

# Fault Detection and Remaining Useful Life Estimation Using Switching Kalman Filters

Reuben Lim and David Mba

**Abstract** The use of condition monitoring (CM) data in degradation modeling for fault detection and remaining useful life (RUL) estimation have been growing with increasing use of health and usage monitoring systems. Most degradation modeling methods requires fault detection thresholds to be established. When the CM measure exceeds the detection threshold, RUL prediction is then performed using a time-invariant dynamical model to represent the degradation path to the failure threshold. Such approaches have some limitations as detection thresholds can vary widely between individual units and a single dynamical model may not adequately describe a degradation path that evolves from slow to accelerated wear. As such, most degradation modeling studies only focuses on segments of their CM data that behaves close to the assumed dynamical model. In this paper, the use of Switching Kalman Filters (SKF) is explored for both fault detection and remaining useful life prediction under a single framework. The SKF uses multiple dynamical models describing different degradation processes from which the most probable model is inferred using Bayesian estimation. The most probable model is then used for accurate prediction of RUL. The proposed SKF approach is demonstrated to track different evolving degradation path using simulated data. It is also applied onto a gearbox bearing dataset from the AH64D helicopter to illustrate its application in a practice.

**Keywords** Switching kalman filter • Fault detection • Degradation modeling • Remaining useful life

---

R. Lim (✉)

Senior Engineer, Republic of Singapore Air Force, Singapore, Singapore  
e-mail: r.limchikeong@cranfield.ac.uk

D. Mba

Head of Turbo-Machinery and Icing Group, Cranfield University, Bedford , UK  
e-mail: d.mba@cranfield.ac.uk

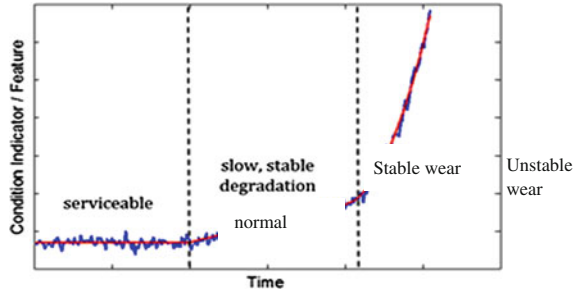
## 1 Introduction

The prevalence of health and usage monitoring systems (HUMS) on aircrafts in the past decades has fuelled the growth of using condition monitoring (CM) data in degradation modeling for health assessment of critical systems [1]. A wide variety of approaches in use of CM data for degradation modeling were comprehensively reviewed by different authors [1–4]. These data driven approaches are broadly classified into three main categories namely, physics-based methods, artificial intelligence methods and statistical methods. Physics-based method uses deterministic model of the system and can be very complex to develop. Artificial intelligence methods such as neural networks and support vector machines can handle highly non-linear problems but it requires huge number of training data which is often not available in practice. Amongst the three, statistical method is the most widely used in industry where conventional statistical process control and trend extrapolation are most commonly applied [3]. Advanced statistical methods such as Hidden Markov Models and cluster analysis can classify faults better but not widely used in practice due again to unavailability of training data. Most of the applications in the literature used experimental or simulated data for model training and little work was done with fielded applications [4]. In this paper, the switching Kalman filters (SKF) is investigated for fault detection and remaining useful life (RUL) of rolling element bearing and applied to both simulated and actual CM data gathered from AH64D helicopter.

## 2 Literature Review

The Kalman filter is a stochastic filtering process, which recursively estimates the state of a dynamic system in the presence of measurement noise and process noise, by minimizing the mean squared error [4]. The Kalman filter has been a widely applied concept in navigation and is also used in fields such as signal processing and econometrics. The Kalman Filter requires less training data compared to other statistical and AI techniques as it relies mainly on individual system's measurement data. However, the dynamical behavior of the system is required and be represented as a state-space model. In prognostic application, the Kalman filter was applied to predict the RUL of electrical connections [5] and electrolytic capacitors [6]. In these applications, the Kalman filter was used to adaptively track changes in the degradation process of the system and the dynamical model describing the degradation process was assumed to be time invariant. However, the degradation process in components can be uncertain and evolve over time as seen in bearing wear tests [7]. For example, in Fig. 1, the vibration measurement of a serviceable bearing can be stationary with measurement noise. When slow stable wear from damage such as surface pitting occurs, the vibration can gradually rise as a linear function. When accumulated damage is severe and unstable, the vibration rises rapidly in higher order functions.

**Fig. 1** Evolution of degradation process across time



As such, a single dynamical model may not adequately represent the different degradation processes. Consequently, this can cause predictions to diverge or fluctuate depending on whether the degradation process is under or over-fitted. This constraint is often seen in works [8, 9] where only measurements above an established threshold are considered in the analysis as those below does not behave according to the assumed dynamical model. For such problems, SKF can track the dynamics of the degradation process as it changes. RUL prediction is then performed based on the most probable dynamical model representing the degradation process. To do this, the SKF consists of multiple linear state-space models; like the basic Kalman filter, and it can switch between these models through a weighted combination across time. It is popularly used to track multiple moving targets but has also been applied in meteorology [10] and econometric [11]. SKF is applied here to track the different bearing degradation processes shown in Fig. 1. By tracking the dynamical behavior of different degradation processes, fault detection can be performed without using pre-established detection thresholds. It also helps maintainers to predict RUL more accurately by distinguishing between stable and unstable wear and performing prediction only when unstable wear is detected.

### 3 Background

This section provides a brief review of the Extended Kalman filter, dynamic Bayesian network, SKF and their application towards fault detection and RUL estimate of rolling element bearing.

#### 3.1 Extended Kalman Filter

As mentioned, the Kalman filter recursively estimates the state mean and covariance of a linear process by minimizing the mean square error. The Extended Kalman filter (EKF) is a non-linear extension which uses linear approximation of the non-linear function to estimate the state mean and covariance [12, 13].

The linear approximation performed through first and second-order Taylor series expansion of the non-linear function is most commonly used and the first-order is adopted here. The discrete state-space model describing a non-linear process is given by:

$$x_t = f(x_{t-1}) + q_{t-1}, y_t = h(x_t) + r_t \quad (1)$$

where  $x_t$  is the true but hidden state of the system and  $y_k$  is the observable measurement of the state.  $f(\cdot)$  is the fundamental matrix describing the system dynamics and  $h(\cdot)$  is the measurement matrix and both are functions assumed to be continuously differentiable.  $q_{t-1} \sim N(0, Q_t)$  is the process noise and  $r_{t-1} \sim N(0, R_t)$  is the measurement noise. The EKF estimates the value of  $x_t$ , given the measurement,  $y_t$  by filtering out the noises. This is carried out using the ‘Prediction’ and ‘Update’ steps also known as the Riccati Equations [13] are shown as follows.

Prediction Step:

$$\begin{aligned} \text{Predicted state estimate : } \hat{x}_t &= f(x_{t-1}, t-1) \\ \text{Predicted estimate covariance } \hat{P}_t &= F(x_{t-1}, t-1)P_{t-1}F'(x_{t-1}, t-1) + Q_{t-1} \end{aligned} \quad (2)$$

Update Step:

$$\begin{aligned} \text{Measurement residual : } v_t &= y_t - h(\hat{x}_{t-1}, t) \\ \text{Residual covariance } C_t &= H(\hat{x}_t, t)\hat{P}_tH'(\hat{x}_t, t) + R_t \\ \text{Kalman Gain } K_t &= \hat{P}_tH'(\hat{x}_t, t)C_t^{-1} \\ \text{Updated state estimate } x_t &= \hat{x}_t + K_tv_t \\ \text{Updated estimate covariance } P_t &= (I - K_tH(\hat{x}_t, t))\hat{P}_t \end{aligned} \quad (3)$$

where  $F(\cdot)$  and  $H(\cdot)$  are the Jacobians of  $f(\cdot)$  and  $h(\cdot)$  are given by

$$F(x_{t-1}, t-1) = \left. \frac{\partial f(x_{t-1}, t-1)}{\partial x} \right|_{\hat{x}_{t-1}|t-1}, H(\hat{x}_t, t) = \left. \frac{\partial h(x_t, t)}{\partial x} \right|_{\hat{x}_t|t-1}, \quad (4)$$

### 3.2 Switching Kalman Filter

The switching Kalman filter may be represented as a dynamic Bayesian network. In each time step, both the model switch variable,  $S_t$  and state variable,  $x_t$  are hidden and have to be inferred from the observations,  $y_t$ . For a system with multiple dynamics which are described with  $n$  Kalman filters, the size of the belief state will increase exponentially at each time step to  $n^t$ . As such, inferring the probability of every state at each time step becomes intractable. To overcome this problem,

approximation method like the Generalised Pseudo Bayesian (GPB) algorithm as described in [12] was adopted. In each time step, the state and covariance estimates from all the filters in the previous time step are combined with weights assigned according to the mix probabilities of the model switch variable,  $S_t^{ij}$  and the model transition probability,  $Z_{ij}$  as shown in Eqs. (5) and (6).

$$\text{Model switching probabilities: } S_t^{ij} = \frac{Z_{ij}S_{t-1}^i}{\sum_{i=1}^n Z_{ij}S_{t-1}^i} \quad (5)$$

Weighted state and covariance estimates:

$$\tilde{x}_{t-1}^j = \sum_{i=1}^n S_t^{ij} x_{t-1}^i, \quad \tilde{P}_{t-1}^j = \sum_{i=1}^n S_t^{ij} \left\{ P_{t-1}^i + [x_{t-1}^i - x_{t-1}^j][x_{t-1}^i - x_{t-1}^j]' \right\} \quad (6)$$

with the weighted state and covariance estimates, the usual Kalman filter as shown in Eqs. (2) and (3) is carried out for each filter model with each yielding a predicted state,  $\hat{x}_{t-1}^j$  and covariance,  $\hat{P}_{t-1}^j$  estimate. The likelihood of each filter is then determined with Eq. (7) using their measurement residual,  $v_t^j$ . The probability of each model at the current time step can then be obtained as shown in Eq. (8). The weighted state and covariance estimate update for the current time can also be determined using Eq. (9). A detailed description of SKF is available in [14] and a good demonstration of SKF with use of GPB is shown in [15].

$$\text{Likelihood of filter from measurement residual: } L_t^i = N(v_t^i; 0, C_t^i) \quad (7)$$

Probability of each model:

$$S_t^i = \frac{L_t^i (\sum_{i=1}^n Z_{ij} S_{t-1}^i)}{\sum_{i=1}^n (L_t^i \sum_{i=1}^n Z_{ij} S_{t-1}^i)} \quad (8)$$

The weighted state and covariance estimate update are computed as follows:

$$x_t = \sum_{i=1}^n S_t^i x_t^i, \quad P_t = \sum_{i=1}^n S_t^i \left\{ P_t^i [x_t^i - x_k] [x_{t-1}^i - x_t]^' \right\} \quad (9)$$

## 4 SKF Formulation for Tracking Varying Degradation Processes

In this analysis, it is assumed that component degradation is monotonically increasing and it evolves from normally operating to stable wear and then unstable wear. For bearings, a linear, polynomial or exponential model is used to describe

the different trends in the vibration-based degradation measure [16–18]. A Kalman filter is built for each of them and they are used together in the SKF. For the exponential filter, extended Kalman filter is applied due to its non-linear form. The state transition  $F_{i,t}$  is obtained from the Jacobian of the state equations using Eq. (4). It is assumed that the process noise entering the system only consists of zero mean white noise  $q_a$  and  $q_b$  which models the wear rate parameters  $a_t$  and  $b_t$  stochastically. The state, transition and process noise covariance for each filter are shown below with subscripts 1, 2 and 3 denoting the zero, first order and exponential Kalman filters respectively.

Zero Order polynomial model (Normal Operation)

$$\begin{aligned}
 \text{State : } x_t &= x_{t-1} \\
 \text{State Transition : } F_{1,t} &= 1 \\
 \text{Process Noise : } Q_{1,t} &= 0, y_t = x_t + r_t \\
 \text{Measurement : } H_{1,t} &= 1
 \end{aligned} \tag{10}$$

1st Order polynomial model (Stable Wear)

$$\begin{aligned}
 \text{State : } x_t &= x_{t-1} + a_{t-1}\Delta t, a_t = a_{t-1} + q_a \\
 \text{State Transition : } F_{2,t} &= \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \\
 \text{Process Noise : } Q_{2,t} &= \begin{bmatrix} 0 & 0 \\ 0 & q_a \end{bmatrix} \\
 \text{Measurement : } y_t &= x_t + r_t, H_{2,t} = [1 \ 0]'
 \end{aligned} \tag{11}$$

Exponential model (Unstable Wear)

$$\begin{aligned}
 \text{State : } x_t &= x_{t-1}e^{b_{t-1}\Delta t}, b_t = b_{t-1} + q_b \\
 \text{State Transition : } F_{3,t} &= \begin{bmatrix} e^{b_{t-1}\Delta t} & x_{t-1}\Delta te^{b_{t-1}\Delta t} \\ 0 & 1 \end{bmatrix} \\
 \text{Process Noise : } Q_{3,t} &= \begin{bmatrix} 0 & 0 \\ 0 & q_b \end{bmatrix} \\
 \text{Measurement : } y_t &= x_t + r_t, H_{3,t} = [1 \ 0]' \\
 \text{Model transition matrix : } Z &= \begin{bmatrix} 0.99 & 0.005 & 0.005 \\ \sim 0 & 0.99 & 0.01 \\ \sim 0 & \sim 0 & \sim 1 \end{bmatrix}
 \end{aligned} \tag{12}$$

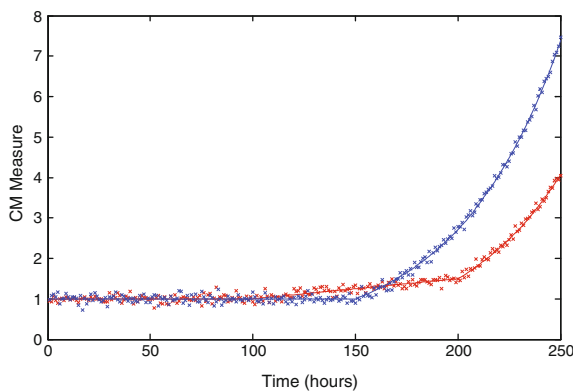
Initial model probabilities, state and covariance estimate:

$$S_0 = [0.98 \quad 0.01 \quad 0.01], x_0 = y_0, a_0 = 0, b_0 = 0, P_0 = I \tag{14}$$

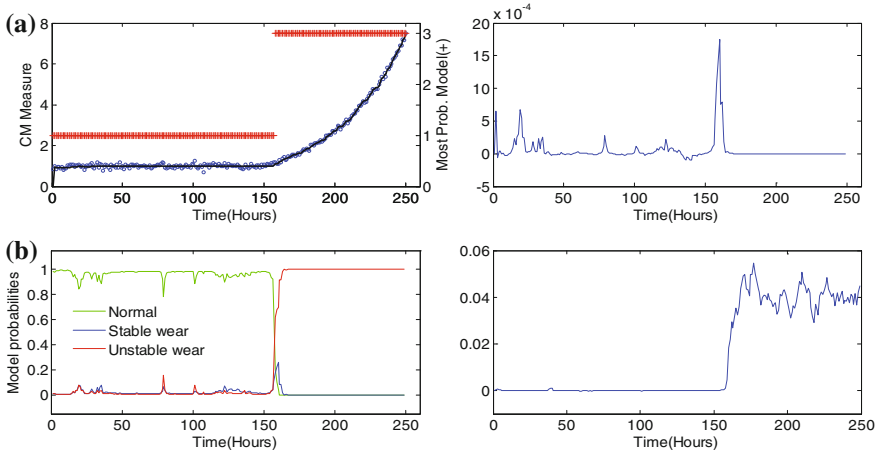
For the SKF, the state transition matrix  $Z$  is set such that the system tends to remain in its own state with  $Z_{ii} \sim 1$ . It is also assumed that the degradation rate can only progress i.e. from normal to stable and unstable degradation but not the reverse. However,  $Z_{ij}$  is assigned a value approximately zero for  $i > j$  as a value of zero can cause underflow problems in Eq. (8) when implemented as a software program. The initial model probability,  $S_0$  is set with high probability that its in normal condition. The initial state estimate,  $x_0$  is initialized to the first measurement and initial parameters  $a_0$  and  $b_0$  are zero. The initial covariance matrix,  $P_0$  is set arbitrarily with an identity matrix,  $I$ .

### 5 Diagnostics of Evolving Degradation Processes Using Simulated Data

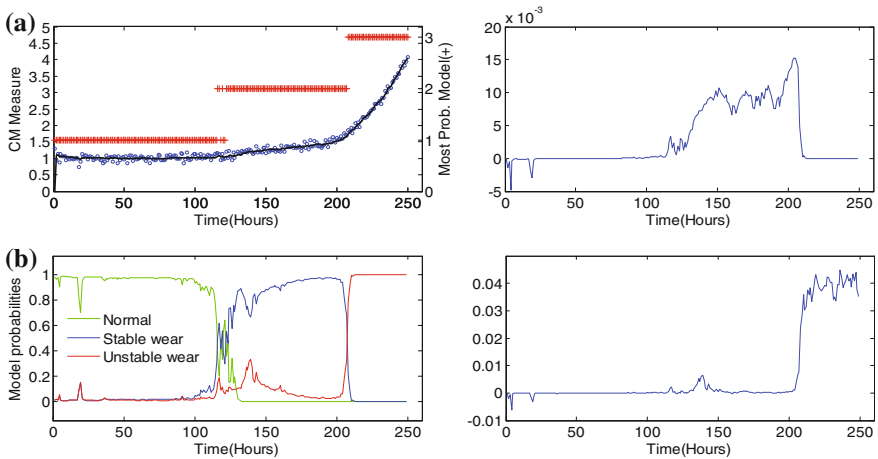
The SKF approach to track the degradation processes is demonstrated here using simulated data. Figure 2 shows different evolving degradation processes; (1) normally operating to unstable wear at  $t = 150$  h and (2) normally operating to stable wear at  $t = 100$  h and then unstable wear at  $t = 200$  h. The simulated degradation measurements are generated using the measurement equations from Eqs. (11–13). An additive measurement noise,  $r \sim N(0, 0.08^2)$  is added all three processes. For stable wear, a wear rate parameter,  $a = 0.01$  is adopted with process noise,



**Fig. 2** Simulated degradation processes with measurement and process noise: (1) normally operating to unstable wear at  $t = 150$  h and (2) normally operating to stable wear at  $t = 100$  h and unstable wear at  $t = 200$  h



**Fig. 3** Normal to unstable wear (*Top left*) Filtered state and most probable model (*Bottom left*) Model probabilities, (*Top and bottom right*) Estimated parameters  $a_t$  and  $b_t$



**Fig. 4** Normal to stable and unstable wear (*Top left*) Filtered state and most probable model, (*Bottom left*) Model probabilities, (*Top and bottom right*) Estimated parameters  $a_t$  and  $b_t$

$q_a \sim N(0, 0.001^2)$ . For unstable wear, a wear rate parameter,  $b = 0.04$  is adopted with process noise,  $q_b \sim N(0, 0.004^2)$ .

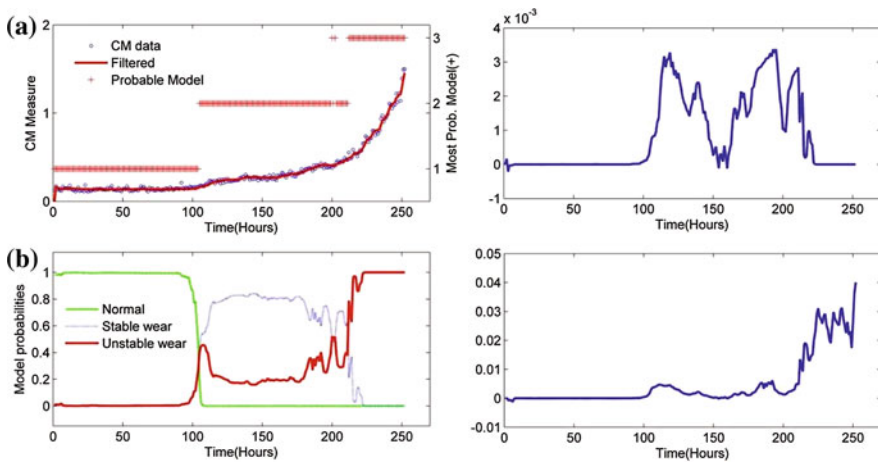
The ideal case where the dynamical models of the degradation processes and their measurement and process noise are known is shown here. Figures 3 and 4 shows the SKF results in tracking the evolving degradation processes. It can be seen that the SKF is able to track and estimate the most probable degradation process well using the dynamical behavior of the measurement. For normal to unsteady



wear, the SKF detects the change at 158 h compared to 150 h. For normal to steady and then unsteady wear, the SKF detects the change at 116 h and 208 h compared to 100 h and 200 h respectively. The SKF lags behind the actual transition times as it is performing the estimation in real-time and requires adequate measurements from the dynamical process. In addition, it can estimate the wear rate parameters,  $a$  and  $b$  well at  $\sim 0.001$  and  $\sim 0.04$ . It should be noted that the estimation will not converge towards the exact parameter value due to inherent noise added to the measurements.

## 6 Case Study on AH64D Helicopter Tail Rotor Gearbox Bearing

The SKF approach is applied to vibration CM data from the AH64D Tail Rotor Gearboxes (TRGB) in a practical scenario. The bearing CM data and results from the SKF is shown in Fig. 5. The measurement error,  $r = 3.2e-4$  is obtained by taking the variance of the stationary measurements when the TRGB is in a good condition and this can vary between individual gearboxes. The process error,  $q_s$  contains the uncertainty of the filters in modeling the real world [13]. It is obtained by tuning the SKF model with similar defect cases and is assumed to be the same across gearbox bearings. The SKF formulations are applied with  $q_s$  set initially as a small percentage of the measurement error,  $R$ . The SKF model is then applied on the CM data and  $q_s$  is tuned till the model is acceptably consistent yet responsive to changes in the degradation processes. In this study,  $q_s = 5e-8$  is obtained by tuning the model using CM data from other TRGB with similar failure. Form Fig. 5, it can



**Fig. 5** TRGB CM data (*Top left*) filtered state and most probable model, (*Bottom left*) Model probabilities, (*Top and bottom right*) Estimated parameters  $a_t$  and  $b_t$

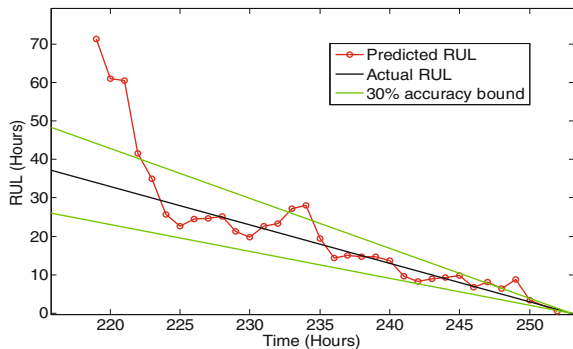
be seen that the SKF can adaptively track the different bearing degradation processes with the process noise tuned from other gearboxes. However, when the CM measurements are not increasing monotonically at  $\sim 200$  h, the SKF has to take a longer time before it converges. Instead of relying on the absolute value of the CM measurements, the SKF uses the dynamic behavior between the current and past measurement to diagnose the degradation state. Therefore, it is not dependent on a fixed threshold which are typically derived from statistical evaluation of large numbers of past failure cases. Another key advantage of this technique for diagnosis is that it provides the probability of which degradation process the bearing is in. In comparison, the widely used, statistical process control (SPC) approach only triggers when the measurement is above a statistical limit and no further information is available. The quantitative probability measure from the SKF allows more support for maintenance engineers as the probabilities of the bearing conditions can be compared in the event of an outlier measurement.

## 7 Prediction of Remaining Useful Life

The SKF infers the most probable dynamic model to be applied at each time step for prediction and the RUL of the bearing is predicted whenever an unsteady wear is detected. The RUL is predicted by propagating the weighted state and covariance estimates obtained from Eq. (8) at each time step using Eq. (2) and determining the time when the degradation state crosses the failure threshold. The  $\alpha$ - $\lambda$  metric [19] is applied to evaluate the performance of this prognostic evaluation as shown in Fig. 6.

The  $\alpha$ - $\lambda$  metric compares the actual RUL to the predicted RUL with converging  $\alpha$  bounds that provides an accuracy region. The  $\alpha$  bounds are application specific and a prediction is correct if it falls within the alpha bounds. From Fig. 6, the prognostic algorithm performs well as its accuracy improves quickly with time within the 30 % bounds. However, there are points on the RUL trajectory that lies outside the accuracy zone towards the end of useful life which is a behavior

**Fig. 6**  $\alpha$ - $\lambda$  performance metric using 30 % accuracy bounds



reportedly observed in [20] as well. This behavior could be attributed to unsteady vibration levels as the accumulated damage in the bearing becomes sizeable and could perhaps be addressed by lowering the failure threshold limit. Besides the RUL estimate, most of the lower confidence bound, which is important for conservative estimate of the RUL prediction are close to the lower 30 % accuracy bound as well.

## 8 Conclusion

In this study, the use of SKF is applied for fault detection and RUL estimation. The method is applied to both simulated data and actual helicopter gearbox bearing with promising results. The SKF model allows for degradation processes to evolve through time from which the underlying dynamical process would be inferred accordingly. The advantages of this approach are that it does not depend on a fixed threshold for fault detection and it can model the different degradation processes as they evolve. This approach also provides maintainers with more information for decision-making as a probabilistic measure of the state of bearing degradation is available. From the prognostic performance metric, it was shown that the RUL estimates have high accuracy when it is inferred that the degradation process is likely to be unstable. This in turn can provide maintainers with higher confidence on the predicted RUL for maintenance planning. A drawback of this method is that it requires frequent acquisition of measurement for the filter estimation which may not be readily available in practice.

## References

1. Gorjian N, Ma L, Mittinty M, Yarlalagadda P, Sun Y (2009) A review on degradation models in reliability analysis. In: Proceedings of the 4th World Congress on Engineering Asset Management. Springer, Athens
2. Heng A, Zhang S, Tan ACC, Mathew J (2009) Rotating machinery prognostics: State of the art, challenges and opportunities. *Mech Syst Signal Process* 23(3):724–739
3. Jardine AKS, Lin D, Banjevic D (2006) A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mech Syst Signal Process* 20(7):1483–1510
4. Sikorska JZ, Hodkiewicz M, Ma L (2011) Prognostic modelling options for remaining useful life estimation by industry. *Mech Systems Signal Process* 25(5):1803–1836
5. Lall P, Lowe R, Goebel K (2010) Prognostics using Kalman-Filter models and metrics for risk assessment in BGAs under shock and vibration loads. In: Proceedings 60th Electronic Components and Technology Conference (ECTC), 2010 pp. 889
6. Jose RC, Chetan K, Gautam B, Sankalita S, Kai G (2011) A Model-based Prognostics Methodology for Electrolytic Capacitors Based on Electrical Overstress Accelerated Aging. In: Annual Conference of the Prognostics and Health Management Society
7. Kotzalas MN, Harris TA (2001) Fatigue failure progression in ball bearings. *J Tribol* 123 (2):238–242

8. Gebraeel N (2006) Sensory-updated residual life distributions for components with exponential degradation patterns. *Autom Sci Eng IEEE Trans* 3(4):382–393
9. Roulias D, Loutas TH, Kostopoulos V (2012) A hybrid prognostic model for multistep ahead prediction of machine condition. *J Phys: Conf Series*. 364(1):012081
10. Manfredi V, Mahadevan S, Kurose J (2005) Switching kalman filters for prediction and tracking in an adaptive meteorological sensing network. In: *Sensor and Ad Hoc Communications and Networks, 2005. IEEE SECON 2005. 2005 Second Annual IEEE Communications Society Conference on*, pp 197
11. Lim Y, Cheng S (2012) Knowledge-driven autonomous commodity trading advisor. 2012 *IEEE/WIC/ACM International Conference on Intelligent Agent Technology, Macau*
12. Särkkä S (2011) Optimal Filtering. In: *Bayesian Estimation of Time-Varying Systems: Discrete-Time Systems* pp 36–38
13. Zarchan P, Musoff H (2005) Polynomial Kalman filters. In: *Fundamentals of Kalman Filtering: A Practical Approach*, 2nd ed, American Institute of Aeronautics and Astronautics, USA, pp 156
14. Kevin M (1998) Learning switching Kalman filter models, 98-10. Compaq Cambridge Research Lab Tech Report
15. John Q SKF Demo. <http://www.cit.mak.ac.ug/staff/jquinn/software.html> (accessed October/21)
16. Gebraeel N, Lawley M, Liu R, Parmeshwaran V (2004) Residual life predictions from vibration-based degradation signals: a neural network approach. *Indus Elect IEEE Trans* 51(3):694–700
17. Shao Y, Nezu K (2000) Prognosis of remaining bearing life using neural networks. In: *Proceedings of the Institution of Mechanical Engineers, Part I. J Syst Control Eng* 214(3):217
18. Sutrisno E, Oh H, Vasan A. SS, Pecht M (2012) Estimation of remaining useful life of ball bearings using data driven methodologies. In: *Prognostics and Health Management (PHM), 2012 IEEE Conference on*, pp 1
19. Saxena A, Celaya J, Balaban E, Goebel K, Saha B, Saha S, Schwabacher M (2008) Metrics for evaluating performance of prognostic techniques. In: *Prognostics and Health Management, 2008. PHM 2008. International Conference on*, pp 1
20. Saxena A, Celaya J, Saha B, Saha S, Goebel K (2009) On applying the prognostic performance metrics. In: *International Conference on Prognostics and Health Management (PHM)*