Condition monitoring and remaining useful life prediction using switching Kalman filters

Reuben Lim*
Republic of Singapore Air Force,
Air Engineering and Logistic Department,
506 Airport Road, #02-01,
534395 Singapore
E-mail: lchikeon@starnet.gov.sg
*Corresponding author

David Mba
School of Engineering,
Department of Power and Propulsion Building,
Cranfield University,
52, Cranfield, Beds MK43 0AL, UK
E-mail: D.Mba@cranfield.ac.uk

Abstract: The use of condition monitoring (CM) data to infer degradation state and remaining useful life (RUL) prediction has grown with increasing use of health monitoring systems. Most degradation modelling requires a detection threshold to be established and can only model a single dynamical behaviour for the degradation. Such approaches have limitations as detection thresholds can vary widely and a single model may not adequately describe a degradation path as it evolves. In this paper, the switching Kalman filter (SKF) is adopted. The SKF uses multiple dynamical models describing different degradation processes and the most probable model is inferred using Bayesian estimation. The advantages are that it does not depend on a fixed threshold for fault detection and can model the different degradation processes as they evolve. The SKF approach is applied to CM data from helicopter gearbox bearings and is shown as a promising tool for maintenance decision-making.

Keywords: switching Kalman filter; SKF; extended Kalman filter; EKF; rolling element bearing; fault detection; degradation modelling; remaining useful life; RUL.


Biographical notes: Reuben Lim is a Senior Engineer from the Republic of Singapore Air Force (RSAF) and is currently a research student in Cranfield University, Department of Power and Propulsion. His research interests include diagnostics and prognostics of helicopter gearboxes, maintenance and reliability engineering. He holds a Bachelor’s in Mechanical Engineering and Master’s in Civil Engineering from the National University of Singapore (NUS). He is also a Chartered Engineer with the RAeS.
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 modelling for health assessment of critical systems (Gorjian et al., 2009). A wide variety of approaches in use of CM data for degradation modelling were comprehensively reviewed by different authors (Gorjian et al., 2009; Heng et al., 2009; Jardine et al., 2006; Sikorska et al., 2011). These data driven approaches are broadly classified into three main categories namely:

1 physics-based methods
2 artificial intelligence methods
3 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 require a huge number of training data which is often not available in practice. Amongst the three, statistical methods are the most widely used in industry where conventional statistical process control (SPC) and trend extrapolation are most commonly applied (Jardine et al., 2006). Advanced statistical methods such as hidden Markov models and cluster analysis can classify faults better but are 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 (Sikorska et al., 2011). In this paper, a recursive Bayesian estimation-based technique, the switching Kalman filter (SKF) is investigated for fault detection and remaining useful life (RUL) of rolling element bearing using actual CM data gathered from AH64D helicopters belonging to the Republic of Singapore Air Force (RSAF). The operational HUMS data from these helicopters are analysed and correlated with tear-down inspection findings from a small number of defective gearboxes.

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
minimising the mean squared error (Sikorska et al., 2011). 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 behaviour of the system is required and be represented as a state-space model. In a way, it can be seen as a hybrid between physics and statistical-based method. Another key advantage of the Kalman filter is that it does not need to store and re-process all past measurements at each time step thus it is suitable for real-time implementation. In prognostic application, the Kalman filter was applied to predict the RUL of ball grid array connections using resistance measurements (Lall et al., 2010). Kalman filtering was also used to estimate the RUL of electrolytic capacitors (Celaya et al., 2011). 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 (Kotzalas and Harris, 2001). For example in Figure 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, the vibration can rise much rapidly in higher order functions.

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 (Gebraeel, 2006; Roulias et al., 2012) 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, also known as linear dynamic models (Kevin, 1998) or interacting multiple models (IMM) (Bar-Shalom et al., 2001) can track the dynamics of the degradation process as it changes. RUL prediction can then be carried out 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 been applied in meteorology (Manfredi et al., 2005) and econometric (Lim and Cheng, 2012). In prognostics, similar concepts have been used to detect sensors and actuator failures (Zhang and Jiang, 2001) changes in non-linear stochastic systems (Wang and Syrmos, 2008). It has also been applied in bearing fault prognosis simulation to combine results from different prediction models (Christoph et al., 2012). SKF is applied in this paper to track the different bearing degradation processes shown in Figure 1. By tracking the dynamical behaviour 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. Although a fault detection threshold is no longer needed using this approach, a failure threshold is still required for RUL prediction. The rest of this paper is organised as follows. Section 3 gives background on the SKF. Section 4 describes the defects of the gearbox bearings and the extraction of the CM measure from the RSAF AH64D helicopters. Section 5 formulates the bearing problem using SKF and Section 6 and 7 discusses the results.
3 Background

This section provides a brief review of the extended Kalman filter (EKF), 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 minimising the mean square error. The EKF is a non-linear extension which uses linear approximation of the non-linear function to estimate the state mean and covariance (Särkkä, 2011; Zarchan and Musoff, 2005). 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:

\[
\begin{align*}
    x_t &= f(x_{t-1}) + q_{t-1} \\
    y_t &= h(x_t) + r_t
\end{align*}
\]

where \(x_t\) is the true but hidden state of the system and \(y_t\) 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 Ricatti equations (Zarchan and Musoff, 2005) are shown as follows.

**Prediction step:**

Predicted state estimate:  
\[
    \hat{x}_t = f(x_{t-1}, t-1)
\]

Predicted estimate covariance  
\[
    \hat{P}_t = F(x_{t-1}, t-1)P_{t-1}F'(x_{t-1}, t-1) + Q_{t-1}
\]
Update step:

Measurement residual: \[ v_t = y_t - h(\hat{x}_{t-1}, t) \]

Residual covariance \[ C_t = H(\hat{x}_t, t) \hat{P}_t H'(\hat{x}_t, t) + R_t \]

Kalman gain \[ K_t = \hat{P}_t H'(\hat{x}_t, t) C_t^{-1} \]

Updated state estimate \[ x_t = \hat{x}_t + K_t v_t \]

Updated estimate covariance \[ P_t = (I - K_t H(\hat{x}_t, t)) \hat{P}_t \]

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

\[
F(x_{t-1}, t-1) = \frac{\partial f(x_{t-1}, t-1)}{\partial x}
\]

\[
H(\hat{x}_t, t) = \frac{\partial h(x_t, t)}{\partial x}
\]

The EKF performs prediction of future states by repeating the prediction step in equation (1) using the last known state and covariance estimate but without updating the state and covariance estimate. However, the accuracy of the estimates will fall with increasing steps with the prediction confidence bounds widening accordingly.

3.2 Switching Kalman filter

The SKF may be represented as a dynamic Bayesian network as shown in Figure 2. 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.

Figure 2  Dynamic Bayesian network representation of a SKF

Source: Kevin (1998)

To overcome this problem, approximation method like the Generalised Pseudo Bayesain (GPB) algorithm as described in Särkkä (2011) 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, \( S_i^t \), and the model transition probability, \( Z_{ij} \) as shown in equations (5) and (6).

Model switching probabilities:

\[
S_i^t = \frac{Z_{ij} S_{i+1}^t}{\sum_{i=1}^n Z_{ij} S_{i+1}^t}
\]

(5)

Weighted state and covariance estimates:

\[
\tilde{x}_{t-1}^i = \sum_{i=1}^n S_i^t x_{t-1}^j
\]

(6)

With the weighted state and covariance estimates, the usual Kalman filter as shown in equations (2) and (3) is carried out for each filter model with each yielding a predicted state, \( \tilde{x}_{t-1}^i \) and covariance, \( \tilde{P}_{t-1}^i \) estimate. The likelihood of each filter is then determined with equation (7) using their measurement residual, \( v_t \). The probability of each model at the current time step can then be obtained as shown in equation (8). The weighted state and covariance estimate update for the current time can also be determined using equation (9). A detailed description of SKF is available in Kevin (1998) and a good demonstration of SKF with use of GPB is shown in John (2011).

Likelihood of filter from measurement residual:

\[
L_i = N(v_t; 0, C_i)
\]

(7)

Probability of each model:

\[
S_i^t = \frac{L_i \left( \sum_{i=1}^n Z_{ij} S_{i+1}^t \right)}{\sum_{i=1}^n \left( L_i \sum_{i=1}^n Z_{ij} S_{i+1}^t \right)}
\]

(8)

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

\[
\tilde{x}_{t} = \sum_{i=1}^n S_i x_{t}^i
\]

\[
\tilde{P}_{t} = \sum_{i=1}^n S_i \left[ P_i^t \left( x_{t}^i - x_{t-1}^i \right) \left( x_{t}^i - x_{t-1}^i \right)^T \right]
\]

(9)

4 AH64D tail gearbox bearing HUMS data and CM extraction

The SKF method is applied here to detect faults in the tail rotor gearbox (TRGB) output shaft thrust bearing of the AH64D helicopter. The TRGB is grease lubricated single stage gearbox and serves to transmit drive torque from the intermediate gearbox to the tail rotor system. From maintenance records, three TRGB were found with widespread pitting on ball elements and spalling in the bearing races as shown in Figure 3. The HUMS data for the TRGBs, which consists of fast Fourier transformed acceleration spectrum, were downloaded for analysis. From the acceleration spectrum of the TRGB, vibration energy
can be observed to rise at the bearing ball pass frequency inner race (BPFI) (Randall, 2011) as shown in Figure 4. The root-mean-square acceleration, $G_{rms}$ of the frequency band surrounding the BPFI and its harmonics was extracted as the CM measurement and the plots from the three TRGBs are shown in Figure 5. A general trend of stationary, linear and then exponential rise can be seen across the TRGBs but the rate and duration in each stage differs between individual gearboxes.

**Figure 3** Severe pitting of rolling elements and spalling of bearing races (see online version for colours)

**Figure 4** TRGB acceleration spectrum plot showing rise in vibration energy (see online version for colours)

**Figure 5** Extracted CM measurement plot for the three TRGBs (see online version for colours)
As the actual time at which the bearing degradation stages transit cannot be physically observed, they are inferred using piece-wise or segmented regression (Ritzema, 1994) on the complete CM data history as shown in Figure 6. The CM history is segmented into the three degradation stages with two transition times or ‘breakpoints’. A zero, first and second-order polynomial regression is then applied to the segmented CM data for each degradation stage and the sum of their residuals is obtained. This procedure is repeated iteratively with different sets of transition times and the optimal piece-wise regression fit is obtained from the set of transition time that minimises the sum of the residual. Figure 6 shows the transition times for the degradation stages are at 87 hrs and 202 hrs for TRGB 3. Although these may not be the real transition time, the piece-wise regression provides the optimal fit for each stage of the degradation process. In practice, the complete CM history will not be available thus the SKF is used as a decision support tool to diagnose the different stages of bearing degradation. When the degradation is diagnosed to be fast and unstable, the RUL of the gearbox bearing is then estimated.

**Figure 6** Degradation stages for TRGB 3 inferred from piece-wise regression (see online version for colours)

---

5 **SKF formulation for tracking bearing degradation process**

As mentioned earlier, the degradation processes has to be known for the SKF method to be applied. For bearing degradation, it is well established that its vibration-based signals is stationary when functioning normally and exhibits linear or exponential trends when it degrades (Gebraeel et al., 2004; Shao and Nezu, 2000; Sutrisno et al., 2012). It is then assumed that the bearing degradation is monotonically increasing and it can evolve from normally operating to stable wear and then unstable wear or from normal operating to unstable wear directly. The dynamics of these processes are represented using a zero and first order polynomial and an exponential model respectively. A Kalman filter is built for each of them and they are used together in the SKF. For the exponential filter, EKF is applied due to its non-linear form. The state transition $F(t)$ is obtained from the Jacobian of the state equations using equation (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)

State: \( x_t = x_{t-1} \)

State transition: \( F_{1,t} = 1 \)  

Process noise: \( Q_{1,t} = 0 \)

Measurement: \( y_t = x_t + r_t \)

1st order polynomial model (stable wear)

State: \( x_t = x_{t-1} + a_{t-1} \Delta t \)
\( a_t = a_{t-1} + q_a \)

State transition: \( F_{2,t} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \)  

Process noise: \( Q_{2,t} = \begin{bmatrix} 0 & 0 \\ 0 & q_a \end{bmatrix} \)

Measurement: \( y_t = x_t + r_t \)

Exponential model (unstable wear)

State: \( x_t = x_{t-1} e^{a_{t-1} \Delta t} \)

State transition: \( F_{3,t} = \begin{bmatrix} e^{a_{t-1} \Delta t} & x_{t-1} \Delta t e^{a_{t-1} \Delta t} \\ 0 & 1 \end{bmatrix} \)

Process noise: \( Q_{3,t} = \begin{bmatrix} 0 & 0 \\ 0 & q_b \end{bmatrix} \)

Measurement: \( y_t = x_t + r_t \)

Model transition matrix

\[
Z = \begin{bmatrix} 0.99 & 0.005 & 0.005 \\ -0 & 0.99 & 0.01 \\ -0 & -0 & 0.999 \end{bmatrix}
\]
Initial model probabilities, state and covariance estimate

\[
S_0 = [0.99 \ 0.005 \ 0.005] \\
x_0 = y_0, \\
a_0 = 0 \\
b_0 = 0 \\
P_0 = 1
\]

For the SKF, the state transition matrix \( Z \) is set such that the system tends to remain in its own state with \( Z_{ii} = 0.99 \) and \( \Sigma Z_{ij} = 0.01 \) for \( i < j \) to transit to other states. It is also assumed that the degradation rate can only progress i.e., from normal to stable or unstable degradation but not the reverse. However, \( Z_{ij} \) is assigned a value approximately zero as a value of zero can cause underflow problems in equation (8) when implemented as a software programme. A high value of \( Z_{ii} \sim 1 \) is set so that the filter only switches when the likelihood of the filter calculated from the measurement update is high. This would make the SKF less responsive but more stable to changes in the measurement trends. The measurement error, \( r \) 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 assumption of stationary measurements from normal operating TRGBs in this study is validated based on measurements from gearboxes of different helicopters. For TRGB 2, the measurement error, \( r_t \) obtained is 3.2e-4. The process noise, \( q_a \) and \( q_b \) contain the uncertainty of the filter in modelling the real world (Zarchan and Musoff, 2005). It is obtained by tuning the SKF model with past similar defect cases and is assumed to be the same across gearboxes. The SKF formulations are applied with \( q_a \) and \( q_b \) set as a small percent of the estimated or guessed value of the parameters (Bar-Shalom et al., 2001). One approach is to estimate the parameters by regression analysis of the segment of the CM data which is inspected to fit the parameter’s model. The SKF model is then simulated and the parameters tuned till the model is acceptably consistent. This is where training of the SKF is required from similar failure cases. Both the measurement and process errors are assumed to be Gaussian when EKF is applied (Särkkä, 2011). In Figure 5 however, a non-Gaussian error distribution can be observed between ~170 – 200 hrs in TRGB 1. In such cases, the use of particle filtering method (Särkkä, 2011) which allows for non-Gaussian errors may be explored for use with the SKF instead of the EKF. Although it is not ideal for TRGB 1, the EKF is applied here for its simplicity and reasonably good results can be obtained despite the Gaussian error assumption. The process noises were thus tuned using CM data from TRGB 1 and 2 before being tested on TRGB 3 where \( q_a \) and \( q_b \) were estimated to be 1e-7 and 1.5e-6 respectively. The initial model probability, \( S_0 \) is set to a serviceable TRGB bearing with stationary CM measurement. The initial state estimate, \( x_0 \) is initialised to the first measurement and initial parameters \( a_0 \) and \( b_0 \) are set to zero. The initial covariance matrix, \( P_0 \) is set with an identity matrix, \( I \). The initial state, parameters and covariance matrix can be set arbitrary as they will converge with measurement updates through the Kalman filtering. With these matrices and parameters defined, the SKF model is established.
6 Tracking of bearing degradation using SKF

The formulated SKF model is applied to the TRGB 3 bearing CM data and the results are shown in Figure 7. The SKF can adaptively track the different bearing degradation processes with the process noise tuned from the other two gearboxes. As the wear rate transits, the probability of the assumed model changes accordingly. Instead of relying on the absolute value of the CM measurements, the SKF uses the dynamic behaviour between the current and past measurement to diagnose the degradation state. Therefore, it is not dependent on a fixed threshold which have to be derived from statistical evaluation of large numbers of past failure cases. In Figure 7, it can be seen that $b_t$ stabilises as it approaches the failure time which is desirable as it allows a more accurate RUL prediction to be made. When compared to the optimal solution from the piece-wise regression, the SKF transition time lags behind at 105 hrs and 212 hrs compared to 87 hrs and 202 hrs. However, it should be noted that the SKF is performing the estimation in real-time and requires adequate measurements from the dynamical process.

Figure 7 (a) TRGB 3 CM data, filtered state and most probable model, (b) Model probabilities, (c and d) Estimated parameters $a_t$ and $b_t$ (see online version for colours)

When the CM measurements are not increasing monotonically at ~200 hrs, the SKF has to take a longer time before it converges as mentioned above. This may cause false alarms of unstable wear in practice and a simple way to reduce it is for maintainers to decide by evaluating the probabilities between stable and unstable wear. The probability of the degradation process is a key advantage as it provides more information. In comparison, the widely used SPC approach only triggers when the measurement is above
a statistical limit and no further information is available. In practice, maintenance engineers using SPC would then wait for future measurements to confirm if it is a fault or false alarm before prescribing any tasks. 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 RUL

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 equation (8) at each time step using equation (2) and determining the time when the degradation state crosses the failure threshold. The bearing is considered to have failed functionally when it reaches the failure threshold and it differs from the fault detection threshold which indicates presence of faults but the component is still functional. Figure 8 shows the RUL forecast at different times when the SKF detects unstable degradation after 210 hrs and it can be seen that the accuracy of the RUL estimate improves as time progresses.

Figure 8  RUL predictions at different times (see online version for colours)

The $\alpha$-$\lambda$ metric (Saxena et al., 2008) is applied to evaluate the performance of this prognostic evaluation as shown in Figure 9. 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 Figure 9, 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 behaviour reportedly observed in (Saxena et al., 2009) as well. This behaviour 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 Conclusions

In this study, the use of SKF as applied for fault detection and RUL estimates of a helicopter gearbox bearing is presented 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. In future work, the use of particle filtering-based method together with SKF will be explored so that the assumption of Gaussian noise can be relaxed.
References


