimated based on previous defect status in one-step-ahead prediction. The estimated defect status is then updated using in-process measurements based on Bayes rule. For multi-step-ahead prediction, recursive process is applied to predict the defect propagation in the desired prediction horizon. Depending on the system type and noise assumption, different methods including Kalman filter (for linear system and Gaussian noise) (Kalman 1960), extended Kalman filter (for weak nonlinear system and Gaussian noise) (Julier and Uhlmann 1997), and particle filter (for nonlinear system and non-Gaussian noise) (Gordon et al. 1993) have been investigated to implement model based prognosis. Different from the above-mentioned two methods, model based approach doesn't require too much data, and the uncertainty could be well quantified by way of probability distribution.

Particle filter (PF) is a numerical approximation method based on Bayesian inference using point mass (or 'particle') representation of probability densities to tackle the nonlinearity and non-Gaussianity in modeling system dynamics (Gordon et al. 1993). It has been investigated for condition prognosis and remaining useful life prediction in different applications. In Orchard and Vachtsevanos (2009), a particle filter framework is investigated to analyse the axial crack growth in a planetary carrier plate. A regularized auxiliary particle filter is presented for battery remaining life prediction in Liu et al. (2011). In the previous work, model parameters are usually pre-determined from lab controlled test or FEA simulation analysis. However, model parameters may vary in practice due to various factors, including different material properties, component-to-component variations, or different operating conditions in the machinery degradation process.

To accommodate parameter variations and improve tracking performance, the level of noise associated with the system state and measurement models for degradation modeling can be increased. The drawback of such an approach, however, is that it also decreases the prediction accuracy and precision. Joint state and parameter estimation techniques have been investigated to address the aforementioned issue, and these methods can be categorized into Bayesian estimation and maximum likelihood estimation. In the Bayesian estimation approach, an auxiliary particle filtering using shrinkage modification of kernel smoothing technique is firstly presented in Liu and West (2001) for sequential Bayesian learning about time-varying state vectors and fixed model parameters simultaneously. Particle filter is then investigated in Storvik (2002) for dynamic state space model by marginalizing the unknown static parameter as part of the state vector. Following the similar idea, a Matlab-based tutorial for model-based prognostic is presented in An et al. (2013) by combining a physical model with observed data to identify model parameter using Bayesian estimation in particle filter. A mean-square filter is investigated in Basin et al. (2013) for joint state and parameter estimation in uncertain stochastic nonlinear polynomial system by incorporating the unknown parameters into extended polynomial state vector. Parameter estimation is investigated for batch processes in Zhao et al. (2013) through the maximum likelihood method which is accomplished by Expectation-Maximization algorithm. Prior efforts on joint state and parameter estimation mainly focus on filtering and control applications.

In line with the above efforts, this study investigates a joint state prediction and parameter estimation framework based on particle filter and expectation-maximization (EM) algorithm for machinery condition prognosis. To account for the stochastic property of machinery degradation process, the system is firstly modelled with unknown parameters. Model parameters are then estimated based on the degradation status (hidden information) and available in-process measurements using expectation-maximization algorithm. Machinery degradation status is then predicted by recursively updating the identified model with in-process measurements, following Bayesian inference scheme in particle filter. A simulation study and an experimental bearing run-to-failure test in a wind turbine gearbox are used to demonstrate the effectiveness of the developed method.

The merits of this study rest on the following. (1) A stochastic state space model is constructed for machinery defect prognosis comparing with predetermined models in prior studies. (2) To the best knowledge, it is the first study of integrative particle filter and EM algorithm investigated for machinery condition prognosis. (3) Performance comparison between Bayesian estimation and presented integrative PF and EM method is discussed. The rest of the paper is constructed as follows. After introducing the theoretical background of particle filter and expectation-maximization algorithm in Sect. 2, the details of the proposed mathematical framework for joint state prediction and parameter estimation is discussed in Sect. 3. Additionally, the construction of state space model based on physical knowledge and in-process data analysis is also discussed respectively. The effectiveness of the developed method is experimentally demonstrated based on bearing run-to-failure data acquired

in a wind turbine gearbox in Sect. 4. To further demonstrate the performance of the presented method, performance comparison with conventional approach is discussed, and their Pros and Cons are summarized in Sect. 5. Finally, conclusions are drawn in Sect. 6.

Theoretical framework

In machinery condition prognosis, defect status is usually impossible to be directly observed unless it is measured offline by costly equipment (e.g., microscope, surface profiler). On the other hand, online sensing techniques, such as vibration, acoustic emission, dynamic force, temperature and wear debris, are readily measured in-process. However, online measurements are usually the indirect indicators of machinery status. Such scenario can be well described using state space model in a *Bayesian framework* (Arulampalam et al. 2002) as follows.

a) A state equation describing the underlying machinery degradation process with time:

$$x_t = f(x_{t-1}, u_{t-1}) \tag{1}$$

where t is time index, and x_t is the defect status at time step t. $f(\bullet)$ describes the state transition function from state x_{t-1} to x_t considering order-one Markov process. u_{t-1} is the process noise representing uncertainty in degradation process. The state transition probability $p(x_t|x_{t-1})$ is the intrinsic property of machinery degradation, and can be derived from Eq. (1). The state equation is usually constructed based on physical knowledge of the degradation process (e.g., Paris' formula).

b) A measurement equation linking online measurement to unobservable defect status:

$$z_t = h(x_t, v_t) \tag{2}$$

where z_t denotes the online measurement at time step t. $h(\bullet)$ is a measurement function representing the relationship between online measurements z_t and unobservable defect status x_t. v_t is the sequence of measurement noise. The measurement z_t is considered to be conditionally independent given the degradation state x_t which is described as measurement probability $p(z_t|x_t)$.

c) A set of available online measurements acquired in the experiments:

$$\mathbf{Z} = \{z_1, z_2, \ldots, z_N\} \tag{3}$$

Equations (1–2) describe the system equation and measurement equation with known model parameters. In case of unknown parameters, the general state space models can be described as follows.

$$x_t = f(x_{t-1}, \theta, u_{t-1}) \tag{4}$$
$$z_t = h(x_t, \theta, v_t) \tag{5}$$

where θ represents the unknown parameters need to be estimated. z_t is the available in-process measurement at time step t. x_t denotes the hidden defect status. The state transition probability is denoted as $p_\theta(x_t|x_{t-1})$ while the measurement probability is described as $p_\theta(z_t|x_t)$.

Particle filter

In machinery defect prognosis, the idea is to find the posterior probability distribution $p(x_{t+l}|z_t)$ of future defect status given the present measurements in the state space model with known parameters. The posterior probability distribution is recursively computed in two stages: *prediction* and *update*. Given the posterior probability distribution $p(x_t|z_t)$ at time t, the prediction stage is to estimate the probability distribution $p(x_{t+1}|z_t)$ via the Chapman–Kolmogorov equation (Arulampalam et al. 2002) as:

$$p(x_{t+1}|z_t) = \int p(x_{t+1}|x_t) p(x_t|z_t) dx_t \tag{6}$$

Upon the new measurement z_{t+1} is available, the posterior probability distribution $p(x_{t+1}|z_t)$ of defect state x_{t+1} is updated via Bayes rule (Arulampalam et al. 2002).

$$p(x_{t+1}|z_{t+1}) = \frac{p(x_{t+1}|z_t) p(z_{t+1}|x_{t+1})}{p(z_{t+1}|z_t)} \tag{7}$$

where $p(z_{t+1}|z_t)$ is the normalizing factor which can be calculated as:

$$p(z_{t+1}|z_t) = \int p(x_{t+1}|z_t) p(z_{t+1}|x_{t+1}) dx_{t+1} \tag{8}$$

As discussed above, Eqs. (6–8) give the exact analytic Bayesian solution for linear system with Gaussian noise. However, for nonlinear system, the integral operation in Eq. (6) is usually intractable due to high dimensional computation. Particle filter is developed using sequential Monte Carlo sampling method based on random samples (particle) representation of probability densities for nonlinear or non-Gaussian system. It has been widely studied in different applications such as mobile robot localization (Kwok et al. 2003), and object tracking (Hue et al. 2002), which involve nonlinear system and/or non-Gaussianity noise. Based on Monte Carlo approximation, the integral operation in Eq. (6) is transformed into the summarization to address complex high-dimensional computation in nonlinear system. In particle filter, the posterior probability distribution $p(x_t|z_t)$ at

time t is represented by a set of random samples or particles x_t^i, $i = 1, 2, \ldots, M$, with the corresponding weight w_t^i. Then, Eq. (6) can be reformulated as Arulampalam et al. (2002):

$$
\begin{aligned}
p(x_{t+1}|z_t) &= \int p\left(x_{t+1}|x_t\right) p\left(x_t|z_t\right) dx_t \\
&\approx \sum_{i=1}^{M} w_t^i \delta\left(x_t - x_t^i\right) p\left(x_{t+1}|x_t\right) \\
&= \sum_{i=1}^{M} w_t^i p\left(x_{t+1}|x_t^i\right)
\end{aligned} \tag{9}
$$

where i is the index of particle, and M is the total number of particles which affects the accuracy of represented probability distribution function and the efficiency of computation. $\delta(\bullet)$ is the delta function. In the update step, the weight of each particle is then updated based on the likelihood of the observation z_{t+1} at time step $t + 1$ as:

$$
w_{t+1}^i \propto w_t^i p\left(z_{t+1}|x_{t+1}^i\right) \tag{10}
$$

In implementation, resampling is applied in every step to obtain equally weighted samples so as to avoid particle degeneracy issue of particle filter (Arulampalam et al. 2002). The particles are resampled from importance distribution with associated importance weights. Those particles with very small weights are eliminated, while those particles with high weights are duplicated. Therefore, particle filter is a recursive numerical method based on the Bayesian inference to estimate posterior probability density function of a state which is represented by a set of random samples (named particles) with associated weights (Wang and Gao 2013). Particle filter gives the complete solution of state estimation in the state space model with known model parameters.

Expectation-maximization algorithm

Because of the stochastic property of machinery degradation process, the parameters of state space model are unknown. To solve this issues, the maximum likelihood estimation (MLE) could be used to deduce the most likely parameters which satisfy the distribution of in-process measurements. However, in practice, the measurements are usually incomplete data which will seriously affect the accuracy of estimation. On the other hand, the likelihood function is sometimes too complicated to solve the MLE of corresponding parameters. Expectation-maximization (EM) algorithm (Dempster et al. 1977; Moon 1996), an iterative algorithm, is developed for obtaining the MLE of parameters using incomplete data. Using this method, the complex problem of maximizing likelihood function is translated into the optimization of a series of simple functions. The essence of EM algorithm is

the postulation of hidden defect state X and in-process measurements Z to form a complete data set $\{X, Z\}$, then the log-likelihood function is expressed through marginalizing the joint distribution $p_\theta(X, Z)$.

$$
L_\theta(X, Z) = \log p_\theta(X, Z) \tag{11}
$$

where X is denoted as the N-dimensional hidden degradation state vector $\{x_1, x_2, \ldots, x_N\}$ while Z represents the in-process measurement vector $\{z_1, z_2, \ldots, z_N\}$. To estimate the parameters, EM algorithm maximizes the cost function $Q(\theta, \theta_k)$ which is formulated as the conditional mean of the log-likelihood function $L_\theta(X, Z)$ (Schon et al. 2011).

$$
\begin{aligned}
Q(\theta, \theta_k) &= \int L_\theta(X, Z) p_{\theta_k}(X|Z) dX \\
&= \int \log p_\theta(X, Z) p_{\theta_k}(X|Z) dX
\end{aligned} \tag{12}
$$

where θ_k is the sequence of estimated parameters. There are mainly two major steps in parameter estimation. In the first step, also known as E-step, cost function $Q(\theta, \theta_k)$ is computed as shown in Eq. (12) by estimating the joint probability density of state and in-process measurements based on initial estimated parameters. Then, in the following M-step, the cost function $Q(\theta, \theta_k)$ is maximized using a gradient ascent algorithm to obtain the new estimated parameters $\theta_{k+1} = \arg \max Q(\theta, \theta_k)$. These two steps are iterated until the parameters converge which means the changes of parameters after each iteration remain within a specified tolerance level.

One challenge in EM algorithm is to calculate the cost function $Q(\theta, \theta_k)$. Based on Bayes' rule and Markov property, the joint distribution $p_\theta(X, Z)$ is derived as:

$$
p_\theta(X, Z) = p_\theta(X) p_\theta(Z|X) \tag{13}
$$

where

$$
p_\theta(X) = p_\theta(x_1) \prod_{t=1}^{N-1} p_\theta(x_{t+1}|x_t) \tag{14}
$$

$$
p_\theta(Z|X) = p_\theta(z_1|x_1) \prod_{t=2}^{N} p_\theta(z_t|x_t) \tag{15}
$$

Similarly, the log-likelihood function can be computed as:

$$
\begin{aligned}
L_\theta(X, Z) &= \log p_\theta(X, Z) = \log p_\theta(Z|X) + \log p_\theta(X) \\
&= \log p_\theta(x_1) + \sum_{t=1}^{N-1} \log p_\theta(x_{t+1}|x_t) \\
&\quad + \sum_{t=1}^{N} \log p_\theta(z_t|x_t)
\end{aligned} \tag{16}
$$

Table 1 Illustration of EM algorithm

Input: hidden defect state X; in-process measurements $\mathbf{Z}$; the joint distribution $p_\theta(X, \mathbf{Z})$; the conditional distribution $p_\theta(X \mid \mathbf{Z})$

Output: estimated parameters ;
 1. Initialize θ_0 randomly or heuristically according to any prior knowledge about where the optimal parameter value might be.
 2. Iteratively improve the estimate of θ by alternating between the following two-steps:
     (a) The E-step (expectation): Compute the cost function $Q(\theta, \theta_k)$ according to Eqs. (12) and (17-20).
     (b) The M-step (maximization): Re-estimate θ by maximizing the $Q(\theta, \theta_k)$:
         $\theta_{k+1} = \arg\max Q(\theta, \theta_k)$
 3. Repeat the step 2 until the log-likelihood $L_\theta(X, \mathbf{Z})$ converges.

According to Eqs. (12) and (16), the cost function $Q(\theta, \theta_k)$ is then transformed as Schon et al. (2011):

$$Q(\theta, \theta_k) = I_1 + I_2 + I_3 \tag{17}$$

where

$$I_1 = \int \log p_\theta(x_1) p_{\theta_k}(x_1 | \mathbf{Z}) dx_1 \tag{18}$$

$$I_2 = \sum_{t=1}^{N-1} \int \int \log p_\theta(x_{t+1} | x_t) p_{\theta_k}(x_{t+1}, x_t | \mathbf{Z}) dx_t dx_{t+1} \tag{19}$$

$$I_3 = \sum_{t=1}^{N} \int \log p_\theta(z_t | x_t) p_{\theta_k}(x_t | \mathbf{Z}) dx_t \tag{20}$$

For linear system with Gaussian noise, the cost function $Q(\theta, \theta_k)$ can be calculated analytically, and the explicit equations for parameter estimation in EM algorithm can be developed. However, for nonlinear state space model in machinery defect prognosis, the integral operation in Eqs. (18–20) is intractable to get a complete solution which poses challenge in EM algorithm. In such case, particle filter plays an import role in computation of the cost function, which will be discussed in the next section of the integrated framework of particle filter and expectation-maximization algorithm for machinery defect prognosis. The details of EM algorithm are given in Table 1.

Formulation of prognosis model

In order to accommodate model dynamics (parameter variations) and reduce prognosis uncertainty, this paper presents a computational framework of joint state prediction and parameter estimation for machinery condition prognosis based on particle filter and EM algorithm as shown in Fig. 2.

According to the inherent physical knowledge of different applications, a state space model consisting of system equation and measurement equation is constructed firstly.

With the available in-process measurements, system identification is performed to identify the model parameters using expectation-maximization algorithm in the learning stage. After the model parameters are identified, the system state is then predicted based on particle filter in the desired prediction horizon. The details of the joint state prediction and parameter estimation framework are discussed below.

Construction of state space model

As discussed in Sect. 2, system equation and measurement equation are essential to construct the state space model which describes the machinery degradation process.

System equation

System equation describes machinery defect state evolving behavior which is usually impossible to be measured directly through online measurements. Take defect prognosis of rolling element bearing as an example, spalling area is a direct indicator of bearing defect severity. Since the bearing defect grows slowly and the physics governing its behavior is known, a physics-based model governing fatigue crack propagation based on Paris' formula is given by Li et al. (1999):

$$\frac{dx}{dt} = \alpha(\Delta k)^m \tag{21}$$

where t is the number of cycles related to bearing running time, and dx/dt is the defect growth rate. The parameters α and m are related to material properties, and Δk is the amplitude of the stress intensity factor. In the case of bearing defect propagation, it is difficult to estimate the stress intensity factor. To circumvent it, an empirical model is developed as Li et al. (1999):

$$\frac{dx}{dt} = cx^m \tag{22}$$

where x represents the spalling area. The defect growth rate is an exponential function of the existing spalling area.
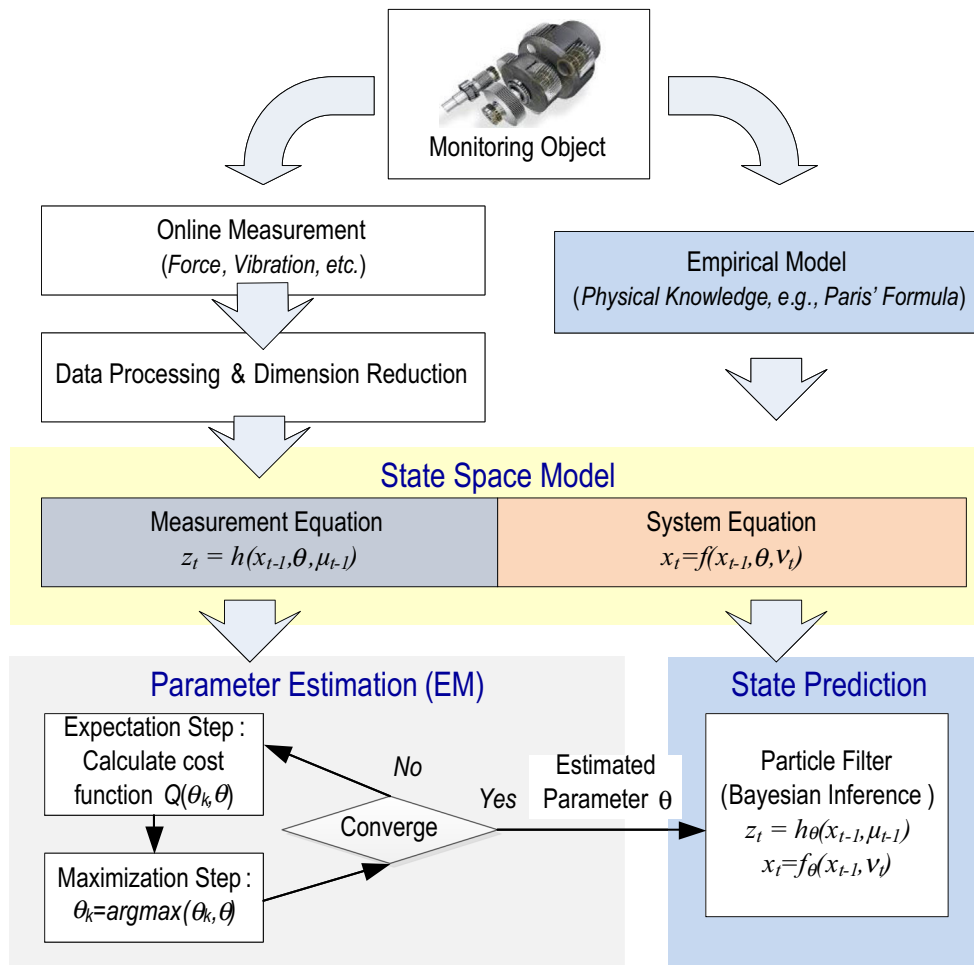
Fig. 2 Computational framework for joint state prediction and parameter estimation in machinery condition prognosis

The model parameters c and m are initialized as unknown variables. Take integration operation on both sides after separating variables, Eq. 22 can be rewritten in a state transition form as:

$$x_t = \left[x_{t-1}^{(1-m)} + c(1-m) \right]^{1/(1-m)} + u_{t-1} \qquad (23)$$

where t is the time index, and u_{t-1} represents the noise in the state evolving process.

Measurement equation

During machinery degradation process, in-process measurements carrying defect signatures have been widely employed. However, due to the complexity and the low signal to noise ratio (SNR) of raw measurements, it is difficult to model the relationship between the measurements and machinery defect status. To tackle this problem, effective data processing and feature representation techniques are performed to reduce data dimensionality without losing the information of defect signature. The relationship between extracted feature and defect state is expected to be modelled using a simple function.

Different features from time domain, frequency domain, and time-frequency domain have been investigated for bearing defect diagnosis and prognosis. Time domain methods involve statistical features such as root mean square (RMS), Kurtosis, skewness, crest factor, peak–peak value, and entropy. The frequency domain methods take advantage of spectrum/time-frequency analysis and extract the energies at bearing defect characteristic frequencies (e.g. the ball-pass frequency of outer ring f_{BPFO}, ball-pass frequency of inner ring f_{BPFI}, and ball-spin frequency f_{BSF}) as features. A number of features are extracted from the measurements to represent the bearing defect. However, these features may contain redundant information. For improved computational efficiency in machinery condition prediction, a proper feature selection/fusion strategy is needed to lower the dimension of feature space. Different feature selection techniques have been investigated in the literature,

including genetic algorithm (Liao 2014), principle component analysis (PCA) (Malhi et al. 2011), and health index (Bechhoefer and Bernhard 2007), etc. Generally, it is difficult to determine which feature is more sensitive to bearing defect propagation. A good feature should present a consistent trend with defect propagation, and is expected to describe the measurement equation using a simple function. The assumption on linear relationship between bearing defect status and selected feature, which is often used in the literature (Pedregal and Carnero 2006), is also adopted in this study.

$$z_t = x_t + v_t \tag{24}$$

where z_t denotes the selected feature from in-process measurements, and v_t represents the measurement noise.

Formulation of joint parameter estimation and state prediction

The state space model with unknown parameters has been constructed. To predict the machinery condition based on available observation, a mathematic framework based on particle filter and expectation-maximization algorithm for joint state prediction and parameter estimation is investigated in this study. To estimate the parameters in the nonlinear state space model, the cost function $Q(\theta, \theta_k)$ described in Eqs. (17–20) is computed based on the sequential importance resampling method in particle filter. From Eqs. (17–20), it is seen that the computation of quantities I_1, I_2, and I_3 depends primarily on the smoothed density function $p_{\theta k}(x_t \mid Z)$. To estimate the density function $p_\theta(x_t \mid Z)$, a forward–backward smoother algorithm is investigated and proofed in Schon et al. (2011).

$$
\begin{aligned}
p_\theta(x_t|\mathbf{Z}) &= \int p_\theta(x_t|x_{t+1}, z_{1:t}) p_\theta(x_{t+1}|\mathbf{Z}) dx_{t+1} \\
&= \int \frac{p_\theta(x_{t+1}|x_t) p_\theta(x_t|z_{1:t})}{p_\theta(x_{t+1}|z_{1:t})} p_\theta(x_{t+1}|\mathbf{Z}) dx_{t+1} \\
&= p_\theta(x_t|z_{1:t}) \int \frac{p_\theta(x_{t+1}|x_t) p_\theta(x_{t+1}|\mathbf{Z})}{p_\theta(x_{t+1}|z_{1:t})} dx_{t+1}
\end{aligned}
\tag{25}
$$

Hence, the smoothed density function can be obtained as the function of filtered state density $p_\theta(x_t|z_{1:t})$ at time t, smoothed density $p_\theta(x_{t+1}|\mathbf{Z})$ at time t+1, and state prediction density $p_\theta(x_{t+1}|x_t)$. Based on particle approximation in particle filter, the smoothed density function $p_\theta(x_t|Z)$ can be obtained following the recursive approximated density function $p_\theta(x_{t+1} \mid \mathbf{Z})$ as:

$$p_\theta(x_t|\mathbf{Z}) \approx \sum_{i=1}^{M} \omega_{t|N}^i \delta\left(x_t - x_t^i\right) \tag{26}$$

where

$$\omega_{t|N}^i = \omega_t^i \sum_{k=1}^{M} \omega_{t+1|N}^k \frac{p_\theta\left(x_{t+1}^k|x_t^i\right)}{\sum_{i=1}^{M} \omega_t^i p_\theta\left(x_{t+1}^k|x_t^i\right)} \tag{27}$$

Similarly, the quantities I_1, I_2 and I_3 which constitute the cost function $Q(\theta, \theta_k)$ can be obtained accordingly as Schon et al. (2011):

$$Q(\theta, \theta_k) = I_1 + I_2 + I_3 \tag{28}$$

where

$$I_1 = \sum_{i=1}^{M} \omega_{1|N}^i \log p_\theta\left(x_1^i\right) \tag{29}$$

$$I_2 = \sum_{t=1}^{N-1} \sum_{i=1}^{M} \sum_{j=1}^{M} \frac{\omega_t^i \omega_{t+1|N}^j p_{\theta_k}\left(x_{t+1}^j|x_t^i\right)}{\sum_{l=1}^{M} \omega_t^l p_{\theta_k}\left(x_{t+1}^j|x_t^l\right)} \log p_\theta\left(x_{t+1}^j|x_t^i\right) \tag{30}$$

$$I_3 = \sum_{t=1}^{N} \sum_{i=1}^{M} \omega_{t|N}^i \log p_\theta\left(z_t|x_t^i\right) \tag{31}$$

By recursively maximizing the cost function $Q(\theta, \theta_k)$ using optimization algorithm, such as gradient descent algorithm (Qian 1999), the parameter vector θ can be determined when it converges.

$$\theta_{k+1} = \theta_k - \gamma \cdot \nabla Q(\theta, \theta_k) \tag{32}$$

where γ is the step size, $\nabla(\cdot)$ is the gradient function. In machinery condition prognosis, we are interested in degradation state prediction with the identified model parameter vector θ. For l-step ahead prediction, the posterior probability distribution $p_\theta(x_{t+l}|z_t)$ can be obtained according to Eqs. (4–6) as Zio and Peloni (2011):

$$p_\theta(x_{t+l}|z_t) = \int \cdots \int \prod_{t=N}^{N+l} p_\theta(x_{t+1}|x_t) p_\theta(x_t|z_t) \prod_{t=N}^{N+l} dx_t \tag{33}$$

To eliminate the computation of integral operation in nonlinear system, particle filter is used to calculate the posterior probability distribution $p_\theta(x_{t+l}|z_t)$ as:

$$p_\theta(x_{t+l}|z_t) \approx \sum_{i=1}^{M} w_{t+l-1}^i p_\theta\left(x_{t+l}|x_{t+l-1}^i\right) \tag{34}$$

Fig. 3 The simulated data and defect state
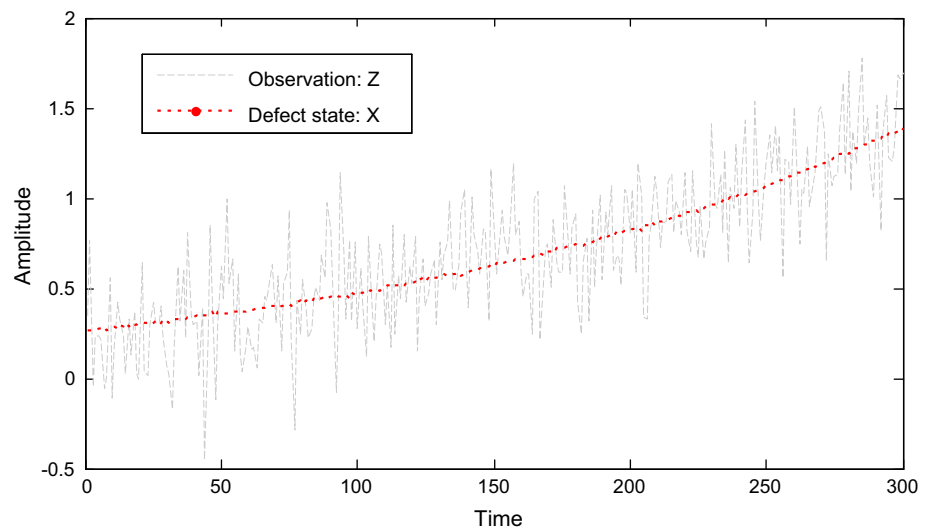
Simulation study

To evaluate the performance of the developed joint state prediction and parameter estimation algorithm, the state space model described in Eqs. (23–24) is used as an example.

$$x_{t+1} = \left[x_t^{(1-m)} + c(1-m) \right]^{1/(1-m)} + u_t \begin{bmatrix} u_t \\ v_t \end{bmatrix}$$
$$z_t = x_t + v_t$$
$$\sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} q & 0 \\ 0 & r \end{bmatrix} \right) \quad (35)$$

here, the parameters q and r are the variances of noises in system equation and measurement equation, respectively. The simulated data is firstly generated by the above model with preset parameters in order to be similar with the experimental data in Sect. 4. The parameters are preset as:

$$\theta^* = [m^*, c^*, q^*, r^*] = [0.9305, 0.0053, 0.0015, 0.252] \quad (36)$$

By setting the initial state x_1 as 0.27, the simulated data z and defect state x can be generated as shown in Fig. 3.

The developed algorithm is investigated to determine the parameters and to predict the future state x on the basis of available observation $z_{1:200}$. Using EM algorithm, it is straightforward to estimate the parameters as described in Table 2. It is seen that the estimated parameters match well with the preset values. It is noted that the estimated variance q of the state vector is less than the preset value with an error of 13.3 % which means the estimated state is smoother than the simulated state.

Next, particle filter is used to estimate the future state based on available observation and estimated parameters.

Table 2 The preset and estimated parameters used in the simulation study

Parameters	Preset value	Estimated value	Error (%)
m	0.9305	0.9301	0.0043
c	0.0053	0.0053	0
q	0.0015	0.0013	13.3
r	0.252	0.2468	2.06

Figure 4 shows the quantified uncertainty of predicted state on the basis of available observation $z_{1:250}$ under 50 steps ahead prediction. Since the estimated states follow probability distributions, the probability distributions of the estimated states $\times 150$, $\times 250$ and $\times 300$ are illustrated in the zoomed regions.

To quantify the prediction accuracy, root mean square error (RMSE) is defined as the square root of the average of the square of all difference between predicted state $\hat{x}$ and simulated state x.

$$RMSE = \sqrt{\frac{1}{N} \sum_{j=1}^{N} (\hat{x}_j - x_j)^2} \quad (37)$$

Figure 5 shows the root mean square error of the predicted state under different-steps-ahead prediction (e.g., 10, 20, 50, 100, and 150 steps) using the developed algorithm. It is found that the prediction error in short term prediction (e.g., around 0.05 % in 10, 20, and 50 steps ahead) is less than that of long term prediction (e.g., around 0.2 % in 100 and 150 steps ahead). The prediction error is relative small in all cases which demonstrates the effectiveness of the developed algorithm in parameter estimation and state prediction.

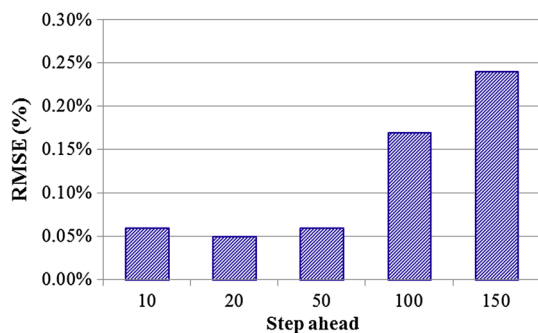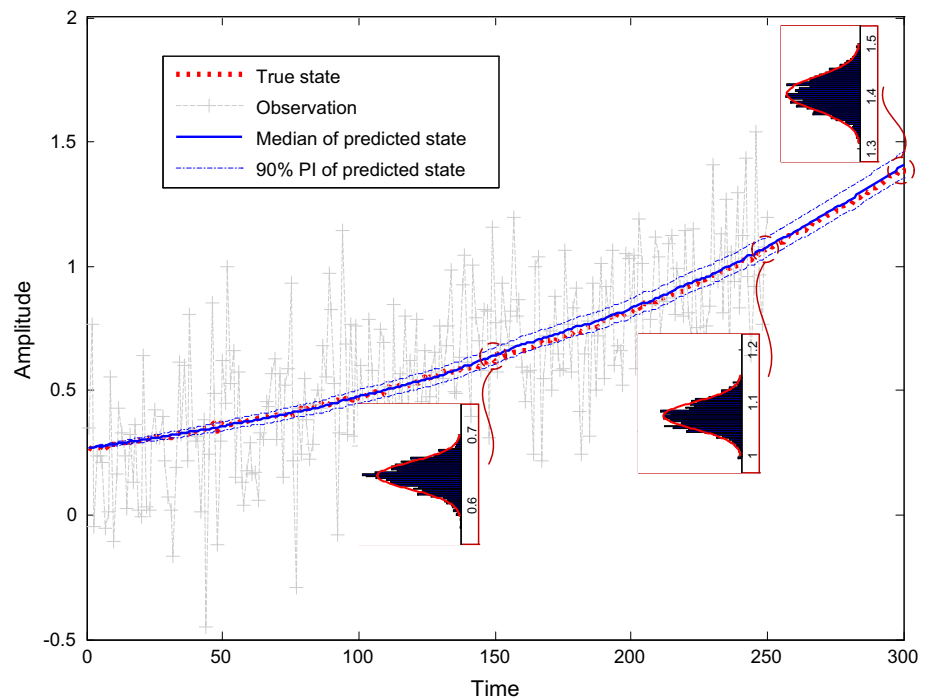**Fig. 4** Uncertainty quantification of predicted state in simulation study

Fig. 5 Prediction error of the developed algorithm under different-step-ahead prediction

Experimental evaluation

To experimentally evaluate the joint state prediction and parameter estimation method based on particle filter and expectation-maximization algorithm, a set of vibration signals measured on a wind turbine gearbox in a run-to-failure test is analyzed, and the analysis results are discussed as below.

Experimental setup

Experimental data measured on a wind turbine gearbox was made available in a field test. The tested turbine is a stall-controlled, three-bladed, upwind machine with 2 MW rated power. The wind turbine gearbox is composed of one low

speed planetary stage and two parallel stages, and its detailed configuration is described in Fig. 6. The defect was initiated from a rolling element bearing A in the high speed shaft as shown in Fig. 6b. The tested bearing run continuously for about 107 million revolutions in the field to the end of service life.

Data preprocessing

With the growth of bearing defect, the energies at bearing defect characteristic frequencies are increased, and thus they can be extracted as the representative features of bearing status. Due to the varying speed operating conditions of the wind turbine, an integrated approach of complex wavelet transform and computed order tracking can be used to extract the energies at the bearing defect characteristic frequencies as representative features (Wang et al. 2014). Figure 7 shows the extracted energies at bearing defect characteristic frequencies as the indicators of bearing status in the run-to-failure test. It is found that the energies at the frequency f_{BPFI} show a clearly increasing trend which means bearing experienced a defect on the inner raceway. The defect may propagate into the cage and the outer raceway since the energies at the frequencies f_{FTF} and f_{BPFO} also show an increasing trend.

The extracted energies at defect characteristic frequencies are denoted as condition indicators (CI) which mean these features indicate the bearing conditions. However, each CI only contains partial information of bearing status or it is sensitive to only one failure mode of bearing. To compre-
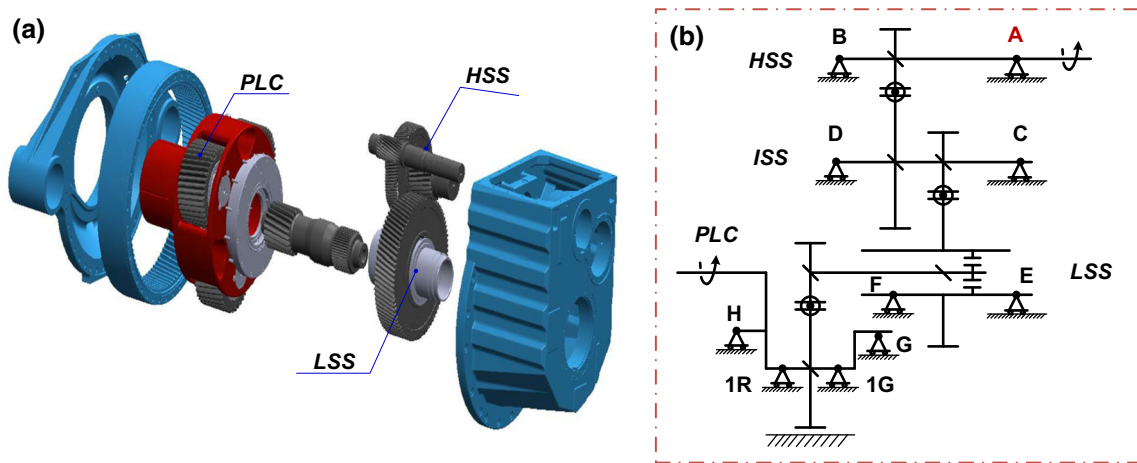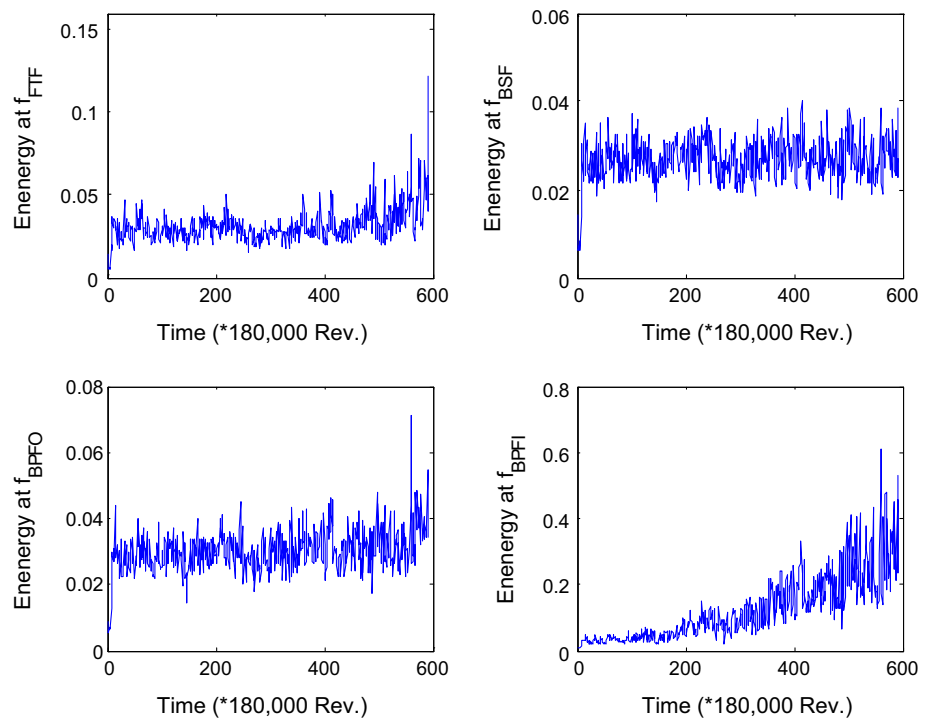
Fig. 6 Illustration of the tested wind turbine gearbox, **a** 3D gearbox structure, **b** internal configuration

Fig. 7 The extracted energies at bearing defect characteristic frequencies

hensively describe the degradation process, the health index (HI) algorithm illustrated in Bechhoefer and Bernhard (2007) is used to fuse all these CIs:

$$HI = \sqrt{CI^T \Sigma^{-1} CI} \times 0.7/v \qquad (38)$$

where vector *CI* represents the energies at bearing defect characteristic frequencies f_{FTF}, f_{BSF}, f_{BPFI} and f_{BPFO}, and Σ is the sample covariance from a set of nominal bearings. The term v is a normalized factor (Bechhoefer and Bernhard 2007). The obtained fused feature describes a clearly trend representing bearing defect growth as shown in Fig. 8.

Based on statistical information (e.g., probability distribution of feature), the fused feature is scaled to represent the defect severity. Based on the rule which has been successfully implemented in aerospace condition monitoring (Beckhoefer et al. 2011), the amplitude in the range of [0 0.5] is set as normal or healthy condition, while the range of [0.5 0.75] is set as mild defective condition, and the amplitude above 0.75 is considered as severely defective condition. The value greater than 1 indicates that the continuous operations may result in collateral damage to other components in the gearbox. Therefore, the threshold is set as 1 to determine the component remaining useful time. It is seen that the fused feature presents dominant noise which causes computational

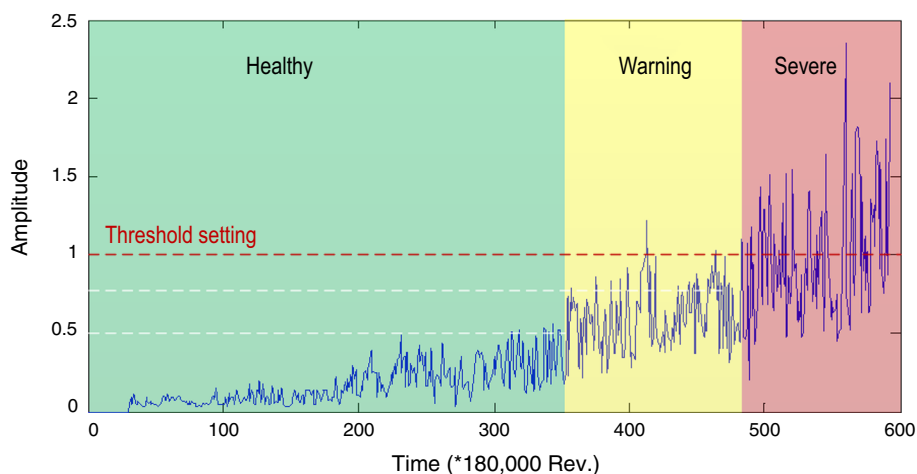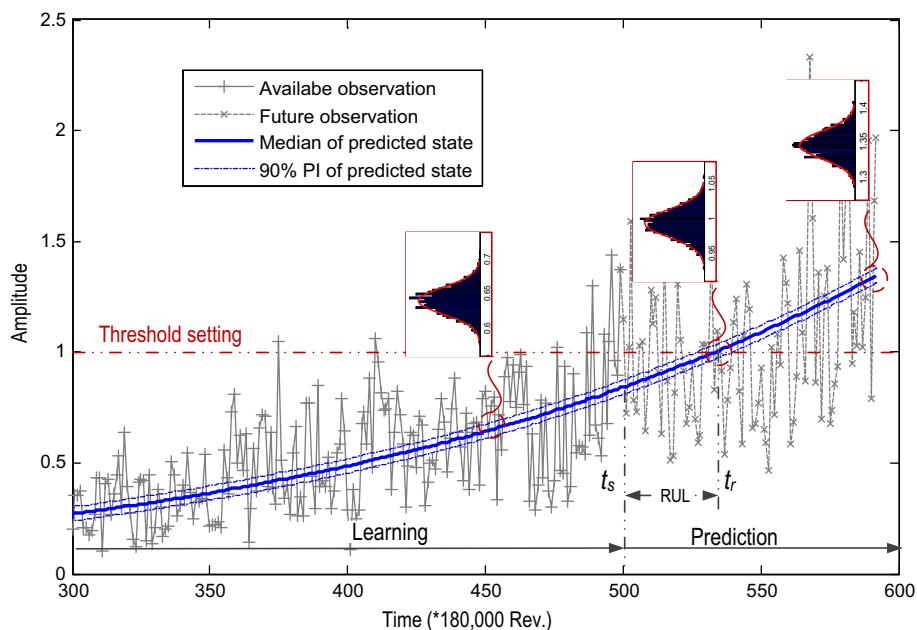**Fig. 8** The fused feature obtained in health index algorithm

Fig. 9 Uncertainty quantification of predicted state using integrative PF and EM method

difficulties of the remaining useful life (RUL). However, due to the dominant noise in fused feature, it is difficult to determine the remaining useful life from the fused feature. The developed joint state prediction and parameter estimation algorithm can quantify the uncertainty of defect growth and reduce the noise variance in fused feature, and thus reduce the false alarm rate in threshold setting as discussed below.

Analysis results

The fused features from time steps [300 500] are chosen as the available observation to predict the bearing defective status in future. The bearing defect growth process is modeled by the system equation and measurement equation in the state space model as described in Eq. 35. According to available observation, the parameters in the state space model are firstly estimated using EM algorithm as:

$$\hat{\theta} = [\hat{m}, \hat{c}, \hat{q}, \hat{r}] = [0.9322, 0.0052, 0.021, 0.2512] \quad (39)$$

Thus, the state space model can be identified as:

$$x_{t+1} = \left[x_t^{0.0678} + 0.0052 * 0.0678 \right]^{1/0.0678}$$
$$\underbrace{}_{z_t = x_t + v_t}$$
$$+ u_t \quad \begin{bmatrix} u_t \\ v_t \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.021 & 0 \\ 0 & 0.2512 \end{bmatrix} \right) \quad (40)$$

Next, particle filter is used to predict the future bearing defective state based on available observation and identified state space model. The predicted bearing defect status with quantified uncertainty in next 100 steps is shown in Fig. 9, and the probability distributions of estimated bearing states are illustrated in zoomed regions. It is found that the predicted bearing defect state using integrative PF and EM method closely follows the trend of observation with highly reduced variance in

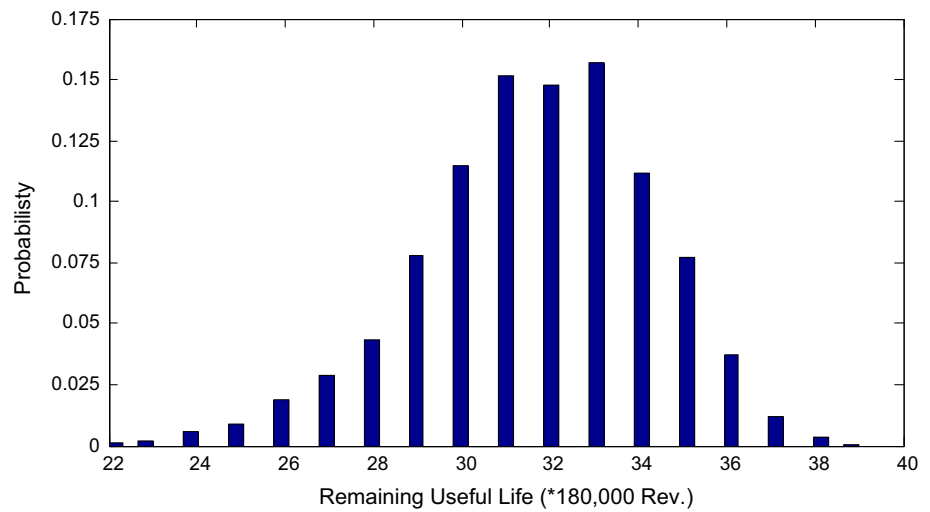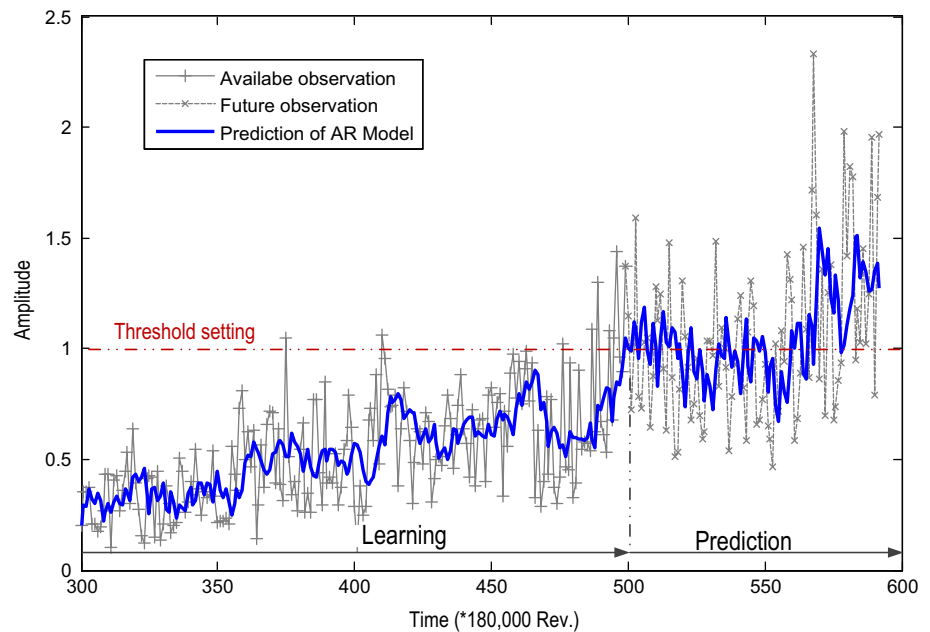**Fig. 10** The probability distribution of bearing remaining useful life

Fig. 11 Predicted bearing defect growth using the ARMA model

uncertainty. Because of the defect accumulating effect in the system equation, the predicted bearing defect status grows monotonically since mechanical defect could not heal itself. It can facilitate the remaining useful life (RUL) calculation of the bearing. Bearing RUL is defined as the service time until the defect state reaches the threshold as illustrated in Fig. 9.

$$RUL(t) = t_r - t_s \qquad (41)$$

where t_s is the present time and t_r is the time when the bearing defect state reaches the threshold. By setting the threshold as 1, the probability distribution of the bearing remaining useful life can be computed as shown in Fig. 10. It is seen that the remaining useful life of bearing is around 33 time steps

(e.g., about 180,000 REV.) by selecting the highest probability distribution.

A data driven approach is also investigated for the machinery condition prediction using the same experimental data for performance comparison. Given the limited number of experimental data set, an autoregressive moving average (ARMA) model is used to predict the bearing defect growth, and the results are shown in Fig. 11. It is found that the prediction results of the ARMA model follow the trend closely; however, it could not quantify prediction uncertainty. The up and down trend also causes the difficulty to determine the remaining useful life of bearing.

From the above analysis, the integrated PF and EM algorithm can incorporate the estimation of model parameters and prediction of hidden defect state in one framework. By taking advantage of physical knowledge, the hidden defect state

Fig. 12 Distribution of model parameters $\bar{m}$ and $\bar{c}$ using conventional PF approach
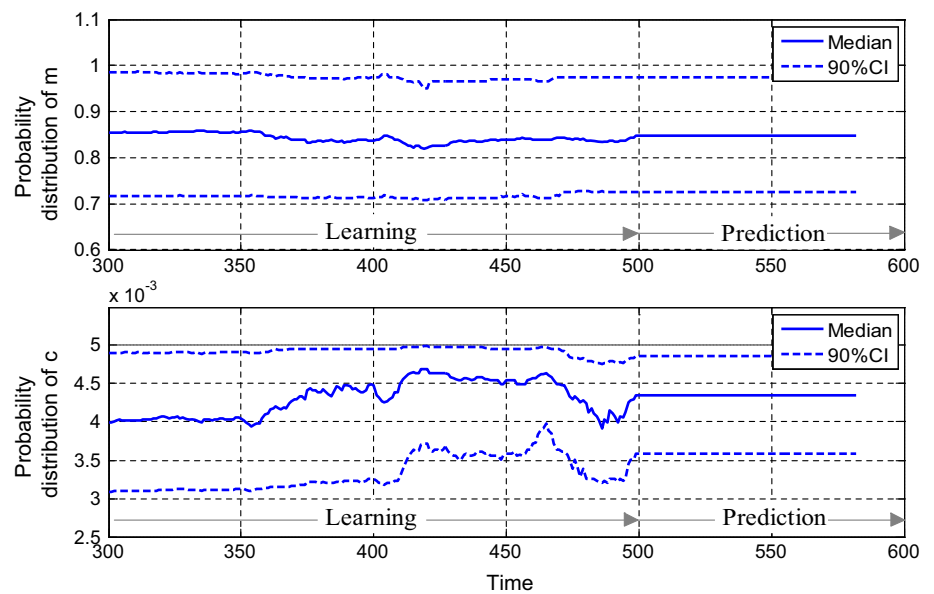
of bearing can be inferred from noisy measurement based on Bayesian inference. The estimated defect state with threshold setting can be used to determine the remaining useful life of bearing. The uncertainty of remaining useful life is also quantified in a probabilistic manner. Therefore, the developed method can facilitate the threshold setting with estimated defect state (reduced variance) instead of noisy measurement to reduce false alarm rate and improve the maintenance strategy in practice.

Discussions

To compare the performance of developed method, the conventional PF approach, which takes the unknown parameters as the part of state vector based on Bayesian estimation, is investigated here. The conventional method still uses the state space model denoted by Eq. 35 as its system equation and measurement equation. The same fused features are chosen as training data of the conventional particle filter to predict the RUL of bearing. In the conventional method, the initial parameters of model are set as probability distributions. If no prior information is available, uniform distribution is usually chosen and the probability parameters of lower and upper bounds are empirically selected. With the prior knowledge about parameter estimated using EM algorithm, the uniform distributions of parameters are determined as:

$$\bar{\theta} = [\bar{m} \sim (0.7, 1), \bar{c} \sim (0.003, 0.005)] \tag{42}$$

Here, the terms $\bar{m}$ and $\bar{c}$ represent the parameters in system equation. The process noise u and the measurement noise

v are ignored since they can be handled through the uncertainty in model parameters. With the available measurement information, model parameters could be updated based on Bayesian estimation and the results are shown in Fig. 12. It is found that the probability distribution of model parameters $\bar{m}$ and $\bar{c}$ fluctuate continuously during the learning stage, but they remain constant during the prediction stage because no measurement is available to update the model parameters. Thus, in the prediction stage, the state space model is determined and particle filter is used for bearing defect prognosis. Figure 13 shows the long-term bearing defect prediction results using the conventional PF approach. It is found that the median of the predicted bearing defect state closely follows the trend of observation, and the uncertainty of prediction result is well demonstrated through the 90 % confidence interval.

Comparing the integrated PF and EM algorithm with the conventional particle filter, it is found that both methods could follow the bearing defect propagation trend closely and quantify the uncertainty of prediction process in the form of confidence intervals. However, the distribution interval of predicted RUL using the integrated PF and EM algorithm is much smaller than that using the conventional particle filter. On the other hand, the prediction accuracy of conventional particle filter approach highly relies on the initial selection of model parameters since the estimation error of model parameters could be propagated in the sequential Bayesian estimation process. In essence, the conventional algorithm and the presented algorithm are based on Bayesian estimation and maximum likelihood estimation, respectively. Bayesian estimation could adapt to the local variation (e.g., the variation of measurement) and update the relevant parameters. Therefore, there are fluctuations during the learning stage

Fig. 13 Uncertainty
quantification of predicted state
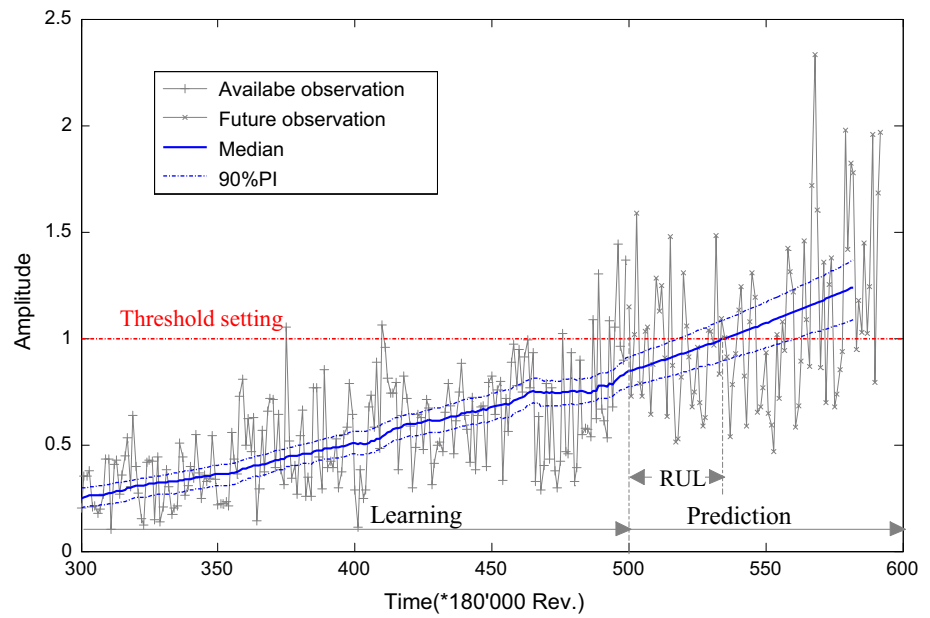using conventional PF approach

Table 3 The comparison between the conventional approach and presented method

Methods	Conventional PF approach (Bayesian estimation)	Integrative PF and EM method (maximum likelihood estimation)
Parameter value	Probability distribution	Deterministic value
Computational efficiency	High	Low
Adapt to local variation	Yes	No
Global property	No	Yes
Error propagation	Yes	No
Effect of initial parameter selection	Probability distribution of parameter needs to be known, inappropriate parameter selection may lead to unsatisfactory results	Good initial selection of parameters improve computational efficiency, but the initial selection of parameters do not affect the prediction accuracy

as shown in Figs. 12 and 13. On the contrary, maximum likelihood estimation calculate model parameters as deterministic value. Once parameters are determined, they will not change during the prediction process. Because of the comprehensive consideration of all available measurements in parameters determination, maximum likelihood estimation has the global property in parameter estimation. The parameters with the maximum likelihood probability are determined during the learning stage, thus the estimation error is not propagated. The Pros and Cons of these two approaches are summarized in Table 3.

Conclusions

Data analytic methods have been extensively investigated to complement sensing techniques for improved systems diagnosis, prognosis, and reliability. This paper presents

a joint state prediction and parameter estimation method to address nonlinearity and stochastic properties that are typically associated with machine degradation processes, based on integrated particle filtering (PF) and expectation maximization (EM). From the results obtained, following conclusion can be drawn:

1. The presented method makes use of prior physical knowledge about the system being estimated as well as new measurement data, and presents solution in a Bayesian framework. Uncertainty associated with the degradation state prediction can be quantified in terms of probability distribution functions.

2. The integration of EM with PF improves the accuracy of system parameter estimation based on measurements. As a result, the presented method can be applied to degradation prediction of systems with limited known model parameters.

3. The presented method can minimize the effect of variance in prediction, caused by measurement noise, thus improving the overall robustness in sensing data thresholding for alarm triggering and maintenance scheduling.

The performance of the presented method has been verified using both simulation and a run-to-failure experimental test on a bearing for wind turbines. Several limitations remain to be further investigated. For example, the computational efficiency of the presented method is low, due to the low convergence rate of the EM algorithm. It was also noted that initial selection of parameters during the model parameters estimation process can significantly affect the computational efficiency. Future research will be directed towards improving the computational efficiency by e.g., replacing the EM algorithm with methods of higher convergence rate, such as Expectation Conditional Maximization (ECM) or Expectation/Conditional Maximization Either (ECME) (Mkhadri 1998). Further systematic experiments will be performed to investigate the robustness of the developed method.

References

An, D., Choi, J. H., & Kim, N. H. (2013). Prognostics 101: A tutorial for particle filter-based prognostics algorithm using Matlab. *Reliability Engineering & System Safety*, *115*, 161–169.

Arulampalam, M. S., Maskell, S., Gordon, N., & Clapp, T. (2002). A tutorial on particle filters for online nonlinear/Non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, *50*(2), 174–188.

Basin, M. V., Loukianov, A. G., & Hernandez-Gonzalez, M. (2013). Joint state and parameter estimation for uncertain nonlinear polynomial systems. *International Journal of Systems Science*, *44*(7), 1200–1208.

Bechhoefer, E., & Bernhard, A. P. F. (2007). A generalized process for optimal threshold setting in HUMS. In *Proceedings of 2007 IEEE Aerospace Conference* (pp. 1–9.). Big Sky, MT, March 3–10.

Beckhoefer, E., He, D., & Dempsey, R. (2011). Gear health threshold setting based on a probability of false alarm. In *Proceedings of Annual Conference of the Prognostics and Health Management Society* (pp. 1–7). Montreal, Quebec, Canada, September 25–29.

Benkedjouh, T., Medjaher, K., Zerhouni, N., & Rechak, S. (2015). Health assessment and life prediction of cutting tools based on support vector regression. *Journal of Intelligent Manufacturing*, *26*, 213–223.

Bousdekis, A., Magoutas, B., Apostolou, D., & Mentzas, G. (2016). Review, analysis and synthesis of prognostic-based decision support methods for condition based maintenance. *Journal of Intelligent Manufacturing*,. doi:10.1007/s10845-015-1179-5.

Dempster, A., Laird, N., & Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, *39*(1), 1–38.

Gokulachandran, J., & Mohandas, K. (2015). Comparative study of two soft computing techniques for the prediction of remaining useful life of cutting tools. *Journal of Intelligent Manufacturing*, *26*, 255–268.

Gordon, N. J., Salmond, D. J., & Smith, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings-F Radar and Signal Processing*, *140*(2), 107–113.

Heng, A., Zhang, S., Tan, A. C. C., & Mathew, J. (2009). Rotating machinery prognostics: State of the art, challenges, and opportunities. *Mechanical Systems and Signal Processing*, *23*, 724–739.

Hue, C., Le Cadre, J. P., & Perez, P. (2002). Tracking multiple objects with particle filtering. *IEEE Transactions on Aerospace and Systems*, *38*(3), 791–812.

Julier, S. J., & Uhlmann, J. K. (1997). A new extension of Kalman Filter to nonlinear systems. In *Proceedings of 11th International Symposium on Aerospace/Defense sensing, simulation and controls, multi sensor fusion, tracking and resource management* (pp. 1–12).

Jurkovie, Z., Cukor, G., Brezocnik, M., & Brajkovie, T. (2016). A comparison of machine learning methods for cutting parameters prediction in high speed turning process. *Journal of Intelligent Manufacturing*,. doi:10.1007/s10845-016-1206-1.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME-Journal of Basic Engineering*, *82*, 35–45.

Kwok, C., Fox, D., & Meila, M. (2003). Adaptive real-time particle filters for robot localization. In *Proceedings of the 2003 IEEE International Conference on Robotics & Automation* (pp. 2836-2841). Taipei, Taiwan, September 14–19.

Lever, P. J. A., Marefat, M. M., & Ruwani, T. (1997). A machine learning approach to tool wear behavior operational zones. *IEEE Transactions on Industry Applications*, *33*(1), 264–73.

Liao, L. (2014). Discovering prognostic features using genetic programming in remaining useful life prediction. *IEEE Transactions on Industrial Electronics*, *61*(5), 2464–2472.

Li, Y., Billington, S., Zhang, C., Kurfess, T., Danyluk, S., & Liang, S. (1999). Dynamic prognostic prediction of defect propagation on rolling element bearings. *Tribology Transactions*, *42*(2), 385–392.

Liu, J., & West, M. (2001). Combined parameter and state estimation in simulation-based filtering. In *Sequential Monte Carlo Methods in Practice* (pp. 197–223). New York: Springer.

Liu, J., Wang, W., & Ma, F. (2011). A regularized auxiliary particle filtering approach for system state estimation and battery life prediction. *Smart Materials and Structures*, *20*, 1–9.

Malhi, A., Yan, R., & Gao, R. X. (2011). Prognosis of defect propagation based on recurrent neural networks. *IEEE Transactions on Instrumentation and Measurement*, *60*(3), 703–711.

Mehta, P., Werner, A., & Mears, L. (2015). Condition based maintenance-systems integration and intelligence using Bayesian classification and sensor fusion. *Journal of Intelligent Manufacturing*, *26*, 331–346.

Mkhadri, A. (1998). On the rate of convergence of the ECME algorithm. *Statistics & Probability Letters*, *37*(1), 81–87.

Moon, T. K. (1996). The expectation-maximization algorithm. *IEEE Signal Processing Magazine*, *13*(6), 47–60.

Orchard, M. E., & Vachtsevanos, G. J. (2009). A particle-filtering approach for on-line fault diagnosis and failure prognosis. *Transactions of the Institute of Measurement and Control*, *31*(3/4), 221–246.

Paris, P. C., Gomez, M. P., & Anderson, W. E. (1961). A rational analytic theory of fatigue. *The Trend in Engineering*, *13*, 9–14.

Pedregal, D. J., & Carnero, M. C. (2006). State space models for condition monitoring: A case study. *Reliability Engineering and System Safety*, *91*, 171–180.

Peng, Y., & Dong, M. (2011). A prognosis method using age-dependent hidden semi-Markov model for equipment health prediction. *Mechanical Systems and Signal Processing*, *25*(1), 237–252.

Peng, Y., Dong, M., & Zuo, M. J. (2010). Current status of machine prognostics in condition-based maintenance: A review. *International Journal of Advanced Manufacturing Technology*, *50*, 297–313.

Pham, H. T., & Yang, B. S. (2010). Estimation and forecasting of machine health condition using ARMA/GARCH model. *Mechanical Systems and Signal Processing*, *24*, 546–558.

Qian, N. (1999). On the momentum term in gradient descent learning algorithms. *Neural Networks*, *12*, 145–151.

Ragab, A., Yacout, S., Ouali, M. S., & Osman, H. (2016). Prognostics of multiple failure modes in rotating machinery using a pattern-based classifier and cumulative incidence functions. *Journal of Intelligent Manufacturing*,. doi:10.1007/s10845-016-1244-8.

Sarkeyli, A., Zain, A. M., & Sharif, S. (2015). A multi-performance prediction model based on ANFIS and new modified-GA for machining processes. *Journal of Intelligent Manufacturing*, *26*, 703–716.

Schon, T. B., Wills, A., & Ninness, B. (2011). System identification of nonlinear state-space models. *Automatica*, *47*(1), 39–49.

Storvik, G. (2002). Particle filters for state-space models with the presence of unknown static parameters. *IEEE Transactions on Signal Processing*, *50*(2), 281–289.

Teti, R., Jemielniak, K., O'Donnell, G., & Dornfeld, D. (2010). Advanced monitoring of machining operations. *CIRP Annals-Manufacturing Technology*, *59*, 717–739.

Vogl, G. W., Weiss, B. A., & Helu, M. (2016). A review of diagnostic and prognostic capabilities and best practices for manufacturing. *Journal of Intelligent Manufacturing*,. doi:10.1007/s10845-016-1228-8.

Wang, J., & Gao, R. X. (2013). Multiple model particle filtering for bearing life prognosis. In *Proceedings of 2013 IEEE International Conference on Prognostics and Health Management (PHM 2013)* (pp. 1–6). Gaithersburg, Maryland, USA, June 24–27.

Wang, J., Gao, R. X., & Yan, R. (2014). Multi-scale enveloping order spectrogram for rotating machine health diagnosis. *Mechanical Systems and Signal Processing*, *46*(1), 28–44.

Zhang, Z., Wang, Y., & Wang, K. (2013). Fault diagnosis and prognosis using wavelet packet decomposition, Fourier transform, and artificial neural network. *Journal of Intelligent Manufacturing*, *24*, 1213–1227.

Zhao, Z., Huang, B., & Liu, F. (2013). Parameter estimation in batch process using EM algorithm with particle filter. *Computers and Chemical Engineering*, *57*, 159–172.

Zio, E., & Peloni, G. (2011). Particle filtering prognostics estimation of the remaining useful life of nonlinear components. *Reliability Engineering and System Safety*, *96*, 403–409.