A Bayesian Approach to Condition Monitoring with Imperfect Inspections

Zhi-Sheng Ye†, Nan Chen‡, and Kwok-Leung TSUI§

†Department of Applied Mathematics, The Hong Kong Polytechnic University
‡Department of Industrial and Systems Engineering, National University of Singapore
§Department of Systems Engineering and Engineering Management, City University of Hong Kong

June 14, 2013

Abstract

Degradation is a common phenomenon for many products. Due to a variety of reasons, the degradation rates of units from the same population are often heterogeneous. In addition, when the degradation process is monitored using dedicated sensors, the measurements are often inaccurate due to various noisy factors. To account for the heterogeneous degradation rate and the non-negligible measurement errors, we model the degradation observations using a random-effects Wiener process with measurement errors. Under the model, direct estimation of current degradation and prediction of future degradation are difficult. We thus develop a filtering algorithm that recursively estimates the joint distribution of the degradation rate and the current degradation levels. Based on the estimates, the distribution of the remaining useful life can be timely predicted. Our method is both computational efficient and storage efficient. Its effectiveness is demonstrated through simulation and real data.
Keywords: Heterogeneous degradation rates, Measurement errors, Wiener process, Recursive filtering

1 Introduction

Degradation signals of a complex system reflect the system health status and are useful for condition monitoring and failure predictions. To help with the predictions, a proper model for the degradation signals is needed. A popular degradation model is the Wiener process with a positive drift because of its meaningful physical interpretations and nice mathematical properties [1]. The Wiener process has extensive applications in degradation data analysis [2, 3, 4, 5, 6], and prediction of remaining useful life (URL), e.g., see [7, 8, 9] for some recent references and see [10] for an overview. The Wiener process is also used in the maintenance scheduling [11, 12].

Most existing studies assume that the degradation level can be observed accurately. Nevertheless, many applications have seen imperfect inspections where the degradation measurements are contaminated with measurement errors. This is because the sensors used for degradation inspections often make use of electronic signals and are prone to non-negligible measurement errors [13]. If the errors from the imperfect inspections are not well taken into account, the estimates of parameters in the Wiener process may have large biases, which deteriorate the prediction accuracies of the future degradation and the RUL. Therefore, it is important to account for measurement errors when we analyze the condition monitoring data in such situations. As a result, imperfect inspection has attracted some attention recently in RUL prediction, e.g., see [14, 15, 16], among others. However, these studies did not consider possible heterogeneities in the degradation process among different units in the product population.

Because of the global supply and distribution channel, product units, even from the same batch of production, may have very different degradation characteristics. This heterogeneity
might be caused by the internal defects during production, differences in materials and components [17], diverse usage patterns and environmental conditions [18, 19, 20], etc. Any of these factors can lead to heterogeneities in the degradation rates among product units. To account for the heterogeneity in the degradation rate, random-effects variants of the Wiener process have been proposed in the literature [18, 21, 22]. The inherent randomness of the degradation rate captured by the random-effects parameter is aleatory in nature. In addition, systems requiring condition monitoring are often expensive. So we may not have enough historical data to accurately estimate the distribution for the random degradation rates. This further introduces epistemic uncertainties into the degradation rate of a specific system.

To capture both the aleatory and epistemic uncertainties, this study uses a Bayesian approach that models the degradation with heterogeneous degradation rates using the Wiener process with measurement errors. We impose a prior distribution on the degradation rate to account for the unit-to-unit heterogeneity and the statistical uncertainty. In the presence of the random degradation rate, direct estimation of the degradation rate and the degradation level is difficult, especially in the applications of real-time monitoring. We thereby propose an online filtering method to recursively estimate the underlying degradation state when noisy degradation measurements are observed sequentially. The algorithm is computational efficient with minimal storage requirement. Extensive numerical studies show that the proposed approach is effective, and the applications can be potentially expanded, e.g., to the scenarios considered in [23, 20, 24].

The remainder of the paper is organized as follows. Section 2 introduces the condition monitoring model and shows its basic properties. Section 3 discusses the filtering algorithm that updates the joint distribution of the current degradation levels and the unknown degradation rate. Distribution of the RUL can then be predicted. Section 4 considers the parameter estimation from historical data when degradation data from multiple systems are available. Sections 5 and 6 demonstrate the effectiveness of the proposed methods using
both simulated and real data. Section 7 concludes the paper.

2 Condition Monitoring with Imperfect Inspections

For a system under surveillance, we assume that the underlying degradation follows the Wiener process

\[ X(t) = \beta \Lambda(t) + \sigma \mathcal{B}(\Lambda(t)), \]  

(1)

where \( \beta \) is the drift rate parameter, \( \sigma \) is the volatility parameter, \( \Lambda(t) \) is the drift function and \( \mathcal{B}(t) \) is the standard Brownian motion. Because of the imperfect inspection, we can only observe a noisy degradation signal at each moment. Specifically, when an inspection is made at time \( t \), the observed degradation signal \( Y(t) \) is modeled as

\[ Y(t) = X(t) + \epsilon, \]  

(2)

where \( \epsilon \) denotes the inspection errors and is assumed to follow an \textit{i.i.d.} normal distribution \( \mathcal{N}(0, \gamma^2) \).

To account for the aleatory and epistemic uncertainties, the drift rate \( \beta \) in (1) is considered as random while the volatility parameter \( \sigma \) is assumed to be constant for all units. We assume \( \beta \) has a prior distribution determined by domain experts or inferred from historical data. A popular choice of the prior distribution is \( \beta \sim \mathcal{N}(\mu, \kappa^2) \), e.g., [2, 25, 21, 24]. For notational ease, let \( \boldsymbol{\theta} \) be the collection of all parameters in this Bayesian model.

When the system is inspected at times \( t_1, t_2, \cdots, t_n \) with observations \( \mathbf{Y} = [Y(t_1) = Y_1, Y(t_2) = Y_2, \cdots, Y(t_n) = Y_n]' \), we can obtain both the conditional distribution of \( (\mathbf{Y}|\beta) \) and the marginal distribution of \( \mathbf{Y} \) according to models (1) and (2). To take advantage of the independent increment property of \( X(t) \), we focus on the degradation increments. Define the increments of the inspection observations as \( \delta_1 = Y_1, \delta_j = Y_j - Y_{j-1}, j = 2, \cdots, n \). Similarly, we define \( \lambda_1 = \Lambda(t_1), \lambda_j = \Lambda(t_j) - \Lambda(t_{j-1}), j = 2, \cdots, n \). According to the
property of the Wiener process and the i.i.d. normal distributions of the measurement errors, 
\[ \mathbf{\delta} = [\delta_1, \delta_2, \ldots, \delta_n]' \] conditioning on \( \beta \) follows a multivariate normal distribution with a joint probability density function (PDF)

\[
f(\mathbf{\delta}|\beta) = (2\pi)^{-n/2} |\Sigma|^{-1/2} \exp \left[ -\frac{1}{2} (\mathbf{\delta} - \beta \mathbf{\lambda})' \Sigma^{-1} (\mathbf{\delta} - \beta \mathbf{\lambda}) \right],
\]

where \( \mathbf{\lambda} = [\lambda_1, \lambda_2, ..., \lambda_n]' \), and \( \Sigma \) is a positive definite matrix with the \((j,k)\)-th element given by

\[
\Sigma_{j,k} = \text{cov}(\delta_j, \delta_k|\beta) = \begin{cases} 
\sigma^2 \lambda_j + \gamma^2, & j = k = 1; \\
\sigma^2 \lambda_j + 2\gamma^2, & j = k > 1; \\
-\gamma^2, & |j - k| = 1; \\
0, & \text{otherwise}. 
\end{cases}
\]

When the prior distribution of \( \beta \) is \( \mathcal{N}(\mu, \kappa^2) \), we can further derive the marginal distribution of \( \mathbf{\delta} \) by integrating \( \beta \) out of (3), which gives

\[
\mathbf{\delta} \sim \mathcal{N}(\mu \mathbf{\lambda}, \Sigma + \kappa^2 \mathbf{\lambda} \mathbf{\lambda}').
\]

From (5), we can see that for any \( j, k > 0 \), \( \delta_j \) and \( \delta_k \) are dependent because they depend on the same random rate \( \beta \).

### 3 On-Line Filtering and RUL Prediction

The underlying degradation process \( X(t) \) is unobservable, but it determines the failure time. Accurate estimation of \( X(t) \) is crucial to characterize current health status and to predict the RUL for maintenance decisions. Therefore, it is important to estimate \( X(t) \) based on the observed \( \mathbf{Y} \). When \( \beta \) is random, however, direct estimation of \( X(t) \) requires all observed data up to time \( t \). We therefore turn to the joint distribution of the degradation level and the unknown \( \beta \), which enables an efficient online filtering algorithm.
In more details, we consider a unit whose underlying degradation and the observed signals follow models (1) and (2), respectively. The unit is inspected at times $0 = t_0 < t_1 < t_2 < \cdots < t_n < \cdots$, with corresponding observations $Y_1, Y_2, \cdots, Y_n, \cdots$. Based on all previous observations up to time $t_n$, the objective of condition monitoring is to compute the joint distribution of $\beta$ and the current degradation level $X_n \equiv X(t_n)$, denoted by $(\beta, X_n|Y_1, \ldots, Y_n)$.

For the purpose of real time processing of the degradation signals, we develop a recursive filtering algorithm to compute the joint distribution of $(\beta, X_n|Y_1, \ldots, Y_n)$, $\forall n = 1, 2, \cdots$. The proposed filtering algorithm is composed of two steps. In keeping with the Kalman filter, we shall call these two steps “predict” and “update”. At the first step, it predicts the state of $(\beta, X_n)$ using the state estimate from the previous iteration, i.e., to compute the conditional distribution of $(\beta, X_n|Y_1, \cdots, Y_{n-1})$. At the second step, it updates the joint distribution of $(\beta, X_n)$ upon the collection of $Y_n$, i.e., to obtain the conditional distribution $(\beta, X_n|Y_1, \cdots, Y_n)$.

It is easy to verify that $(\beta, X(t))'$ follows a bi-variate normal distribution because every linear combination of $\beta$ and $X(t)$ is normal. In the following, we denote $\omega^{j:k}$ and $Q^{j:k}$ as the mean and covariance of the conditional distribution of $(\beta, X_j|Y_1, \cdots, Y_k)$, $j \geq k$. It is readily found that $(\beta, X_0)' \sim \mathcal{N}(\omega^{0:0}, Q^{0:0})$, where $\omega^{0:0} = (\mu, 0)'$ and $Q^{0:0}$ is a diagonal matrix with diagonal vector $(\kappa^2, 0)'$. Based on these set-ups, the “predict” step can be computed according to Theorem 1. The proof is included in the Appendix.

**Theorem 1** Given $(\beta, X_n|Y_1, \ldots, Y_{n-1}, Y_n) \sim \mathcal{N}(\omega^{n:n}, Q^{n:n})$, the conditional joint distribution of $(\beta, X_{n+1}|Y_1, \ldots, Y_{n-1}, Y_n)$ follows a bi-variate normal distribution $\mathcal{N}(\omega^{n+1:n}, Q^{n+1:n})$, where $\omega^{n+1:n} = S_n \cdot \omega^{n:n}$ and $Q^{n+1:n} = S_n (Q^{n:n} + P_n) S_n'$. The matrices $S_n$ and $P_n$ are defined as

$$S_n = \begin{bmatrix} 1 & 0 \\ \lambda_{n+1} & 1 \end{bmatrix}, \quad P_n = \begin{bmatrix} 0 & 0 \\ 0, \sigma^2 \lambda_{n+1} \end{bmatrix}.$$ 

Theorem 1 can be easily extended to predict the joint distribution of the future degradation levels at multiple time points. When $Y_{n+1}$ is observed at the next inspection, the
The joint distribution of \((\beta, X_{n+1})'\) can be updated. The updated distribution provides a more accurate estimate of the rate \(\beta\) and the underlying degradation level. Theorem 2 summarizes the result of the “update” step. The proof can be found in the Appendix.

**Theorem 2** Given the joint distribution \((\beta, X_{n+1}|Y_1, \cdots, Y_n) \sim \mathcal{N}(\omega^{n+1:n}, Q^{n+1:n})\) and the observation \(Y_{n+1} = Y(t_{n+1})\), we have \((\beta, X_{n+1}|Y_1, \cdots, Y_n, Y_{n+1}) \sim \mathcal{N}(\omega^{n+1:n+1}, Q^{n+1:n+1})\), where \(\omega^{n+1:n+1} = A_n \omega^{n+1:n} + B_n\), \(Q^{n+1:n+1} = A_n Q^{n+1:n}\). The matrices \(A_n, B_n\) are defined as

\[
A_n = \begin{bmatrix}
1, & -Q_{1,2}^{n+1:n}/(\gamma^2 + Q_{2,2}^{n+1:n}) \\
0, & \gamma^2/(\gamma^2 + Q_{2,2}^{n+1:n})
\end{bmatrix}, \quad B_n = \frac{Y_{n+1}}{\gamma^2 + Q_{2,2}^{n+1:n}} \begin{bmatrix}
Q_{1,2}^{n+1:n} \\
Q_{2,2}^{n+1:n}
\end{bmatrix}.
\]

The subscript of a matrix \(M_{j,k}\) represents the \((j,k)\)-th element of the matrix \(M\).

Because of the recursive nature of the algorithm, it can be effectively used in real time applications. Only the latest observation and the updated distribution of \((\beta, X(t))'\) from the last iteration are needed. A storage of all historical observations is not required. In addition, the “predict” and “update” steps only require additions and multiplications of matrices. The algorithm is thus free of the expensive matrix inverse operation.

After obtaining the up-to-date estimate of the degradation level and the drift rate \(\beta\), we can use it to predict the RUL distribution. A typical degradation-threshold failure is defined as the event that \(X(t)\) first hits a pre-specified failure threshold \(D\). Based on the degradation-threshold failure, the RUL \(R(t_n)\) at time \(t_n\) is defined as the time needed for \(X(t)\) to hit the threshold \(D\) from current degradation level \(X_n\). It is readily seen that after observing \(Y_n\) at \(t_n\), the CDF of \(R(t_n)\) can be computed as

\[
F_{R(t_n)}(\tau|Y_1, Y_2, \cdots, Y_n) = \mathbb{E}_{(\beta,X_n)} \left[ F_{IG} \left( \Lambda(\tau + t_n) - \Lambda(t_n); \frac{D - X_n}{\beta}, \frac{(D - X_n)^2}{\sigma^2} \right) \right],
\]

where the expectation is taken using the distribution \(\mathcal{N}(\omega^{n:n}, Q^{n:n})\), and \(F_{IG}(x; a, b)\) is the CDF of an inverse Gaussian distribution with mean \(a\) and variance \(a^3/b\).

Although there is no closed-form expression for the distribution of \(R(t_n)\), it can be
effectively computed using some well-developed numerical methods, such as the Riemann-Stieljes method [26]. The updated distribution of the RUL can then be used for subsequent reliability decision-makings such as preventive maintenance scheduling.

4 Parameter Estimation

In Section 3, we assume that the model parameters $\theta$, i.e., $\mu, \kappa^2, \gamma^2, \sigma$ and parameters involved in $\Lambda(\cdot)$, are known. In practice, $\theta$ might be unknown but historical degradation data of similar units can be used to estimate $\theta$. In this section, we discuss how to estimate $\theta$ from historical data.

Assume that degradation data of $N$ units are available. Throughout the section, the subscript $i$ is used to index the units, and $j$ is used to index the time epochs. For unit $i$, it was inspected at ordered inspection times $t_i = \{t_{i1}, t_{i2}, \ldots, t_{im_i}\}$, with observed degradation levels $\{Y_i(t_{i1}), Y_i(t_{i2}), \ldots, Y_i(t_{im_i})\}$, where $Y_i(t_{ij}) = X(t_{ij}; \beta_i) + \epsilon_{ij}$, and $\beta_i$ is the degradation rate of unit $i$. The measurement errors $\epsilon_{ij}$ are assumed to be i.i.d. realizations of $\epsilon$. It is noted that for any $j, k > 0$, $Y_i(t_{ij})$ and $Y_i(t_{ik})$ are dependent because they depend on the same random rate $\beta_i$. We assume that the parametric form of $\Lambda(t)$ has been specified in advance, either through expert knowledge or through analysis of the degradation physics.

As discussed in Section 2, it is often easier to work with the degradation increments. So we define $\delta_i = [Y_i(t_{i1}), Y_i(t_{i2}) - Y_i(t_{i1}), \ldots, Y_i(t_{im_i}) - Y_i(t_{i,m_i-1})]$, and $\lambda_i = [\Lambda(t_{i1}), \Lambda(t_{i2}) - \Lambda(t_{i1}), \ldots, \Lambda(t_{im_i}) - \Lambda(t_{i,m_i-1})]$. $\Sigma_i$ is similarly defined as in (4). Based on (5), the log-likelihood function (up to a constant) given the degradation observations, according to (5), can be expressed as

$$l(\theta | \delta_i, \cdots, \delta_N) = -\frac{1}{2} \sum_{i=1}^{N} \left[ \ln |\Sigma_i + \kappa^2 \lambda_i \lambda_i'| + (\delta_i - \mu \lambda_i)'(\Sigma_i + \kappa^2 \lambda_i \lambda_i')^{-1}(\delta_i - \mu \lambda_i) \right]. \quad (6)$$

Direct maximization of (6) can be done numerically, but it might be trapped in some local optimum. Instead, if we consider the realization $\beta_i$ of each unit as a missing datum, we can
use the EM algorithm for the inference. Let $\beta = [\beta_1, \beta_2, \cdots, \beta_N]'$ be the missing data. The log-likelihood given the complete data $\delta_i$ and $\beta$ can be expressed as

$$l(\theta|\delta_i, \beta_i, i = 1, 2, \cdots, N) = -\frac{1}{2} \sum_{i=1}^{N} \left[ \ln |\Sigma_i| + (\delta_i - \beta_i'\lambda_i)'\Sigma_i^{-1}(\delta_i - \beta_i'\lambda_i) \right]$$

$$-N \ln \kappa - \frac{1}{2\kappa^2} \sum_{i=1}^{N} (\beta_i - \mu)^2. \tag{7}$$

The EM algorithm executes the E-step and M-step iteratively. We use the superscript $(s)$ to denote the iteration in the EM algorithm, with $\theta^{(0)}$ being the starting point.

**E-Step:** The $Q$-function is obtained by taking the conditional expectation of (7) given the observed data $\delta_i$ and the current estimates of the parameters $\theta^{(s)}$. Since both the prior distribution of $\beta_i$ and the conditional distribution of $(\delta_i|\beta_i)$ are normal, it is easy to verify that $(\beta_i|\delta_i)$ follows a normal distribution as

$$[\beta_i|\delta_i] \sim \mathcal{N}\left(\frac{\kappa^2\lambda_i'\Sigma_i^{-1}\delta_i + \mu}{\kappa^2\lambda_i'\Sigma_i^{-1}\lambda_i + 1}, \left(\lambda_i'\Sigma_i^{-1}\lambda_i + \frac{1}{\kappa^2}\right)^{-1}\right). \tag{8}$$

As a result, we can obtain

$$a_i^{(s)} \equiv \text{E}(\beta_i|\delta_i, \theta^{(s)}) = \frac{(\kappa^2)^{(s)}(\lambda_i^{(s)})'\Sigma_i^{(s)}(\delta_i^{(s)})^{-1}\delta_i + \mu^{(s)}}{(\kappa^2)^{(s)}(\lambda_i^{(s)})'\Sigma_i^{(s)}(\lambda_i^{(s)})^{-1} + 1},$$

$$b_i^{(s)} \equiv \text{E}(\beta_i^2|\delta_i, \theta^{(s)}) = (a_i^{(s)})^2 + \left(\lambda_i^{(s)}'\Sigma_i^{(s)}(\lambda_i^{(s)})^{-1} + 1/(\kappa^2)^{(s)}\right)^{-1}.$$ 

Therefore, the $Q$-function can be expressed as

$$\text{E}[l(\theta)|\delta_i, \theta^{(s)}] = -\frac{1}{2} \sum_{i=1}^{N} \left[ \ln |\Sigma_i| + \delta_i'\Sigma_i^{-1}\delta_i - 2a_i^{(s)}\delta_i'\Sigma_i^{-1}\lambda_i + b_i^{(s)}\lambda_i'\Sigma_i^{-1}\lambda_i \right]$$

$$-N \ln \kappa - \frac{1}{2\kappa^2} \sum_{i=1}^{N} (b_i^{(s)} - 2a_i^{(s)}\mu + \mu^2). \tag{9}$$

**M-Step:** We maximize the $Q$-function (9) with respect to $\theta$ to obtain $\theta^{(s+1)}$ for the next
iteration. As we can see, \( \kappa \) and \( \mu \) in (9) are only involved in the second term on the right hand side of (9) and this term can be analytically optimized as

\[
\mu^{(s+1)} = \frac{\sum_{i=1}^{N} a_i^{(s)}}{N}, \quad (\kappa^2)^{(s+1)} = \frac{\sum_{i=1}^{N} \left( b_i^{(s)} - 2a_i^{(s)} \mu^{(s+1)} + (\mu^{(s+1)})^2 \right)}{N}.
\]

The solution for the other parameters can be numerically obtained by optimizing the first term of (9) on the right hand side. In particular, when all the units have the same inspection times and \( \Lambda(t) \) takes a certain parametric form such as \( \Lambda(t) = t^p \), (9) can be easily maximized. The E-step and M-step are applied iteratively until the solution converges.

5 Simulation Studies

5.1 Performance of online filtering

First, we use simulation studies to demonstrate the effectiveness of our filtering algorithm. To begin with, we consider the degradation model (1) with \( \beta \sim N(2, 1) \), \( \gamma = 2, \sigma = 1.5 \) and \( \Lambda(t) = t^{1.1} \). Then a degradation path is generated according to the model with randomly sampled \( \beta \). We sample the degradation path at times \( t = 1, 2, 3, \ldots \). Upon the observation of each new sample \( Y_n \), we can execute the predict and update steps of the filtering algorithm to obtain a new estimate of \( \beta \) and the current degradation state \( X_n \).

To illustrate the performance of the filtering algorithm, we plot the estimated \( \beta \) and \( X_n \) and the corresponding confidence intervals against the true values in Figure 1. It shows that the estimated degradation levels are able to track the real values quite well. In addition, the estimated degradation rate \( \beta \) approaches the true value as the observation process goes on.

We further run 1000 replications of the simulation, including degradation data generation, filtering at each observation time, and calculation of the estimation error. The box plots of the estimation errors of \( X_n \) and \( \beta \) at different times are summarized in Figure 2. The simulation demonstrates that the estimation error of \( X_n \) converges to \( \gamma \) as the majority of
the estimation errors lie between $-2\gamma$ and $2\gamma$. On the other hand, the error of $\beta$ decreases progressively as more observations are available.

5.2 Remaining useful life estimation

Section 5.1 demonstrates that the online filtering algorithm is able to estimate the degradation rate and the degradation state quite well. In this subsection, we further demonstrate the performance of RUL prediction based on the estimation of $(\beta, X_n|Y_1, \cdots, Y_n)$. In addition to the previous simulation settings, we set the failure threshold $D = 40$. We predict the RUL at different times $t$ using all the observations up to $t$. Figure 3 shows an example of the density of the RUL predicted at times $t = 5$ and $t = 10$ respectively. It clearly reveals that as more data are available, the predicted RUL becomes more accurate with a narrower spread.

To evaluate the accuracy of the RUL prediction, we replicate the simulation 500 times.
Figure 2: Box plot of the estimation errors at different times using 1000 replications. The boundary of the box indicates the inter-quartile ranges.

Figure 3: Density of the RUL predicted at two different times. The dashed line indicates the location of the true RUL.
In each replication, the prediction interval (PI) of the RUL with confidence level $\alpha = 0.1$ is constructed. We compare the length of the PI and the correct coverage probability of the PI when the prediction is made at $t = 5$ and $t = 10$. Figure 4 summarizes the length of the PI against the true RUL at the time of prediction. It reveals that when the prediction is made at the early stage of the unit’s degradation (corresponding to larger true RUL), the length of PI is longer and has larger variability. Moreover, the actual coverage probability of the PI, as estimated from the 500 simulations, is 0.947 when $t = 5$ and 0.908 when $t = 10$, while the nominal coverage should be $1 - \alpha = 0.9$. These results show that with even a little more observations, the PI is shorter and has more accurate coverage probabilities.

### 5.3 Performance of parameter estimation

We also find that with reasonable size of historical samples, the parameters can be estimated quite accurately. To illustrate the effectiveness of the EM algorithm, we compare the estimation performance using two simulations with identical parameter settings yet different
Figure 5: Box plots of the estimated parameters using different sample sizes. \( N \) is the number of degradation path and \( m_i \) is the number of observations from each path.

The true parameters in both simulations are set as \( \mu = 2, \kappa^2 = 0.64, \gamma^2 = 1, \sigma^2 = 0.25, \rho = 1.1 \). The box plot of the parameter estimates from 1000 replications are shown in Figure 5. The result illustrates that the estimation is almost unbiased, and the estimation error decreases as the sample size increases. Even though the parameters \( \mu \) and \( \kappa^2 \) of the prior distribution of \( \beta \) have relatively large variances initially, the prediction errors decrease sharply after we get a few in-situ observations from the system in operation.

6 Condition Monitoring of Rotation Bearings

In this section, we use a real example to demonstrate the application and effectiveness of the proposed methods. The dataset comes from an accelerated degradation test of rotational bearings, where vibration signals of 25 bearings are collected [27]. When the vibration magnitude exceeds a certain threshold, the bearing is considered as failed. Many researchers have demonstrated their condition monitoring methods using this dataset as a benchmark,
It has been found in previous studies that a log-linear model provides a good fit to the degradation data, and the bearings exhibit heterogeneity in the degradation rate [27]. Therefore, the degradation dataset analyzed in this section is a logarithm transformation of the original data in [27]. After the transformation, however, we find from the degradation plot that the empirical mean degradation path by direct average over the 25 paths is not linear. In our example, we use the general function $\Lambda(t) = t^p$ to capture the curvature, and the power parameter $p$ can be estimated from data (hereafter, the degradation data refer to a log-transformation of the original data in [27]). In addition, we include the observation error term which has not been considered before for this dataset. Among the 25 units with complete degradation observations, we randomly select 22 of them as training units to estimate the parameters of the degradation model, i.e., $\mu, \kappa^2, \gamma^2, \sigma^2$ and $p$. Using the estimated parameters, we apply the online filtering method to the remaining 3 bearings to jointly estimate their degradation rates and future degradation levels when new observations are made available during condition monitoring, after which the RUL distribution can be updated.

The parameters estimated from the selected 22 training units are shown in Table 1. Clearly, $\hat{p} = 0.711$ suggests that a linear model might not be a good choice, as also indicated in Figure 6. Based on the estimated parameters, we monitor the other 3 bearings using the proposed online filtering algorithm. Figure 7 plots the degradation levels and the drift rates $\beta$ estimated at different times $t$ using all observations up to $t$. Although the degradation paths exhibit large variability, the filtering algorithm tracks the degradation trajectories and updates the degradation rates promptly. Based on the filtering results, the RUL distribution (with failure threshold $D = 1.5$) can be further predicted, as shown in Figure 8. It shows that the predicted RUL is consistent with the degradation paths. When a unit has a large degradation rate $\beta$, the PDF of the RUL will be more concentrated. In addition, when the degradation level approaches the failure threshold, the PDF curve moves towards 0.
Table 1: Estimated parameters of the degradation model.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$\mu$</th>
<th>$\kappa^2$</th>
<th>$\gamma^2$</th>
<th>$\sigma^2$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimates</td>
<td>0.0243</td>
<td>0.00335</td>
<td>0.00634</td>
<td>0.0158</td>
<td>0.711</td>
</tr>
</tbody>
</table>

Figure 6: Degradation paths of the three testing bearing units.

Figure 7: Estimation of the degradation states and the drifts of the three units using the filtering algorithm. (a) The points represent the observed vibration signals, and the lines represent the estimated degradation levels, (b) the estimated mean of $\beta$. 
7 Conclusion

This paper studied condition monitoring of units with heterogeneous degradation rates and imperfect inspections. We used a Wiener-process model with measurement errors to characterize the observed degradation signals over time. To account for the aleatory uncertainty (i.e., the inherent heterogeneities) and possible epistemic uncertainties, a Bayesian approach was adopted where a prior distribution was imposed on the degradation rates $\beta$. An online filtering algorithm was developed to jointly estimate the distribution of the degradation rate and the underlying degradation levels. The filtering algorithm is efficient with a minimal storage requirement. When the model parameters are unknown, they can be effectively estimated from historical data using the EM algorithm.

There are some possible extensions to further enrich the applications of the proposed methods. One possibility is to generalize the model to a general hidden degradation process, where the observations might be a nonlinear function of the degradation level. Additionally, since all model parameters are estimated from data or from expert knowledge, they are
subject to statistical uncertainties. How to quantify the effects of the uncertainties, i.e., the epistemic uncertainties, is worth investigation. This investigation may help with the design of more robust condition-monitoring methods.

Acknowledgements: We would like to thank the editor and two anonymous reviewers for their constructive comments that have greatly helped in the revision of an earlier version of the paper.

Appendix

Proof of Theorem 1

Because $X_{n+1} = X_n + \Delta_n$, where $\Delta_n \sim \mathcal{N}(\beta \lambda_n + 1, \sigma^2 \lambda_n + 1)$, we have $(X_{n+1} | \beta, X_n, Y_1, ..., Y_n) \sim \mathcal{N}(X_n + \beta \lambda_n + 1, \sigma^2 \lambda_n + 1)$. Recall that $(\beta, X_n | Y_1, Y_2, ..., Y_n) \sim \mathcal{N}(\omega^{n:n}, Q^{n:n})$. It is readily seen that

$$
\ln f(\beta, X_n, X_{n+1} | Y_1, Y_2, ..., Y_n) \propto -\frac{1}{2} \left[ (v - \mu)' T_1 (v - \mu) + v' T_2 v \right], \quad (A.1)
$$

where $v \equiv (\beta, X_n, X_{n+1})'$,

$$
\mu = \begin{bmatrix} \omega^{n:n} \\ \omega^{n:n} \\ 0 \end{bmatrix}, \quad T_1 = \begin{bmatrix} (Q^{n:n})^{-1} & 0 \\ 0 & 0 \end{bmatrix}, \quad T_2 = \frac{1}{\sigma^2 \lambda_n + 1} \begin{bmatrix} \lambda^2_{n+1} & \lambda_{n+1} & -\lambda_{n+1} \\ \lambda_{n+1} & 1 & -1 \\ -\lambda_{n+1} & -1 & 1 \end{bmatrix}. \quad (A.2)
$$

From (A.1), we see that $(\beta, X_n, X_{n+1} | Y_1, Y_2, ..., Y_n)$ follows a multivariate normal distribution with covariance matrix $\Sigma^{n+1:n} = (T_1 + T_2)^{-1}$, and mean $\mu^{n+1:n} = \Sigma^{n+1:n} \cdot T_1 \mu$. By manipulating the matrix, we can simplify the mean and covariance matrix as

$$
Q^{n+1:n} = \begin{bmatrix} Q^{n:n}_{1,1} & \lambda_{n+1} Q^{n:n}_{1,1} + Q^{n:n}_{1,2} \\ \lambda_{n+1} Q^{n:n}_{1,1} + Q^{n:n}_{1,2} & \lambda^2_{n+1} Q^{n:n}_{1,1} + 2\lambda_{n+1} Q^{n:n}_{1,2} + Q^{n:n}_{2,2} + \sigma^2 \lambda_{n+1} \end{bmatrix} = S_n (Q^{n:n} + P_n) S_n'.
$$
\[ \mu^{n+1:n} = \Sigma^{n+1:n} \cdot T \mu = \begin{bmatrix} Q^{n:n}, & Q^{n:n}D' \\ DQ^{n:n}, & DQ^{n:n}D' + \sigma^2 \lambda_{n+1} \end{bmatrix} \cdot \begin{bmatrix} (Q^{n:n})^{-1} \omega^{n:n} \\ 0 \end{bmatrix} = \begin{bmatrix} \omega^{n:n} \\ D\omega^{n:n} \end{bmatrix}, \]

where \( D = [\lambda_{n+1}, 1] \) is a row vector. Taking the first and third element of \( \mu^{n+1:n} \), we can obtain \( \omega^{n+1:n} = S_n \omega^{n:n} \).

**Proof of Theorem 2**

Using the Bayes theorem, we have

\[ f(\beta, X_{n+1}| Y_1, ..., Y_n, Y_{n+1}) \propto \exp \left\{ -\frac{1}{2} \left[ (u - \omega^{n+1:n})' (Q^{n+1:n})^{-1} (u - \omega^{n+1:n}) + (u - \nu)' T (u - \nu) \right] \right\}, \]

where \( u = (\beta, X_{n+1})' \), \( \nu = \begin{bmatrix} 0 \\ Y_{n+1} \end{bmatrix} \), \( T = \begin{bmatrix} 0 & 0 \\ 0 & \gamma^{-2} \end{bmatrix} \).

By combining the quadratic forms, we can find that \( f(\beta, X_{n+1}| Y_1, ..., Y_n, Y_{n+1}) \) follows a bi-variate normal distribution with covariance matrix \( Q^{n+1:n+1} = \left( (Q^{n+1:n})^{-1} + T \right)^{-1} \), and mean vector \( \omega^{n+1:n+1} = Q^{n+1:n+1} \cdot \left( (Q^{n+1:n})^{-1} \omega^{n+1:n} + T \cdot \nu \right) \). After some manipulation, we have

\[ Q^{n+1:n+1} = \left( (Q^{n+1:n})^{-1} + T \right)^{-1} = (I_{2 \times 2} + Q^{n+1:n} T)^{-1} \cdot Q^{n+1:n} = A_n \cdot Q^{n+1:n}, \]

and \( \omega^{n+1:n+1} = A_n \cdot \omega^{n+1:n} + A_n \cdot Q^{n+1:n} T \cdot \nu \). It can be easily shown that the last term is equal to \( B_n \) defined in Theorem 2.
References


