Online Robustness Degradation Analysis with Measurement Outlier

Xingchen Liu, Carman K.M. Lee*, Jingyuan Huang, Qiuzhuang Sun

Abstract—Online degradation analysis requires an adaptive model parameter estimation. In addition, measurement errors and outliers are inevitable in real applications of degradation analysis. However, existing online models ignore the measurement error or assume the measurement error to be distributed as Gaussian for mathematical simplicity, which is vulnerable to measurement outliers. To deal with such problems, an online degradation analysis technique with robustness to measurement outliers is developed. More specifically, the underlying degradation is modeled with the Wiener process and the measurement error is modeled by constructing a modified Huber density to enhance the robustness against the outlier. For the adaptive estimation of model parameters, an online Expectation-Maximization (EM) algorithm is developed. Furthermore, procedures are provided for recursive degradation state identification by maximizing a posteriori based on the Laplace approximation. Numerical and two real case studies are carried out to validate the efficacy of the proposed model.

Index Terms—Wiener process, modified Huber density, online expectation-maximization, maximizing a posterior, Laplace approximation.

I. Introduction

Degradation is a common phenomenon for modern physical products, e.g., mechanical components [1] and electronic systems [2]. Degradation will eventually lead to the failure of products, which causes substantial operation and maintenance costs [3]. The utilization of degradation information for the health status evaluation and the remaining useful life prediction of a product holds practical value. Degradation modeling, which effectively incorporates degradation information, has gathered significant attention from academia and industry.

A variety of methods are proposed for degradation modeling, including model-based approaches and data-driven approaches [4]. Model-based approaches formulate the degradation process based on physical or mechanism models. For data-driven approaches, the degradation pattern is directly extracted from the recorded data. There are some frequently utilized data-driven techniques, such as stochastic process models [5], Bayesian deep learning [6], and recurrent neural networks [7]. With some advantages such as being statistically interpretable and flexible, stochastic processes are widely adopted. Three widely adapted stochastic process models include the inverse Gaussian process model [8], the Gamma process model [9], and the Wiener process model [10]. The degradation processes that are not monotonic, such as the battery's capacity and the LED's luminosity, can be modeled with the Wiener process. Monotonic degradation processes, e.g., the wear and the crack propagation, can be formulated by the other two models.

Most traditional models estimate parameters based on all the degradation observations in a batch manner. For example, Whitmore et al [11] proposed a bivariate Wiener process model, where one component represents marker and the second latent component determines the failure time, and estimated the model parameters with a batch maximum likelihood estimation. Mishra et al. [12] utilized principal component regression to extract damage sensitive features of a Lamb wave sensor signal and predicted future damage with the Wiener process, whose parameters are obtained based on all the historical data. Muhammad et al. [13] considered the unit-to-unit variability of the drift rate of the Wiener process model and estimated model parameters with the Gibbs sampling technique and the Metropolis–Hastings algorithm. However, since the observed degradation data is usually obtained in a sequential manner, re-estimating model parameters from scratch based on all historical datasets is time and storage-consuming.

For an effective real-time analysis, some extensions on the Wiener process are developed for the online degradation modeling. For example, Si [14] proposed a nonlinear Wiener process model with a time-varying drift coefficient and adopted the Kalman filter to online update model parameters. Wang et al. [15] considered the dynamic environment and formulated the drift rate of a Wiener process model as an autoregressive one. Zhai et al. [16] developed an adaptive Wiener process model, whose drift rate is modeled by a continuous Brownian motion to get rid of the problems in the autoregressive model of order one and the value can be updated online. Omishi et al. [17] proposed a condition-based maintenance policy based on the remaining useful life prediction with a stochastic

This work was supported in part by Hong Kong Innovation and Technology Commission (InnoHK Project CAiRS); in part by the National Natural Science Foundation of China under Grant 72471144.

Xingchen Liu and Carman K.M. Lee are with the the Department of Industrial and Systems Engineering, Hong Kong Polytechnic University, Hong Kong SAR, China, and also with the Center for Advances in Reliability and Safety, Hong Kong Science Park, Hong Kong SAR, China (e-mail: xingchenliu@u.nus.edu, ckm.lee@polyu.edu.hk).

Jingyuan Huang is with the Center for Advances in Reliability and Safety, Hong Kong Science Park, Hong Kong SAR, China (e-mail: jenny.huang@cairs.hk).

Qiuzhuang Sun is with the School of Mathematics and Statistics, The University of Sydney, Camperdown, NSW 2006, Australia, and also with the National University of Singapore Suzhou Research Institute, Suzhou 215000, China (e-mail: qiuzhuang.sun@sydney.edu.au).

model-based degradation model, whose parameters can be updated based on the Bayes' theorem. By linking the features measured in the reliability tests to the variables measurable online, Djeziri et al. [18] proposed to model the trend of the health indices with the Wiener process model, where the drift parameter was also updated online. To jointly analyze the system degradation and the lifetime data, Hu et al. [19] adopted a random-effects Wiener process to model the degradation process and developed a particle filter method to update the model parameters online. Zhao et al. [20] combined the Box–Cox transformation with the Wiener process with linear drift to model the degradation of aluminum electrolytic capacitor and also updated the drift coefficient based on similarity measurement. However, all these works regard the observed degradation data as perfect without measurement error. Due to reasons such as the limited precision of sensors and measurement variations, measurement errors are inevitable.

To investigate the measurement errors, Ye et al. [21] developed a mixed-effects Wiener process model and provided efficient estimation procedures. Some other works combine the Wiener process model and some filtering techniques for the analysis of observed degradation data with measurement errors. For example, Lei et al. [22] considered the variability from multiple sources and achieved the degradation state estimation based on a Kalman particle filtering technique. To mitigate the impact of measurement errors, Peng et al. [23] proposed a batch particle filter for the multivariate degradation processes. Veloso et al. [24] proposed a dynamic linear degradation model with a local linear approximation for the true degradation path and assumed the measurement error is distributed as Gaussian. Liu et al. [25] integrated the Box-Cox transformation with the state-space modeling as a unified prognostic framework in the presence of Gaussian-distributed measurement error. These models mentioned above assume that the distribution of measurement error is Gaussian, which is mathematically tractable for the following estimation and inference procedures. However, the Gaussian assumption may cause misleading results with the presence of large disturbances or outliers in practical applications. To deal with such a problem, Zhai et al. [26] adopted the student t-distribution for the measurement error of a Wiener process and estimated the model parameters and other hidden values with the variational Bayes technique. Similarly, Ge et al. [27] extended this work by modeling the measurement error with scale-mixture normal distributions. Liu et al. [28] proposed a framework based on the distributionally robust optimization that is robust to measurement outlier as well as the uncertainty of model parameters. However, all these robust models are based on offline parameter estimation and degradation state identification. It is unknown how they can be extended to the online case.

To our certain knowledge, there is no existing literature or technique for online degradation analysis that is robust to the measurement outlier. To fill such a research gap, a technique is proposed for online degradation analysis in the presence of measurement outliers. More specifically, the challenge of online degradation analysis with outliers is solved by constructing a modified Huber density based on the modified Huber loss for measurement errors. This distribution is heavy-tailed and second-order differentiable with a simple form. For the adaptive estimation of model parameters, an online Expectation-Maximization (EM) algorithm is provided. Based on the Laplace approximation, recursive procedures are also proposed for the identification of degradation status by maximizing a posteriori.

The main contributions can be listed below. (1) An online robust degradation analysis technique is proposed by considering the measurement outlier in the online degradation analysis. (2) To enhance robustness against the outlier, a modified Huber density is constructed, which is heavy-tailed and second-order differentiable with a simple form. (3) An online EM algorithm is developed for the sequential adjustment of model parameters and achieve recursive degradation status identification based on the Laplace approximation.

The remaining part of this article can be structured below. The formulated problem and the basic Wiener process model are shown in Section II. Section III supplies procedures for the estimation of model parameters and the identification of the degradation status. The numerical and two real case studies are respectively provided in Section IV and V to illustrate the performance of the proposed method. Finally, Section VI gives the conclusion part.

II. Basic Wiener Process Model and Problem Formulation

Suppose a product degrades according to a Wiener process as:

$$X(t) = X_0 + \nu\Lambda(t) + \sigma\mathcal{B}(\Lambda(t)), \tag{1}$$

where ν represents the rate of drift, σ is the coefficient of diffusion, $\Lambda(t)$ represents a transformed time scale function, $\mathcal{B}(t)$ denotes the standard Brownian motion, and X_0 is the initial degradation state. The function $\Lambda(t)$ monotonically increases with respect to t and $\Lambda(0) = 0$. Widely used forms of $\Lambda(t)$ include the power law t^α and the exponential $\exp(\alpha t) - 1$ for $\alpha > 0$. The increments of $\mathcal{B}(t)$ are independent and Gaussian-distributed, i.e., $\mathcal{B}(t+\tau) - \mathcal{B}(t) \sim \mathcal{N}(0, \tau)$ for $t > 0$ and $\tau > 0$.

Consider observations $Y_1, Y_2, \cdots, Y_n$ that are measured in an online manner at discrete time points $t_1, t_2, \cdots, t_n$. The associated measurement errors are denoted as $e_1, e_2, \cdots, e_n$. The observations can be written as

$$Y_i = X_i + e_i, \tag{2}$$

for $i = 1, 2, \cdots, n$. The degradation process (1) can also be denoted in an incremental form as

$$X_i = X_{i-1} + \nu\lambda_i + \omega_i, \tag{3}$$

where $\lambda_i = \Lambda_i - \Lambda_{i-1}$, $\Lambda_i = \Lambda(t_i)$, and $\omega_i = \sigma\mathcal{B}(\Lambda_i) - \sigma\mathcal{B}(\Lambda_{i-1})$. The unknown model parameters include ν, σ,

and parameters regarding the distribution of measurement noise e_i. Based on observations $Y_{1:i}$, the objective is to estimate model parameters in an online manner as well as identify the degradation status X_i of the product recursively.

In the existing literature, the measurement noise is modeled with some heavy-tailed distributions, e.g., Student-t, Laplace, and Huber distributions [26, 29], to enhance model robustness to the outlier. The form of Student-t distribution is relatively complex, which complicates the estimation of model parameters and identification of degradation status. The form of Laplace and Huber distribution is simple. However, since there is an absolute function in the exponential part of the Laplace distribution, it is not always differentiable making its use in an online setting difficult. The Huber distribution is constructed based on the Huber loss, a combination of ℓ_1-norm and ℓ_2-norm losses, which is not second-order differentiable, causing inconvenience in the identification of degradation status.

To overcome such a deficiency, a modified Huber density is constructed based on the modified Huber loss [30], which is second-order differentiable. The definition of the probability density function (PDF) of the modified Huber density is given as

$$\mathcal{H}(x|\mu,\delta) := \frac{1}{\delta C_0}\exp\left\{-\rho_\epsilon\left(\frac{x-\mu}{\delta}\right)\right\}, \tag{4}$$

where μ and δ respectively denote the mean and standard error of the distribution. The value of C_0 is given as

$$C_0 = \int_{-\infty}^{+\infty}\exp\{-\rho_\epsilon(x)\}dx. \tag{5}$$

Here, $\rho_\epsilon(\cdot)$ denotes the modified Huber loss and is defined as

$$\rho_\epsilon(x) = \begin{cases}\epsilon^2[1-\cos(x/\epsilon)] & |x| \le \frac{\pi}{2}\epsilon;\\ \epsilon|x| + \frac{1}{2}\epsilon^2(2-\pi) & |x| > \frac{\pi}{2}\epsilon,\end{cases} \tag{6}$$

where ϵ is a tuning parameter. The advantages of the modified Huber density are twofold. Firstly, it has a simple form with a few parameters. In addition, it is second-order differentiable, which is convenient for developing techniques for online parameter estimation and degradation status identification.

The shapes of different loss functions, i.e., L_1-norm loss, L_2-norm loss, Huber loss, and modified Huber loss, as well as the corresponding first and second order derivatives (if available) are illustrated in Figure 1. The shape and derivative of the modified Huber loss are very close to that of the Huber loss. Similar to the L_1-norm loss, both of the Huber loss and the modified Huber loss are less affected by large deviations than L_2-norm loss. Unlike the L_1-norm loss and the Huber loss, which are not always second-order differentiable, the modified Huber loss has continuous second-order derivatives.

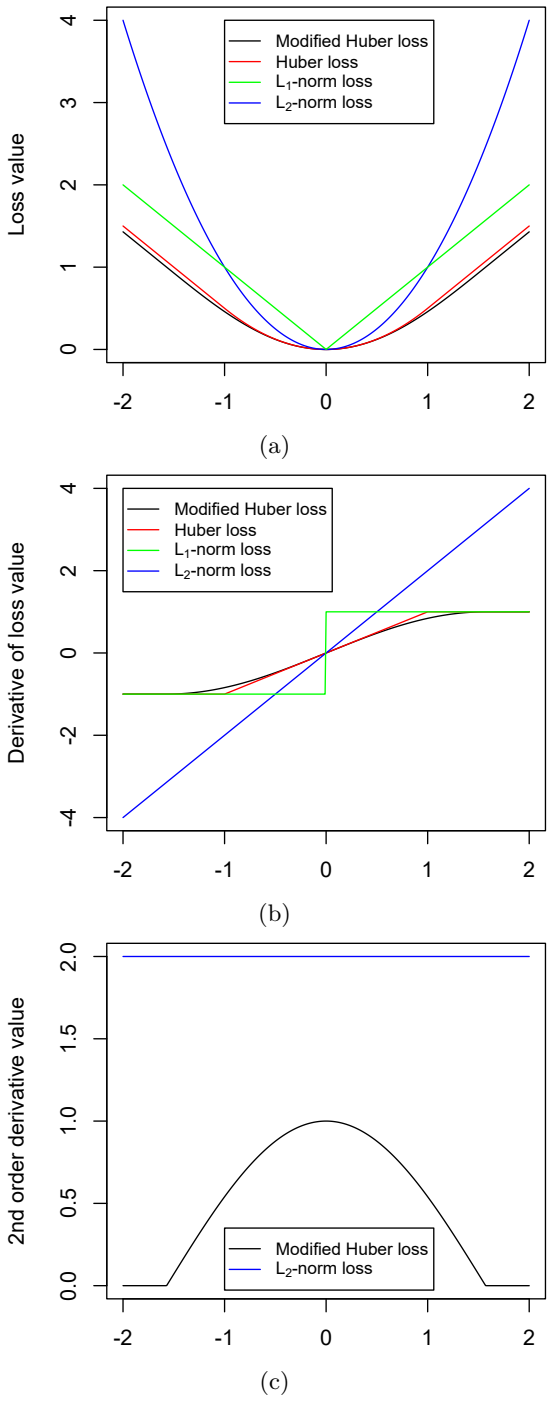


Fig. 1: The shapes of different loss functions, i.e., L_1-norm loss, L_2-norm loss, Huber loss, and modified Huber loss, as well as the corresponding first and second order derivatives (if available): (a) loss functions; (b) first derivatives; (c) second-order derivatives.

III. Estimation of Model Parameters and Identification of Degradation Status

A. Parameter Estimation with an Online Expectation Maximization Algorithm

$\Lambda(t)$ is firstly treated as given and the concentration is on the estimation of model parameters $\boldsymbol{\theta} = (\nu, \sigma^2, \delta)^\top$. The determination of $\Lambda(t)$ is discussed in Section III-C. As a regular procedure of an EM algorithm, the unobservable degradation status $X_{1:n} = (X_1, X_2, \cdots, X_n)^\top$ is regarded as missing data. For the complete data $\{X_{1:n}, Y_{1:n}\}$, it is recognized that

$$\begin{aligned} Y_i - X_i &\sim \mathcal{H}(0, \delta); \\ \Delta X_i = X_i - X_{i-1} &\sim \mathcal{N}(\nu\lambda_i, \sigma^2\lambda_i). \end{aligned} \tag{7}$$

The complete data log-likelihood function can be denoted as

$$\ell(\boldsymbol{\theta}; X_i, Y_i) = \ell_1(\nu, \sigma^2; X_i) + \ell_2(\delta; Y_i), \tag{8}$$

where

$$\begin{aligned} \ell_1(\nu, \sigma^2; X_i) &= -\frac{\ln 2\pi}{2} - \frac{\ln \sigma^2}{2} - \frac{1}{2}\ln\lambda_i - \frac{1}{2\sigma^2}\frac{(\Delta X_i - \nu\lambda_i)^2}{\lambda_i}, \\ \ell_2(\delta; X_i, Y_i) &= -\ln\delta - \ln C_0 - \rho_\epsilon\Big(\frac{Y_i - X_i}{\delta}\Big). \end{aligned} \tag{9}$$

Let $\mathbb{P}_{X_i|Y_{1:i}}$ denote the conditional probability distribution of X_i given $Y_{1:i}$. Q-function is defined as

$$Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} \mathbb{E}_{X_i \sim \mathbb{P}_{X_i|Y_{1:i};\hat{\boldsymbol{\theta}}}}[\ell(\boldsymbol{\theta}; X_i, Y_i)]. \tag{10}$$

Based on the conditional probability distribution $\mathbb{P}_{X_i|Y_{1:i}}$, the expectation can be taken with respect to X_i at the E-step. For simplification, $\mathbb{E}_{X_i;\hat{\boldsymbol{\theta}}}[\ell(\boldsymbol{\theta}; X_i, Y_i)]$ is utilized to denote $\mathbb{E}_{X_i \sim \mathbb{P}_{X_i|Y_{1:i};\hat{\boldsymbol{\theta}}}}[\ell(\boldsymbol{\theta}; X_i, Y_i)]$ in the remaining part of this paper. The estimation of parameters can be updated at the M-Step by maximizing the Q-function in (10), namely,

$$\tilde{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}). \tag{11}$$

Since the observation Y_i is obtained sequentially, the estimate of parameters is to be recursively updated. Let $\hat{\boldsymbol{\theta}}_{1:n}$ denote the estimate of $\boldsymbol{\theta}$ based on $Y_{1:n}$. Assume a new observation Y_{n+1} is obtained at the next time point t_{n+1}, then the M-step in (11) can be replaced by some gradient-descent-based updates. One possible choice is to directly move toward the direction of the negative gradient. As such, all parameters share the same learning rate. However, Duchi et al. [31] suggested that it is preferred if the learning rate can be adjusted individually for different parameters. Borrowing the idea from [31], the parameters are updated with an adaptive learning rate $\boldsymbol{\eta}_n$ in the following manner

$$\hat{\boldsymbol{\theta}}_{1:n+1} = \hat{\boldsymbol{\theta}}_{1:n} - \boldsymbol{\eta}_n \circ \nabla_{\boldsymbol{\theta}} \mathbb{E}_{X_{n+1};\hat{\boldsymbol{\theta}}_{1:n}}[\ell(\boldsymbol{\theta}; X_{n+1}, Y_{n+1})]\big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{1:n}}, \tag{12}$$

where '$\circ$' denotes the element-wise product of two vectors. For convenience of notation, $\boldsymbol{\theta}$ is denoted as $(\theta_1, \theta_2, \cdots, \theta_m)^\top$. The learning rate $\boldsymbol{\eta}_n = (\eta_{1,n}, \eta_{2,n}, \cdots, \eta_{m,n})^\top$ is defined as

$$\eta_{j,n} = \frac{\gamma}{\sqrt{\sum_{i=1}^{n}\Big(\frac{\partial \mathbb{E}_{X_{i+1};\hat{\boldsymbol{\theta}}_{1:i}}[\ell(\boldsymbol{\theta}; X_{i+1}, Y_{i+1})]}{\partial\theta_j}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{1:i}}\Big)^2 + \varepsilon}}, \tag{13}$$

where γ denotes the coefficient of decay. Here, a small number ε, e.g., $\varepsilon = 10^{-8}$, is added to avoid possible numerical issue. The above equation reduces the learning rate for parameters with big gradients while raising the learning rate for parameters with small gradients.

With the Leibniz rule, the derivative part in (12) can be exchanged with integration or expectation, i.e.,

$$\nabla_{\boldsymbol{\theta}} \mathbb{E}_{X_{n+1};\hat{\boldsymbol{\theta}}_{1:n}}[\ell(\boldsymbol{\theta}; X_{n+1}, Y_{n+1})] = \mathbb{E}_{X_{n+1};\hat{\boldsymbol{\theta}}_{1:n}}[\nabla_{\boldsymbol{\theta}}\ell(\boldsymbol{\theta}; X_{n+1}, Y_{n+1})]. \tag{14}$$

Similarly, the partial derivative part in (13) can also be written as

$$\frac{\partial \mathbb{E}_{X_{i+1};\hat{\boldsymbol{\theta}}_{1:i}}[\ell(\boldsymbol{\theta}; X_{i+1}, Y_{i+1})]}{\partial\theta_j} = \mathbb{E}_{X_{i+1};\hat{\boldsymbol{\theta}}_{1:i}}\Big[\frac{\partial\ell(\boldsymbol{\theta}; X_{i+1}, Y_{i+1})}{\partial\theta_j}\Big]. \tag{15}$$

More specifically, it can be derived that

$$\begin{aligned} \frac{\partial\ell(\boldsymbol{\theta}; X_i, Y_i)}{\partial\theta_1} &= \frac{\partial\ell_1(\nu, \sigma^2; X_i)}{\partial\nu} = -\frac{1}{\sigma^2}(\nu\lambda_i - \Delta X_i), \\ \frac{\partial\ell(\boldsymbol{\theta}; X_i, Y_i)}{\partial\theta_2} &= \frac{\partial\ell_1(\nu, \sigma^2; X_i)}{\partial\sigma^2} = -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4}\frac{(\Delta X_i - \nu\lambda_i)^2}{\lambda_i}, \\ \frac{\partial\ell(\boldsymbol{\theta}; X_i, Y_i)}{\partial\theta_3} &= \frac{\partial\ell_2(\delta; X_i, Y_i)}{\partial\delta} = -\frac{1}{\delta} + \rho'_\epsilon\Big(\frac{Y_i - X_i}{\delta}\Big)\frac{Y_i - X_i}{\delta^2}, \end{aligned} \tag{16}$$

where $\rho'_\epsilon(x)$ is the derivative of the modified Huber loss function, i.e.,

$$\rho'_\epsilon(x) = \begin{cases} \epsilon\sin(x/\epsilon) & |x| \le \frac{\pi}{2}\epsilon; \\ \epsilon & x > \frac{\pi}{2}\epsilon; \\ -\epsilon & x < -\frac{\pi}{2}\epsilon. \end{cases} \tag{17}$$

For the successful implementation of the above online update procedure, it is still necessary to get the conditional distribution of X_i given $Y_{1:i}$, i.e., $\mathbb{P}_{X_i|Y_{1:i}}$, for $i = 1, 2, \cdots, n$, which is corresponds to the degradation status identification. The technical details are provided in Section III-B, which approximate $\mathbb{P}_{X_i|Y_{1:i}}$ as a Gaussian distribution $\mathcal{N}(\mu_{i|i}, \sigma^2_{i|i})$. Then the expectation with respect to X_i or X_i^2 in (16) can be obtained analytically.

B. Degradation Status Identification with Laplace Approximation

In this subsection, the objective is to identify the unobservable underlying degradation status $X_{1:n} = (X_1, X_2, \cdots, X_n)^\top$ in a recursive manner. For the likelihood function based on some heavy-tailed distributions, e.g., Huber density, the corresponding posterior probability distribution is typically not analytically evaluable. Time-consuming sampling procedures are required for those numerical approximation techniques, such as Markov chain Monte Carlo and particle filter. For a more efficient alternative, the Laplace approximation [32] is adopted, which provides an analytical expression for a posterior probability distribution using a Gaussian distribution whose mean equals the mode of a posterior solution

and precision equals the observed Fisher information. Since such a method requires the log-likelihood function to be second-order differentiable, the traditional Huber density is not applicable, which is the main motivation to construct a modified Huber density in this paper. The details are outlined below.

Assume the distribution of X_n is approximated by a Gaussian conditional on the observations up to time t_n, namely,

$$\mathbb{P}_{X_n|Y_{1:n}} \approx \mathcal{N}(\mu_{n|n}, \sigma^2_{n|n}). \tag{18}$$

Based on (18), the objective is to approximate $\mathbb{P}_{X_{n+1}|Y_{1:n+1}}$ with another Gaussian distribution $\mathcal{N}(\mu_{n+1|n+1}, \sigma^2_{n+1|n+1})$.

To this end, a one-step prediction is made toward the state at time t_{n+1} with (3). The distribution of this prediction can be derived as

$$\mathbb{P}_{X_{n+1}|Y_{1:n}} = \mathcal{N}(\mu_{n|n} + \nu\lambda_{n+1}, \sigma^2_{n|n} + \sigma^2\lambda_{n+1}). \tag{19}$$

Based on our modeling framework, the likelihood function of observing Y_{n+1} is equal to the modified Huber density function in (4), i.e., $p(Y_{n+1}|X_{n+1}) = \mathcal{H}(Y_{n+1} - X_{n+1}|0, \delta)$.

With the Bayes formula, it is recognized that

$$\begin{aligned} &p(X_{n+1}|Y_{1:n}, Y_{n+1}) \propto p(X_{n+1}|Y_{1:n})p(Y_{n+1}|X_{n+1}) \\ &\propto \exp\left\{-\frac{(X_{n+1} - \mu_{n|n} - \nu\lambda_{n+1})^2}{2(\sigma^2_{n|n} + \sigma^2\lambda_{n+1})}\right\} \cdot \exp\left\{-\rho_\epsilon\big(\frac{Y_{n+1} - X_{n+1}}{\delta}\big)\right\} \\ &= \exp\left\{-\frac{(X_{n+1} - \mu_{n|n} - \nu\lambda_{n+1})^2}{2(\sigma^2_{n|n} + \sigma^2\lambda_{n+1})} - \rho_\epsilon\big(\frac{Y_{n+1} - X_{n+1}}{\delta}\big)\right\}. \end{aligned} \tag{20}$$

Denote

$$f(X) = -\frac{(X - \mu_{n|n} - \nu\lambda_{n+1})^2}{2(\sigma^2_{n|n} + \sigma^2\lambda_{n+1})} - \rho_\epsilon\big(\frac{Y_{n+1} - X}{\delta}\big), \tag{21}$$

where the two terms on the right side are both concave in X. It concludes that $f(X)$ is a concave function of X. From the point of maximizing a posterior, the estimation of X_{n+1}, i.e., $\mu_{n+1|n+1}$, can be obtained as

$$\mu_{n+1|n+1} = \arg\max_X f(X). \tag{22}$$

As $f(X)$ is differentiable and concave, the maximization problem can be easily solved with gradient-based algorithms such as L-BFGS and MMA. With the Laplace approximation, the variance of the state estimation can also be obtained as

$$\begin{aligned} &(\sigma^2_{n+1|n+1})^{-1} = -f''(X)|_{X=\mu_{n+1|n+1}} \\ &= (\sigma^2_{n|n} + \sigma^2\lambda_{n+1})^{-1} + \delta^{-2}\rho''_\epsilon\big(\frac{Y_{n+1} - \mu_{n+1|n+1}}{\delta}\big) \end{aligned} \tag{23}$$

where the function ρ''_ϵ is the second-order derivative of the modified Huber loss, i.e.,

$$\rho''_\epsilon(x) = \begin{cases} \cos(x/\epsilon) & |x| \le \frac{\pi}{2}\epsilon; \\ 0 & |x| > \frac{\pi}{2}\epsilon. \end{cases} \tag{24}$$

The main procedures of the proposed technique are provided in Algorithm 1.

Algorithm 1: Online estimation of model parameters and identification of degradation status

Input: Sequential measurement values of the degradation $\{Y_1, Y_2, \cdots, Y_n\}$.

Output: The estimate of model parameters $\hat{\boldsymbol{\theta}}_{1:i}$ and the identified distributions of degradation status $\mathbb{P}_{X_i|Y_{1:i}}$ for $i = 1, 2, \cdots, n$.

1: Initialization: model parameters $i = 1$ and $\hat{\boldsymbol{\theta}}_{1:i} = \boldsymbol{\theta}_0$.
2: while $i <= n$ do
3: Get the predictive distribution $\mathbb{P}_{X_{i+1}|Y_{1:i}}$ with Eq. (19).
4: Get the mean and variance of the distribution $\mathbb{P}_{X_{i+1}|Y_{1:i+1}}$, i.e., $\mu_{i+1|i+1}$ in Eq. (22) and $\sigma^2_{i+1|i+1}$ in Eq. (23).
5: Calculate the expectation of the partial derivative of the complete data log-likelihood function with Eq. (15).
6: Update the learning rate $\boldsymbol{\eta}_i$ with Eq. (13).
7: Update the model parameters $\hat{\boldsymbol{\theta}}_{1:i+1}$ with Eq. (12).
8: Set $i = i + 1$.
9: end while

C. Determination of the Time Scale Function $\Lambda(t)$ and Initialization of Model Parameters

In the previous section, the time scale function $\Lambda(t)$ is fixed. To determine this function in real applications, the form of function $\Lambda(t)$ is firstly discussed. Two methods can be utilized. The first method determines the parametric form of $\Lambda(t)$ with prior knowledge or the practical degradation physics [26]. The other is to adopt some empirical forms, such as the exponential $\Lambda(t) = \exp(\alpha t) - 1$ and the power-law $\Lambda(t) = t^\alpha$. After determining the form of $\Lambda(t)$, the parameters in $\Lambda(t)$ can be estimated by maximizing the complete data log-likelihood function in (8) with respect to these parameters. For the online update of the parameters in $\Lambda(t)$, the partial derivative of these parameters can also be derived as that in (16). For example, the partial derivative for the parameter α in the power-law time scale function can be obtained as

$$\begin{aligned} &\frac{\partial \ell(\boldsymbol{\theta}; X_i, Y_i)}{\partial \alpha} = \frac{\partial \ell(\boldsymbol{\theta}; X_i, Y_i)}{\partial \lambda_i}\frac{\partial \lambda_i}{\partial \alpha} \\ &= \Big(-\frac{1}{2\lambda_i} - \frac{\nu^2}{2\sigma^2} + \frac{(\Delta X_i)^2}{2\sigma^2\lambda_i^2}\Big) \cdot \big(i^\alpha \ln i - (i-1)^\alpha \ln(i-1)\big). \end{aligned} \tag{25}$$

Similarly, the partial derivative for the parameter α in the exponential form can also be obtained as

$$\begin{aligned} &\frac{\partial \ell(\boldsymbol{\theta}; X_i, Y_i)}{\partial \alpha} = \frac{\partial \ell(\boldsymbol{\theta}; X_i, Y_i)}{\partial \lambda_i}\frac{\partial \lambda_i}{\partial \alpha} \\ &= \Big(-\frac{1}{2\lambda_i} - \frac{\nu^2}{2\sigma^2} + \frac{(\Delta X_i)^2}{2\sigma^2\lambda_i^2}\Big) \cdot \Big(ie^{\alpha i} - (i-1)e^{\alpha(i-1)}\Big). \end{aligned} \tag{26}$$

For the initialization of model parameters, the proposed online algorithm can be modified into a batch version. For the batch version, the backward-smoothing procedure [14] can also be utilized at the E-step. Based on the above

procedures, the E-step and M-step can be iterated until convergence.

IV. Numerical Study

To verify the efficacy of the developed technique, Monte Carlo simulations are carry out in this section. The duration of observations, i.e., the length of the degradation path, is set as $N = 200$. Additionally, the initial status is set as $X_0 = 200$, with a drift rate $\nu = -1$ and a diffusion coefficient $\sigma^2 = 1$. To simulate the degradation observations with possible outliers, the measurement error is set as a mixture of a Gaussian distribution and a large deviation, i.e.,

$$e_t = (1 - \kappa_t)e_{t,1} + \kappa_t e_{t,2}, \tag{27}$$

where $e_{t,1} \sim \mathcal{N}(0, \delta^2)$, $e_{t,2} = 15$. Here, κ_t is a binary variable taking value 0 or 1. A small proportion, such as 0.02, of κ_t values are randomly assigned as 1, while the remaining values are set as 0. For the measurement error $e_{t,1}$, its standard deviation is set as $\delta = 1$. The form of the power law, i.e., t^α and $\alpha = 0.8$, is set for the transformed time scale $\Lambda(t)$. Therefore, $\boldsymbol{\theta} = [\nu, \sigma^2, \delta, \alpha]^\top = [-1, 1, 1, 0.8]^\top$ represent the true values. For illustrative purposes, a simulated degradation path is shown in Figure 2 as an example. Here, the actual degradation status as well as the corresponding observation are denoted with black and green lines respectively.

The estimation of model parameters is initialized relying on data $Y_{1:n}$ for $n = 50$, which is denoted as $\hat{\boldsymbol{\theta}}_{1:n}$. Denote $\hat{X}_{n|n}$ the estimate of X_n given $Y_{1:n}$. After obtaining a new observation Y_{n+1} at the time point t_{n+1}, the estimation of the degradation status can be updated as $\hat{X}_{n+1|n+1}$ based on the proposed technique. Then the model parameters can also be updated as $\hat{\boldsymbol{\theta}}_{1:n+1}$. The above procedures are repeated until the end of the whole degradation path. As shown in Figure 2, outliers cause large deviations to the degradation status identification by the Gaussian assumption model, while the proposed method is less affected by outliers. With the same degradation path in Figure 2, the adaptively estimated model parameters based on the proposed technique is presented in Figure 3. With the increasing of observations, the estimation of model parameters converges to the actual values. In addition, the effect of outliers on the estimation is insignificant. The parameter estimation result of a benchmark model is also provided, where the assumption is made that the measurement noise follows a Gaussian distribution. After each new observation is obtained, this benchmark model re-estimate parameters from scratch. For the Gaussian assumption model, there are sharp changes of the estimation results due to the existence of the measurement outliers. As a comparison, the developed technique is more robust against these outliers, which is especially obvious for the parameter δ^2. Since δ^2 denotes the variance of measurement error, the outliers cause this value to be overestimated.

The absolute errors between the identified degradation status and the actual values are also provided in Figure 4.

TABLE I: Mean absolute error of the estimation of degradation status on the overall degradation path (with standard deviation in the parenthesis) as well as the running time (in Seconds) of different techniques.

	Benchmark 1	Benchmark 2	Benchmark 3	Proposed
MAE	0.696(0.741)	0.666(0.565)	0.616(0.537)	0.531(0.390)
Time	495.68	172.28	110.06	5.77

For quantitative evaluation of the effectiveness of the proposed technique, the mean absolute error (MAE) of the overall degradation path is adopted, whose definition is given as

$$\text{MAE} = \frac{1}{n - 49} \sum_{i=50}^{n} \left| X_i - \hat{X}_{i|i} \right|^2. \tag{28}$$

To exhibit the superiority of the proposed model, we conduct a comparative analysis against several benchmark models. Benchmark 1 denotes the basic Wiener process model with Gaussian measurement error, where the parameters are estimated with maximum likelihood estimation in a batch manner. Benchmark 2 is the model proposed by Zhai et al. [26], where the measurement error is modeled with the Student-t distribution and model parameters are estimated with variational inference under a framework of expectation maximization algorithm in a batch manner. Benchmark 3 represents the bootstrap particle filter [33] for Wiener process with Laplace measurement errors, where the degradation status is identified with the particle filter and the model parameters are estimated by maximizing the likelihood function in a batch manner. The degradation status identification results as well as the corresponding absolute error of these benchmark models are given in Figure 2 and 4, respectively. The MAE and running time taken by these models are presented in Table I. They indicate that the developed method gives a faster and more accurate degradation status identification than these benchmarks. As a comparison, the benchmark model is too time-consuming for an online application. The above procedures are repeated 1000 times and the MAEs of these results are summarized with a boxplot in Figure 5, which further indicates that the developed method gives a more precise degradation status identification than these benchmark models.

V. Engineering Illustrations

A. Lithium-Ion Battery Degradation Data

A dataset about the degradation of lithium-ion batteries is accessible from the website of NASA [34]. In the experiments, charge, discharge, and impedance profiles were executed at a temperature of 24 °C. Discharges were conducted at various current load levels until the battery voltage reached predetermined voltage thresholds. For each charge and discharge cycle, the capacity of the battery was measured. The capacity of lithium-ion batteries would decrease with usage due to various complicated

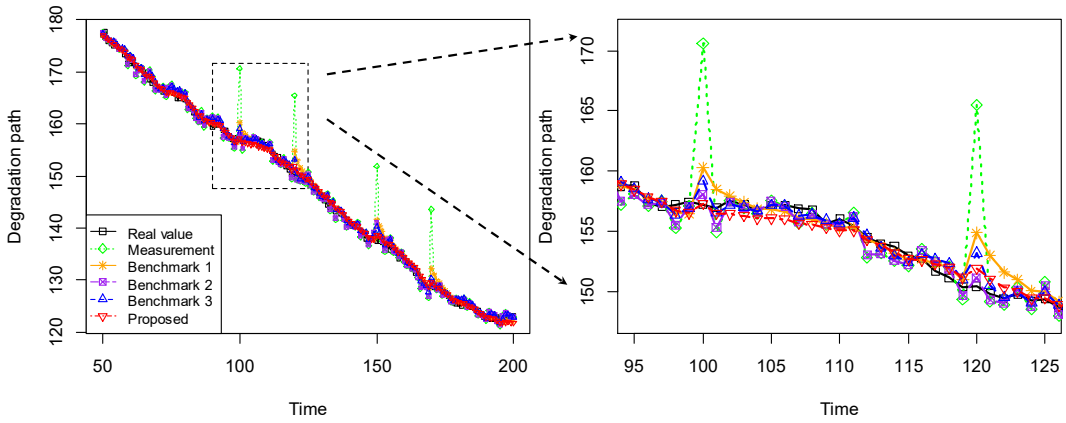


Fig. 2: A simulation of the degradation path, the observed degradation values, as well as the identified degradation status based on different techniques.

mechanisms. When the capacity of a battery diminishes to 70% of the nominal value, i.e., from 2 to 1.4 Ah, it is considered failed. Figure 6 illustrates measurement values of the degradation path of capacity for four distinct batteries, which are labeled as #5, #6, #7, and #18, from the provided dataset.

The developed technique is employed for analyzing all these four degradation paths. The form of the power law is set for the transformed time scale function both in this case study and in the case study in Section V-B. The same procedures as that in the numerical study are carried out to deal with each degradation path. The model parameters are initialized based on data $Y_{1:n}$ for $n = 10$ and let $\hat{\boldsymbol{\theta}}_{1:n}$ denote the estimated values of the model parameter. Relying on the developed technique, the degradation status can be identified as $\hat{X}_{n+1|n+1}$ based on the observation Y_{n+1} at the time point t_{n+1}. The model parameters can also be updated and represented as $\hat{\boldsymbol{\theta}}_{1:n+1}$. The above procedures are repeated for the whole degradation path.

Due to the presence of measurement errors in the observation value of the battery capacity, the actual value of the capacity is unavailable. Different methods are not directly comparable in the identification of degradation status like that in the numerical study, where the actual values of degradation status are given. To address this challenge, several criterion, i.e., Akaike information criterion (AIC), Bayesian information criterion (BIC), and minimum description length (MDL), are utilized for comparing different models. The definitions of these criterion are

$$\mathrm{AIC} = 2s - 2l, \tag{29}$$

$$\mathrm{BIC} = s\ln(n) - 2l, \tag{30}$$

$$\mathrm{MDL} = \frac{s}{2}\log_2(n) - \log_2 L, \tag{31}$$

where s denotes the number of model parameters, L is the likelihood, l represents the log-likelihood, and n is the number of samples. As more complex model tends to fit the data better, but it will also more likely to cause over-fitting. AIC, BIC, and MDL all handle the trade-off between the goodness of fit of the model and its simplicity. In other words, they deal with both the risk of over-fitting and the risk of under-fitting [35].

Since the distribution of degradation status is obtained in Section III-B, the value of the log-likelihood can be derived as

$$l = \sum_{i=1}^{n} \mathbb{E}_{X_i \sim \mathbb{P}_{X_i|Y_{1:i}}} \left[\ln p(Y_i|X_i) \right]. \tag{32}$$

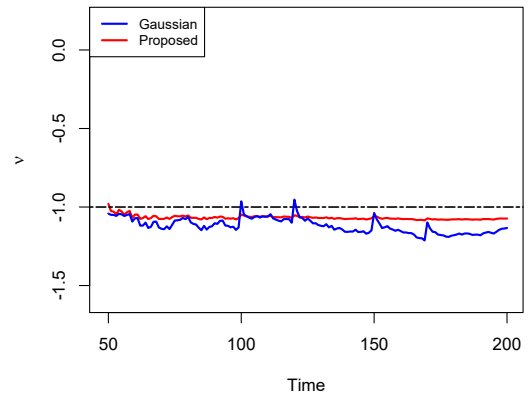
Here, the expectation can be calculated with some numerical integration techniques.

The degradation path of Battery 5 is analyzed and the result is presented in the following part as an illustration. Figure 7 displays the degradation status identification results obtained using the developed technique and the benchmark models. The evaluation criterion for the developed technique and the benchmark models can be calculated and are shown in Table II. These results indicate that the developed technique shows better performance than the benchmark model. Furthermore, the execution time of different techniques is also given in Table II as a comparison. It suggests that the developed technique is far more effective than the comparative.

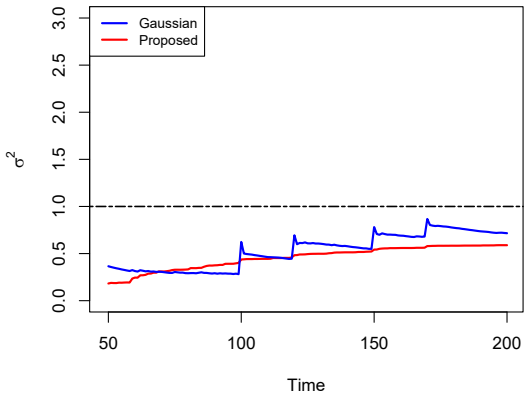
Next, the same procedures are applied to the measured degradation trajectories of the remaining three batteries. Table II presents the evaluation criterion and corresponding execution times of the degradation status identification using the developed technique and the benchmark techniques. The results demonstrate that the developed technique gets the smallest criterion and running time for all batteries, surpassing the performance of the comparison techniques. These outcomes indicate that the developed technique is more efficient and offers more accurate results in the identification of degradation status compared to the alternative techniques.

B. Rolling Bearing Degradation Data

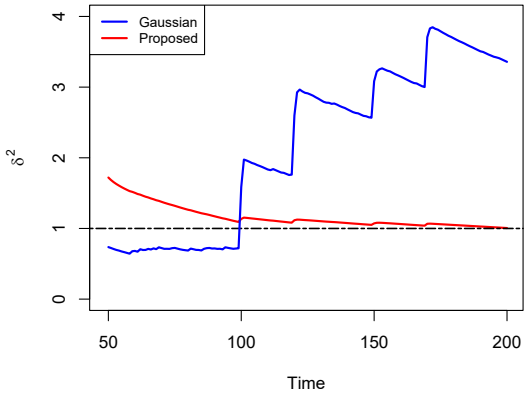
Another dataset about the rolling bearing degradation provided by [36] is analyzed in this section. For the



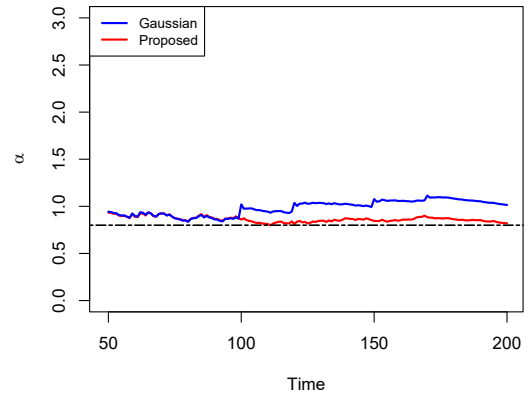
(a)



(b)



(c)



(d)

Fig. 3: Estimated model parameters: (a) ν; (b) σ^2; (c) δ^2; (d) α. The red lines are the results obtained by the proposed online technique. The blue lines denote the estimation result with Gaussian assumption in a batch manner.

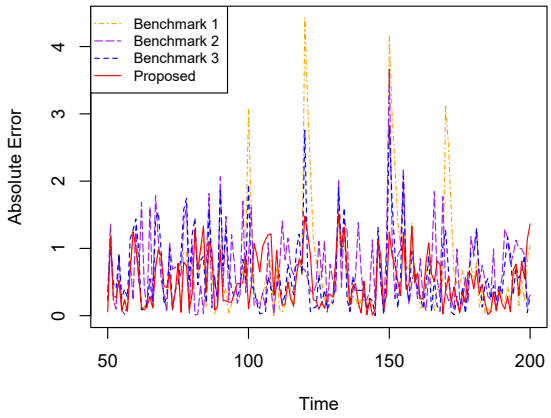


Fig. 4: The absolute error between the identified degradation status and the real values using the proposed/benchmark methods for the degradation trajectory shown in Figure 2.

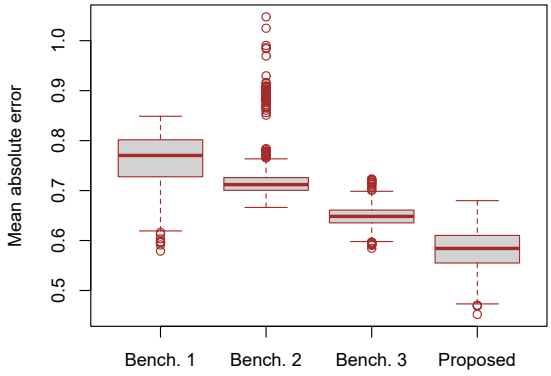


Fig. 5: Boxplot of MAEs of different techniques in degradation status identification based on the repetition for 1000 times.

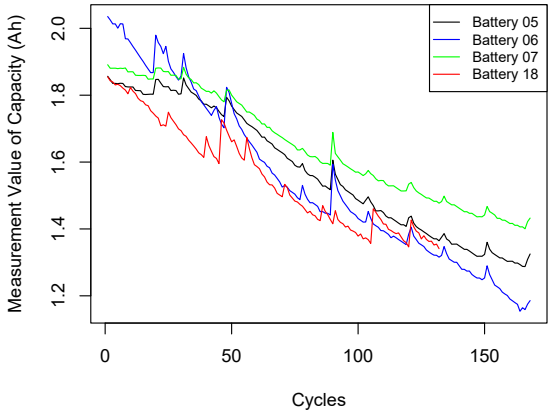


Fig. 6: Measurement values of the capacity degradation path of four different batteries from NASA.

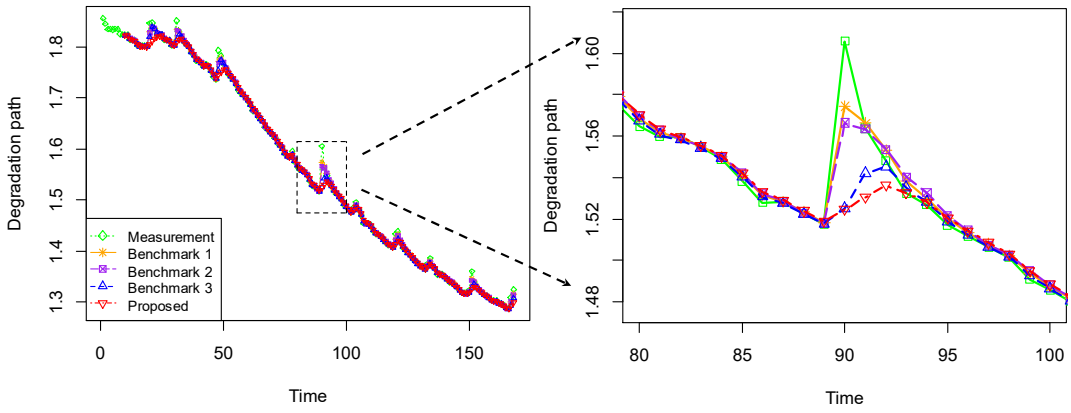


Fig. 7: Observed degradation path of #5 and the identified degradation status with different methods.

TABLE II: AIC, BIC, MDL values, and the corresponding running times (Seconds) of the battery degradation data.

		Bench. 1	Bench. 2	Bench. 3	Proposed
#05	AIC	-852.3	-478.9	-741.6	-954.4
	BIC	-846.1	-469.6	-735.3	-945.1
	MDL	-610.3	-338.7	-530.4	-681.7
	Time	298.33	180.37	141.49	6.04
#06	AIC	-793.3	-663.3	-820.3	-883.7
	BIC	-784.1	-653.8	-814.1	-874.3
	MDL	-567.7	-471.7	-587.2	-630.7
	Time	359.83	160.28	142.21	5.94
#07	AIC	-863.3	-567.5	-702.9	-987.9
	BIC	-857.1	-558.1	-696.6	-973.6
	MDL	-618.2	-402.6	-502.5	-702.3
	Time	387.94	160.7	142.02	6.02
#18	AIC	-667.3	-521.9	-626.2	-734.6
	BIC	-661.6	-513.3	-620.5	-725.9
	MDL	-477.2	-370.3	-447.6	-523.6
	Time	261.47	140.06	80.26	5.46

dataset, there are four run-to-failure experiments based on bearings of the same type, i.e., the Nachi 6205-2NSE9 single-row deep-groove ball bearing, which was lubricated with grease. All four tests were started from a bearing with a circular seeded defect on its raceway. There are different loading conditions and spall sizes for different tests. A commonly used indicator of bearing degradation is the acceleration's root mean square (RMS). The measurement values of RMS with respect to million cycles of these four testings are shown in Figure 8.

Similar to the previous section, all these four degradation paths are analyzed. As an illustration, the degradation path and the identified degradation status with different techniques of Test 1 are provided in Figure 9. It shows that the developed technique gives a smoother degradation path and is robust to these outliers. Table III shows the evaluation criterion and corresponding execution times using the developed technique and the benchmark techniques. It also shows the superiority of the proposed technique than the alternative one.

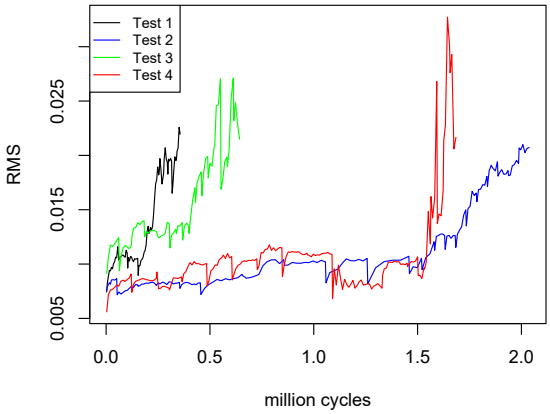


Fig. 8: Measurement values of degradation path of four run-to-failure tests.

C. Insulated Gate Bipolar Transistor (IGBT) degradation data

To further show the wide applicability of the developed technique, the IGBT aging data obtained from a thermal aging test [37] is further analyzed here. A square wave signal is applied to the gate to conduct a thermal cycle experiment, and transient data is collected by switching the device on and off continuously. The collector-emitter saturation voltage V_{CE}, which is defined as voltage drop from collector to emitter for a specified gate voltage and collector current, is very sensitive to the bond-wire lifting off and solder layer fatigue of IGBT. Hence it can well describe the aging process of IGBT. Figure 10 shows measurement values of the degradation path of V_{CE} for four distinct devices from the provided dataset.

Similarly, all these four degradation paths are analyzed. The degradation path and the identified degradation status with different techniques of Device 2 are provided in Figure 11 as an illustration. The developed technique presents a smoother degradation path and is robust against measurement outliers. The evaluation criterion

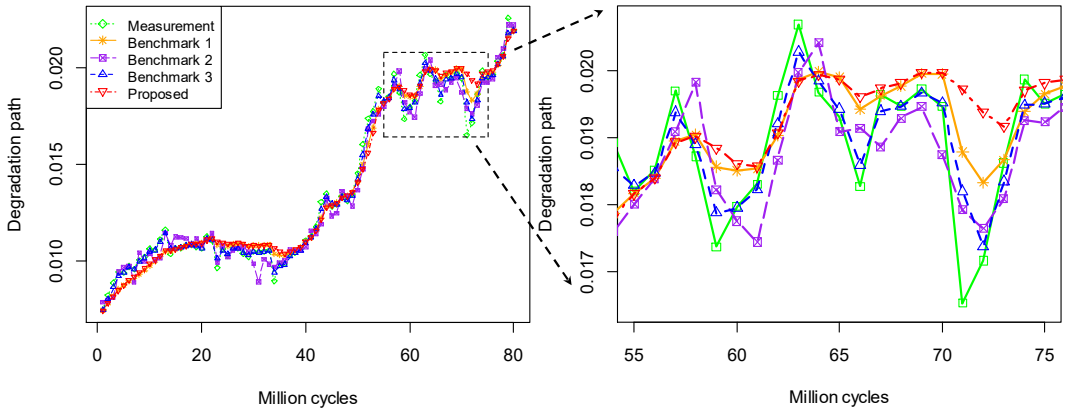


Fig. 9: Observed degradation path of Test #1 and the identified degradation status with different methods.

TABLE III: AIC, BIC, MDL values, and the corresponding running times (Seconds) of the bearing degradation data.

		Bench. 1	Bench. 2	Bench. 3	Proposed
#01	AIC	-517.3	-266.1	-467.4	-529.2
	BIC	-510.2	-258.9	-462.7	-524.4
	MDL	-368.1	-186.8	-333.7	-378.3
	Time	333.35	80.25	70.52	2.78
#02	AIC	-1112.7	-562.5	-981.4	-1120.6
	BIC	-1103.4	-553.1	-975.2	-1114.4
	MDL	-795.9	-398.9	-703.5	-803.9
	Time	192.49	140.28	131.99	6.21
#03	AIC	-531.5	-263.4	-460.7	-540.1
	BIC	-524.4	-256.3	-455.9	-535.4
	MDL	-378.3	-184.9	-328.9	-386.2
	Time	146.2	60.66	60.35	2.51
#04	AIC	-1214.2	-630.4	-1090.1	-1221.9
	BIC	-1204.6	-620.7	-1083.6	-1215.5
	MDL	-868.9	-447.7	-781.6	-876.8
	Time	195.01	90.21	102.78	6.39

TABLE IV: AIC, BIC, MDL values, and the corresponding running times (Seconds) of the IGBT degradation data.

		Bench. 1	Bench. 2	Bench. 3	Proposed
#02	AIC	-567.8	-463.6	-23.6	-636.5
	BIC	-575.9	-455.5	-29.1	-641.9
	MDL	-415.5	-328.5	-20.9	-463.1
	Time	226.73	100.99	182.73	3.97
#03	AIC	-843.2	-452.7	-140.5	-919.1
	BIC	-839.1	-444.8	-145.7	-914.5
	MDL	-608.9	-320.9	-105.1	-663
	Time	217.86	80.88	125.48	3.28
#04	AIC	-718.4	-477.4	-106.5	-787.8
	BIC	-712.1	-469.3	-111.8	-779.8
	MDL	-518.7	-338.6	-80.7	-562.4
	Time	230.43	100.57	114.92	3.84
#05	AIC	-571.2	-465.6	-79.5	-640.2
	BIC	-563.1	-457.6	-84.8	-636.1
	MDL	-412.6	-330.1	-61.2	-462.1
	Time	226.14	100.43	184.46	3.61

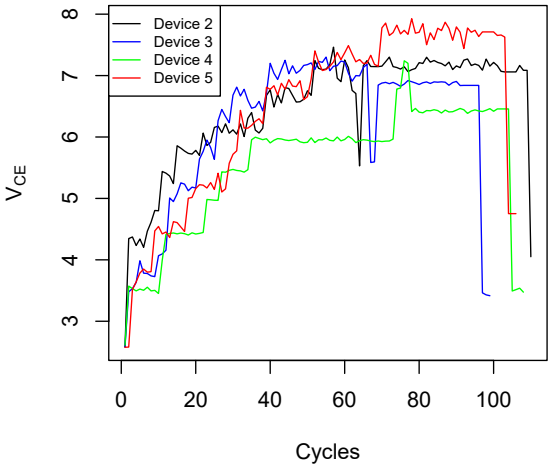


Fig. 10: Measurement values of the degradation path of four IGBT devices.

and corresponding execution times using the developed technique and the benchmark techniques are provided in Table IV, which also shows the superiority of the proposed technique than these alternatives.

In these real case studies, the degradation paths identified by the proposed technique are smoother than these benchmark models and less affected by measurement outliers. The evaluation criterion also shows that the proposed technique achieves the best trade off between the goodness-of-fit and model complexity. Since the proposed technique could update the model parameters online, it is much more efficient than these alternative models.

VI. Conclusion

In this study, an online and robust degradation analysis technique in the presence of measurement outlier was proposed. A modified Huber density, which was heavy-tailed and robust to outliers, was constructed. It was also second-order differentiable with a simple form and

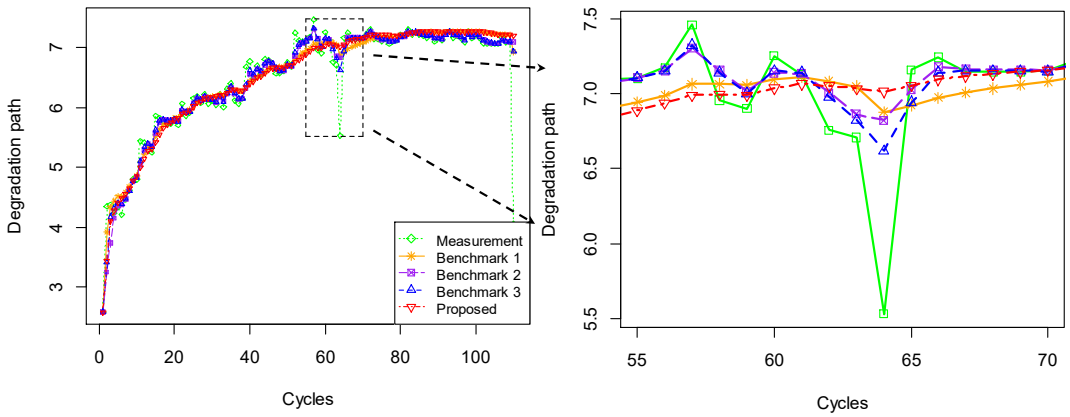


Fig. 11: Observed degradation path of Device #2 and the identified degradation status with different methods.

was convenient for inference. An online EM algorithm was developed for the adjustment of model parameters in a sequential manner. Procedures based on Laplace approximation for recursive degradation status identification by maximization of a posterior were also proposed. Comprehensive numerical and two real case studies have successfully demonstrated the efficacy and advantage of the developed technique. As for future work, it is of practical value to learn the patterns of outliers, such as periodicity, that exist in real-world applications and understand their effects on degradation analysis.

Acknowledgment

This work is supported by Centre for Advances in Reliability and Safety (CAiRS) admitted under AIR@InnoHK Research Cluster.

Special thanks to Professor Michael Pecht and Dr. Chak-Nam Wong for making helpful comments and suggestions on an earlier version of this paper.

References

[1] J. Qi, R. Zhu, C. Liu, A. Mauricio, and K. Gryllias, "Anomaly detection and multi-step estimation based remaining useful life prediction for rolling element bearings," Mech. Syst. Signal Process., vol. 206, p. 110910, 2024.

[2] M. Sudarshan, A. Serov, C. Jones, S. M. Ayalasomayajula, R. E. García, and V. Tomar, "Data-driven autoencoder neural network for onboard bms lithium-ion battery degradation prediction," J. Energy Storage, vol. 82, p. 110575, 2024.

[3] T. Yan, Y. Fu, M. Lu, Z. Li, C. Shen, and D. Wang, "Integration of a novel knowledge-guided loss function with an architecturally explainable network for machine degradation modeling," IEEE Trans. Instrum. Meas., vol. 71, pp. 1–12, 2022.

[4] A. F. Shahraki, O. P. Yadav, and H. Liao, "A review on degradation modelling and its engineering applications," Int. J. Perform. Eng., vol. 13, no. 3, p. 299, 2017.

[5] Z. Zhang, X. Si, C. Hu, and Y. Lei, "Degradation data analysis and remaining useful life estimation: A review on wiener-process-based methods," Eur. J. Oper. Res., vol. 271, no. 3, pp. 775–796, 2018.

[6] W. Peng, Z.-S. Ye, and N. Chen, "Bayesian deep-learning-based health prognostics toward prognostics uncertainty," IEEE Trans. Ind. Electron., vol. 67, no. 3, pp. 2283–2293, 2019.

[7] M. Catelani, L. Ciani, R. Fantacci, G. Patrizi, and B. Picano, "Remaining useful life estimation for prognostics of lithium-ion batteries based on recurrent neural network," IEEE Trans. Instrum. Meas., vol. 70, pp. 1–11, 2021.

[8] W. Peng, S.-P. Zhu, and L. Shen, "The transformed inverse gaussian process as an age-and state-dependent degradation model," Appl. Math. Modell., vol. 75, pp. 837–852, 2019.

[9] F. Corset, M. Fouladirad, and C. Paroissin, "Imperfect and worse than old maintenances for a gamma degradation process," Appl. Stochastic Models Bus. Ind., 2024.

[10] Z. Wang, Z. Chen, T. Xia, and E. Pan, "Degradation modeling considering the dependency of rate and volatility for real-time prognostics with error correction," IEEE Trans. Instrum. Meas., 2024.

[11] G. Whitmore, M. Crowder, and J. Lawless, "Failure inference from a marker process based on a bivariate wiener model," Lifetime Data Anal., vol. 4, pp. 229–251, 1998.

[12] S. Mishra and O. A. Vanli, "Remaining useful life estimation with lamb-wave sensors based on wiener process and principal components regression," J. Nondestr. Eval., vol. 35, pp. 1–13, 2016.

[13] I. Muhammad, X. Wang, C. Li, M. Yan, M. Mukhtar, and M. Muhammad, "Reliability analysis with wiener-transmuted truncated normal degradation model for linear and non-negative degradation data," Symmetry, vol. 14, no. 2, p. 353, 2022.

[14] X.-S. Si, "An adaptive prognostic approach via nonlinear degradation modeling: Application to battery data," IEEE Trans. Ind. Electron., vol. 62, no. 8, pp. 5082–5096, 2015.

[15] Z. Wang, Q. Zhai, and L. Shen, "Degradation modeling and rul prediction in dynamic environments using a wiener process with an autoregressive rate," IEEE Trans. Reliab., 2023.

[16] Q. Zhai and Z.-S. Ye, "Rul prediction of deteriorating products using an adaptive wiener process model," IEEE Trans. Ind. Inform., vol. 13, no. 6, pp. 2911–2921, 2017.

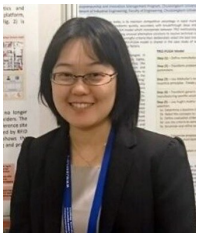
[17] E. M. Omshi, A. Grall, and S. Shemehsavar, "A dynamic auto-adaptive predictive maintenance policy for degradation with unknown parameters," Eur. J. Oper. Res., vol. 282, no. 1, pp. 81–92, 2020.

[18] M. Djeziri, S. Benmoussa, M. S. Mouchaweh, and E. Lughofer, "Fault diagnosis and prognosis based on physical knowledge and reliability data: Application to mos field-effect transistor," Microelectron. Reliab., vol. 110, p. 113682, 2020.

[19] J. Hu, Q. Sun, Z.-S. Ye, and Q. Zhou, "Joint modeling of degradation and lifetime data for rul prediction of deteriorating products," IEEE Trans. Ind. Inf., vol. 17, no. 7, pp. 4521–4531, 2020.
[20] J. Zhao, Y. Zhou, Q. Zhu, Y. Song, Y. Liu, and H. Luo, "A remaining useful life prediction method of aluminum electrolytic capacitor based on wiener process and similarity measurement," Microelectron. Reliab., vol. 142, p. 114928, 2023.
[21] Z.-S. Ye, Y. Wang, K.-L. Tsui, and M. Pecht, "Degradation data analysis using wiener processes with measurement errors," IEEE Trans. Rel., vol. 62, no. 4, pp. 772–780, 2013.
[22] Y. Lei, N. Li, and J. Lin, "A new method based on stochastic process models for machine remaining useful life prediction," IEEE Trans. Instrum. Meas., vol. 65, no. 12, pp. 2671–2684, 2016.
[23] W. Peng, Z.-S. Ye, and N. Chen, "Joint online rul prediction for multivariate deteriorating systems," IEEE Trans. Ind. Inform., vol. 15, no. 5, pp. 2870–2878, 2018.
[24] G. A. Veloso and R. H. Loschi, "Dynamic linear degradation model: Dealing with heterogeneity in degradation paths," Reliab. Eng. Syst. Saf., vol. 210, p. 107446, 2021.
[25] J. Liu, B. Hou, M. Lu, and D. Wang, "Box-cox transformation based state-space modeling as a unified prognostic framework for degradation linearization and rul prediction enhancement," Reliab. Eng. Syst. Saf., vol. 244, p. 109952, 2024.
[26] Q. Zhai and Z.-S. Ye, "Robust degradation analysis with non-gaussian measurement errors," IEEE Trans. Instrum. Meas., vol. 66, no. 11, pp. 2803–2812, 2017.
[27] R. Ge, Q. Zhai, H. Wang, and Y. Huang, "Wiener degradation models with scale-mixture normal distributed measurement errors for rul prediction," Mech. Syst. Signal Process., vol. 173, p. 109029, 2022.
[28] X. Liu, X. Wang, M. Xie, and Z. Ye, "Robust degradation state identification in the presence of parameter uncertainty and outliers," IEEE Trans. Ind. Inform., 2023.
[29] P. J. Huber, Robust statistics. John Wiley & Sons, 2004, vol. 523.
[30] W. J. Rey, Introduction to robust and quasi-robust statistical methods. Springer Science & Business Media, 2012.
[31] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization." J. Mach. Learn. Res., vol. 12, no. 7, 2011.
[32] C. M. Bishop and N. M. Nasrabadi, Pattern recognition and machine learning. Springer, 2006, vol. 4, no. 4.
[33] N. Branchini and V. Elvira, "An adaptive mixture view of particle filters," Found. Data Sci., pp. 0–0, 2024.
[34] B. Saha and K. Goebel, "Battery data set," 2010, https://c3.nasa.gov/dashlink/resources/133/, Last accessed on 2023-05-19.
[35] J. Zhang, Y. Yang, and J. Ding, "Information criteria for model selection," Wiley Interdiscip. Rev. Comput. Stat., vol. 15, no. 5, p. e1607, 2023.
[36] H. Zhang, P. Borghesani, R. B. Randall, and Z. Peng, "A benchmark of measurement approaches to track the natural evolution of spall severity in rolling element bearings," Mech. Syst. Signal Process., vol. 166, p. 108466, 2022.
[37] L. Fan, W. Lin, X. Chen, H. Yin, and Y. Chai, "Degradation path approximation for remaining useful life estimation," Adv. Eng. Inf., vol. 60, p. 102422, 2024.

Xingchen Liu received the B.E. degree in Measurement and Control Technology and Instruments from Hunan University, Changsha, China, in 2014, the M.S. degree in Instruments Science and Technology from the University of Science and Technology of China, Hefei, China, in 2017, and the Ph.D. degree in Industrial Systems Engineering and Management from the National University of Singapore, Singapore, in 2021.

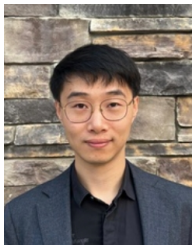
He is currently a Research Assistant Professor at the Department of Industrial and Systems Engineering, Hong Kong Polytechnic University, Hong Kong SAR, China. His research interests include machine learning, process monitoring, reliability engineering, and prognostics and health management.



Carman K.M. Lee is currently an associate professor in the Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hong Kong. She is the program leader of BSc(Hons) Enterprise Engineering with Management. She obtained her PhD and BEng degree from The Hong Kong Polytechnic University. Her main research areas include Logistics and Supply Chain Management, Industrial Internet of Things (IIoT), Cyber Physical System, Data Analytics and Swam Intelligence Optimization. As of now, Dr Lee has published over 130 articles in various international journals and seminars. She was awarded Silver Medal in the 47th International Exhibition of Inventions of Geneva in 2019 and Outstanding Paper Award of Emerald Network Awards in 2019. Dr Lee also serves as the Lab-in-Charge of PolyU's Cyber-Physical Systems Laboratory.

Jingyuan Huang received the BEng degree in Product Analysis with Design from The Hong Kong Polytechnic University in 2010, and MPhil and PhD degrees in Mechanical Engineering from the Hong Kong University of Science and Technology in 2012 and 2022, respectively.

She is currently the assistant program manager at the Center for Advances in Reliability and Safety. Her research interests include product design, electronic packaging and testing, advanced materials, heat and mass transfer and its theoretical modeling, big data analysis, and reliability and safety.



Qiuzhuang Sun received the joint B.E. degree in industrial engineering and in computer science fromShanghai JiaoTong University, Shanghai, China, in 2015, and the Ph.D. degree in industrial and systems engineering from the National University of Singapore, Singapore, in 2019.

He is currently a Lecturer at the University of Sydney. His research interests include maintenance optimization, degradation modeling, and data-driven decision making.