Reliability Engineering and System Safety 216 (2021) 107928

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

A hierarchical modeling approach for degradation data with mixed-type covariates and latent heterogeneity

Xuxue Sun [a], Wenjun Cai [b], Mingyang Li [c,*]

[a] College of Media Engineering, Communication University of Zhejiang, 998 Xueyuan St, Hangzhou, Zhejiang 310019, China
[b] Department of Materials Science and Engineering, Virginia Tech, 460 Old Turner St., Blacksburg VA, 24061, USA
[c] Department of Industrial and Management Systems Engineering, University of South Florida, 4202 E. Fowler Avenue, Tampa, FL 33620, USA

ARTICLE INFO

Keywords:
Degradation data
Mixed-type covariates
Latent heterogeneity
Data augmentation
Functional data analysis

ABSTRACT

Successful modeling of degradation data with covariates is essential for accurate reliability assessment of highly reliable product units. Due to the influences of different types of covariates, such as the external factors (e.g. accelerated operating conditions) and the internal factors (e.g. material microstructure characteristics), as well as latent heterogeneity due to the influences of the unobserved or unknown factors shared within each product unit, the degradation measurements of product units are highly heterogeneous over time. Many of existing degradation models often failed to simultaneously consider the influences of (i) both external accelerated conditions and internal material information, (ii) latent heterogeneity, and (iii) multiple material types. In this work, we propose a generic degradation modeling approach with mixed-type (e.g. both scalar and functional) covariates and latent heterogeneity to account for both the influences of observed internal and external factors as well as their interaction, and the influences of unobserved factors. Effective estimation algorithm is developed under expectation–maximization framework to jointly quantify the influences of mixed-type covariates and individual latent heterogeneity. The proposed algorithms further enables closed-form updating of model parameters at each iteration to ensure the estimation convenience. A real case study is provided to illustrate the proposed modeling approach and to demonstrate its effectiveness from both model prediction and interpretation perspectives.

1. Introduction

Accurate modeling of degradation data is of great importance for achieving accurate reliability assessment and failure prediction for highly reliable product units with few and zero failure observations. Due to the varied product characteristics and the influences of the environmental conditions, such as accelerated operating conditions (e.g. elevated voltage, load, temperature), product units often exhibit highly heterogeneous performance degradation over time. To account for such heterogeneity of degradation data and to improve accuracy of reliability assessment, many of existing degradation models focused on modeling degradation data with covariates by incorporating various external influencing factors and quantifying their influences [1,2]. However, there is limited research about extracting and incorporating reliability relevant material characteristics of product units as internal factors, and further integrating them with external accelerated conditions to improve prediction accuracy of performance degradation.

Rich material information of product units, such as microstructure characteristics, become readily available or can be easily accessible with the advancement of sensing technologies and material property characterization techniques, such as scanning electron microscope and transmission electron microscope [3]. To extract rich characteristics information from material microstructure images, engineers often utilize informative material descriptors [4], such as two-point correlation function, radial distribution function and lineal-path function. Many of these material descriptors are represented in functional form rather than scalar form. For instance, the two-point correlation is a functional feature curve over spatial distance to describe the spatial heterogeneity of material microstructure at microscopic scale, which often reflects the reliability-related product properties at macroscopic level (e.g. strength, hardness). Incorporating such functional covariates as internal reliability influencing factors and integrating them together with other external factors has a great opportunity for improving the accuracy of degradation modeling as compared to existing models which only consider external factors.

In the existing literature of degradation modeling, different data-driven models have been developed, such as continuous stochastic

* Corresponding author.
E-mail address: mingyangli@usf.edu (M. Li).

https://doi.org/10.1016/j.ress.2021.107928
Received 26 January 2021; Received in revised form 10 July 2021; Accepted 13 July 2021
Available online 8 August 2021

Nomenclature

$\alpha_{ls}(\cdot)$	The coefficient function of the sth functional covariate contributing to the coefficient η_{li}
$\boldsymbol{\beta}_l$	The coefficient vector associated with scalar covariates that contributes to the coefficient η_{li}
$\boldsymbol{\epsilon}_i$, ϵ_{ij}	The vector of error term and its jth element for test unit i
$\boldsymbol{\gamma}_i$, γ_{li}	The latent vector and its lth element for test unit i
$\boldsymbol{t}_i$, t_{ij}	The vector of measurement time and its jth element for test unit i
$\boldsymbol{x}_i$, x_{ip}	The vector of scalar covariates and its pth element for test unit i
$\boldsymbol{y}_i$, y_{ij}	The vector of degradation measurements and its jth element of unit i
η_{li}	The lth basis coefficient of test unit i at response level model
ν_l	The intercept parameter contributing to the coefficient η_{li}
$\phi_l(\cdot)$	The lth basis function of test unit i at response level model
$\psi_k(\cdot)$	The kth basis function in the approximation of functional data
$\rho_{lps}(\cdot)$	The coefficient function of interaction term between the sth functional covariate and the pth scalar covariate that contributes to coefficient η_{li}
σ_ϵ^2	The variance of error term ϵ_{ij}'s
Σ_γ	The variance matrix of latent vector $\boldsymbol{\gamma}_i$'s
$\sigma_{\gamma l}^2$	The variance of latent variable γ_{li}'s
c_{isk}, b_{lsk}, b'_{lpsk}	The kth basis coefficient of $Z_{is}(\cdot)$, $\alpha_{ls}(\cdot)$ and $\rho_{lps}(\cdot)$, respectively
R	The range of spatial distance measured in pixel unit
r	The spatial distance measured in pixel unit
$Z_{is}(\cdot)$	The sth functional covariate for test unit i

process [5,6], Markov-based models [7,8]. Many of these modeling approaches mainly focused on characterizing the heterogeneity of degradation data as a whole without explicitly incorporating covariates as additional inputs to explain part of the heterogeneity. To characterize the heterogeneity of degradation data with covariates, existing degradation models often considered scalar covariates which represented the external factors, such as environmental conditions [9,10]. There are limited studies which account for the influence of material characteristics on performance degradation. Park et al. [11] incorporated a scalar covariate into degradation modeling to represent the aggregate-level material information. Si et al. [12] recently considered functional covariate in their degradation model and incorporated detailed microstructure information. However, these approaches failed to jointly consider both the mixed-type (i.e., functional and scalar) covariates and their potential interaction. Moreover, the latent heterogeneity caused by the influence of many unobserved or unknown factors shared within each product unit was not addressed by the above models either.

In recent years, there were several degradation models which considered latent heterogeneity. The existing stochastic process-based degradation models with latent heterogeneity [13] mainly used random drift and/or random diffusion parameters to capture unit-specific heterogeneity, such as Wiener process model with random drift rate [14,15] and random diffusion parameter [16,17], inverse Gaussian process model with random parameter [18], and gamma degradation process with random parameter [19]. However, most of these stochastic process-based models with latent heterogeneity either failed to incorporate individual covariates to explain part of observed heterogeneity [14,15,20], or only used single type covariate of external accelerated conditions [19] without incorporating material characteristics. Moreover, the existing linear or nonlinear degradation path models with latent heterogeneity often used random parameters to capture unit-to-unit variation [1,21]. These degradation path models with latent heterogeneity either failed to consider the influence of covariates, such as general path models [1,22], or only considered external factors [23,24] without considering internal material information. Overall, most of existing approaches failed to simultaneously account for the influences of both external accelerated conditions and internal material information, their potential interaction, and individual latent heterogeneity as well.

To address the above research gaps, we propose a generic statistical degradation modeling framework to account for both the influences of observed mixed-type covariates and the latent heterogeneity. The mixed-type covariates consist of (i) the functional covariates which capture the internal material microstructure characteristics of product units, and (ii) the scalar covariates which capture the external environmental conditions elevated in the context of accelerated degradation test. Moreover, a model estimation algorithm is developed to jointly quantify the influences of mixed-type covariates and the latent heterogeneity, and further to examine the potential interaction between functional and scalar covariates. Functional data analysis and data augmentation techniques are employed to address a series of estimation challenges, such as the infinite dimensionality of functional covariates and joint estimation of the influences of observed factors and latent variables. To demonstrate the effectiveness of the proposed approach, we present a real case study using accelerated tribological degradation data of test units of copper alloys. We also conduct a comprehensive comparison study to demonstrate the appealing modeling performance of the proposed approach.

The rest of this paper is organized as follows. Section 2 first provides the notations, then describes the formulation of proposed degradation modeling framework and introduces the concepts of functional material descriptor, followed by a detailed elaboration of the developed estimation algorithm. Section 3 presents a real case study to illustrate the proposed work and further to demonstrate its effectiveness via comparison with several alternative models. Section 4 draws the conclusive remarks of this paper.

2. Methodology

To capture the influences of both accelerated conditions and material characteristics as well as the influence of unobserved factors, we propose a novel degradation modeling framework with mixed-type covariates and latent heterogeneity. The proposed framework consists of three modules, such as data collection, data processing, and degradation model development, as shown in Fig. 1. The multi-source reliability relevant data for a sample of test units, such as their degradation measurements, accelerated conditions and material microstructure images, are first collected, as shown in Module I on the top-left of Fig. 1. Given such multi-source data, the proposed framework begins with extracting and quantifying rich material characteristics information from microstructure images as functional covariates, as shown in Module II on the top-right of Fig. 1. Then, a generic model formulation is established to integrate multi-source data by comprehensively incorporating both mixed-type (e.g. scalar and functional) covariates and their potential interaction as well as unobserved factors shared with each test unit. Based on the established model, estimation algorithm is further developed to address a series of theoretical difficulties during model estimation, such as the infinite dimensionality issue of functional data and the quantification of latent heterogeneity of individual test unit, as shown in Module III on the bottom of Fig. 1. The technical details of the proposed work will be elaborated in the following subsections.

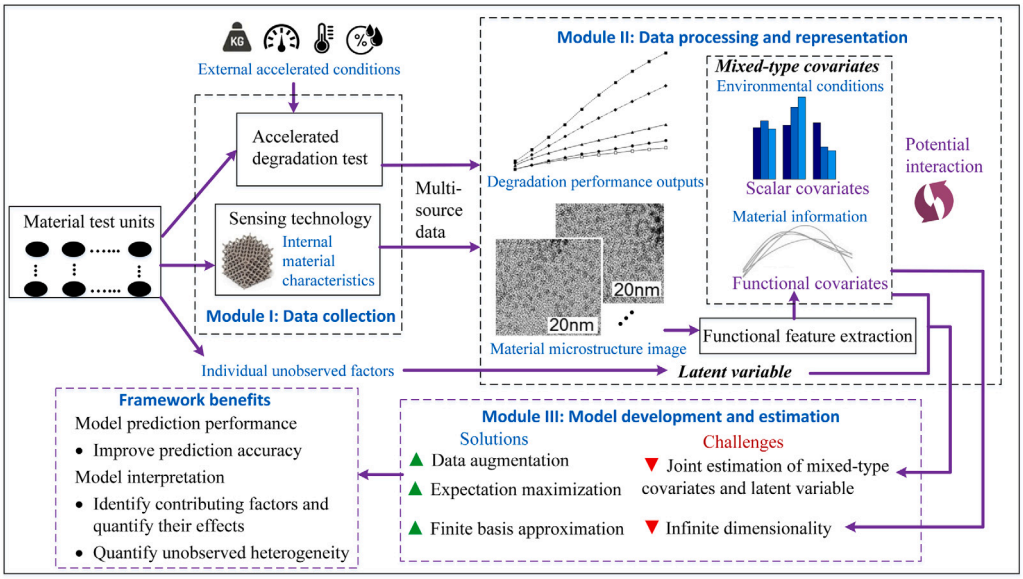

Fig. 1. Overview of proposed degradation modeling framework.

2.1. Model formulation

Considering a population of N test units, the degradation measurement of test unit i at time t_{ij} is denoted as $y_{ij}, \forall i = 1, \ldots, N,\ j = 1, \ldots, m_i$ where m_i is total number of degradation measurements of test unit i. The performance degradation of each test unit may be influenced by both the external factors (e.g. accelerated operating conditions) and the internal factors (e.g. material characteristics). In this paper, we use scalar covariates to represent a specific stress level of the accelerated conditions and further use functional descriptors to quantify the influence of baseline material characteristics due to their rich representation of microstructure information. Moreover, it is possible that the individual latent heterogeneity may exist due to the influence of unobserved or unknown factors shared within each test unit. To simultaneously quantify the influences of mixed-type (i.e., scalar and functional) covariates and their potential interaction as well as the latent heterogeneity, the proposed hierarchical degradation model can be generically formulated as

Response level:

$$y_{ij} = g_i(t_{ij}, \boldsymbol{\theta}_i, \boldsymbol{\Psi}) + \epsilon_{ij} \approx \sum_{l=0}^{L} \eta_{li}\phi_l(t_{ij}) + \epsilon_{ij}, \quad i = 1, \ldots, N, j = 1, \ldots, m_i \tag{1a}$$

Coefficient level:

$$\eta_{li} = \nu_l + \boldsymbol{\beta}_l^{\mathrm{T}}\boldsymbol{x}_i + \sum_{s=1}^{S}\int_{\mathbb{R}_s} \alpha_{ls}(r)Z_{is}(r)\mathrm{d}r + \sum_{p=1}^{P} x_{ip}\Big(\sum_{s=1}^{S}\int_{\mathbb{R}_s} \rho_{lps}(r)Z_{is}(r)\mathrm{d}r\Big) + \gamma_{li}, \quad i = 1, \ldots, N, l = 0, \ldots, L \tag{1b}$$

At response level, the degradation responses over time of test unit i can be captured by a unit-specific nonlinear function mapping $g_i(\cdot)$ and an error term $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$ where σ_ϵ^2 is the variance of measurement error. Among the parameters of $g_i(\cdot)$, $\boldsymbol{\Psi}$ is a set of fixed unknown parameters that are common among all test units, and $\boldsymbol{\theta}_i$ is a set of random unknown parameters of test unit i. To improve the model interpretation and estimation tractability, the nonlinear function $g_i(\cdot)$ can be further approximated by a set of basis functions $\{\phi_l(\cdot), \forall l = 0, \ldots, L\}$ and unit-specific basis coefficients $\{\eta_{li}, \forall l = 0, \ldots, L\}$ [25]. The linear additive form of basis functions and linear basis coefficients ensures model interpretability. When $l = 0$, $\phi_0(\cdot) = 1$ and η_{0i} is the intercept. When $l > 0$, different basis functions, such as polynomial basis and spline basis, can be considered for $\phi_l(\cdot)$ to capture the nonlinear curvature of performance degradation over time [26]. The unit-specific basis coefficients capture the individual heterogeneity of degradation data.

At coefficient level, the individual heterogeneity captured by each basis coefficient η_{li} can then be decomposed into five linear additive components, namely, (i) population-level component ν_l contributing to the lth coefficient, which captures the population average degradation pattern among all test units; (ii) part of the individual heterogeneity contributing to the lth coefficient, which is explained by the marginal effect of the observed scalar covariates $\boldsymbol{x}_i = [x_{i1}, \ldots, x_{iP}]^{\mathrm{T}}$ and covariates coefficients $\boldsymbol{\beta}_l = [\beta_{l1}, \beta_{l2}, \ldots, \beta_{lP}]^{\mathrm{T}}$; (iii) part of the individual heterogeneity contributing to the lth coefficient, which is explained by the marginal effect of the observed functional covariates $Z_{is}(r)$ with support space $\mathbb{R}_s$ and covariate coefficient functions $\alpha_{ls}(r), \forall s = 1, \ldots, S$; (iv) part of the individual heterogeneity contributing to the lth coefficient, which is explained by the interaction between the observed scalar and functional covariates and covariate coefficient functions $\rho_{lps}(r), \forall p = 1, \ldots, P, s = 1, \ldots, S$, and (v) the latent heterogeneity contributing to the lth coefficient due to the influence of unobserved or unknown factors shared within each test unit i, captured by continuous latent variable, i.e., $\gamma_{li} \sim N(0, \sigma_{\gamma l}^2)$ where $\sigma_{\gamma l}^2$ is the variance of latent variable.

To better explain Eq. (1) (b), we assume that the nonlinear degradation pattern can be approximated by polynomial basis functions with finite order, i.e., $g_i(t_{ij}, \boldsymbol{\theta}_i, \boldsymbol{\Psi}) \approx \sum_{l=0}^{L} \eta_{li} t_{ij}^l$. The basis coefficient η_{1i} essentially captures the slope of degradation path for test unit i. Based on Eq. (1) (b), the slope coefficient can be decomposed into five parts, i.e., $\eta_{1i} = \nu_1 + \boldsymbol{\beta}_l^{\mathrm{T}}\boldsymbol{x}_i + \sum_{s=1}^{S}\int_{\mathbb{R}_s} \alpha_{1s}(r)Z_{is}(r)\mathrm{d}r + \sum_{p=1}^{P} x_{ip}(\sum_{s=1}^{S}\int_{\mathbb{R}_s} \rho_{1ps}(r)Z_{is}(r)\mathrm{d}r) + \gamma_{1i}$. ν_1 captures average slope of performance degradation among all test units. The second, third and fourth additive terms capture the influences of individual observed covariates on the slope of degradation path of test unit i, including the influence of scalar covariates, the influence of functional covariates and the influence of interaction between scalar and functional covariates, respectively. γ_{1i} captures the individual latent heterogeneity on the slope of degradation path of test unit i.

The proposed model formulation is generic and several of the existing degradation models can be treated as the special cases of the proposed model. For instance, by neglecting the mixed-type covariates and their potential interaction, i.e., $\beta_{lp} = 0, \alpha_{ls}(\cdot) = 0, \rho_{lps}(\cdot) = 0, \forall l =$

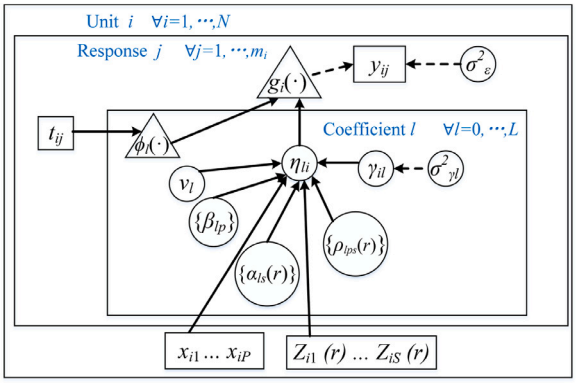

Fig. 2. The hierarchical structure of proposed degradation model.

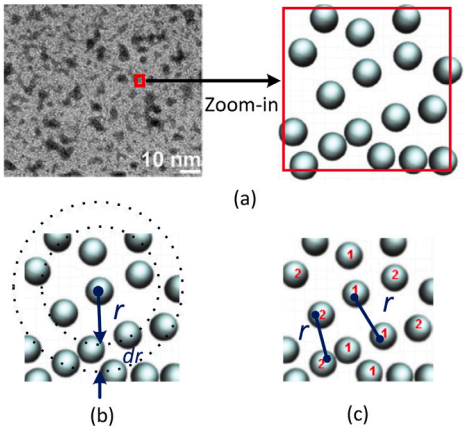

Fig. 3. TEM image and its functional descriptor, (a) raw TEM image and zoom-in view, (b) diagram of calculating radial distribution function, (c) diagram of calculating two-point correlation function.

$0,\ldots,L, p=1,\ldots,P, s=1,\ldots,S$, the proposed model is reduced into the degradation path model in [1]. In addition, by neglecting the functional covariates and the interaction term as well as latent heterogeneity, i.e., $\alpha_{ls}(\cdot)=0, \rho_{lps}(\cdot)=0, \gamma_{li}=0, \forall i=1,\ldots,N, l=0,\ldots,L, p=1,\ldots,P, s=1,\ldots,S$, the proposed model becomes the typical accelerated degradation model with constant stress factor. Moreover, by neglecting scalar covariates, potential interaction and latent heterogeneity, i.e., $\beta_{lp}=0, \rho_{lps}(\cdot)=0, \gamma_{li}=0, \forall l=0,\ldots,L, p=1,\ldots,P, s=1,\ldots,S$, the proposed model becomes the degradation model with functional covariates introduced in [12]. Fig. 2 further summarizes the hierarchical structure of the proposed model. As shown in the graphical representation, the rectangle nodes refer to the observed data, including degradation responses and various observed individual characteristics of each test unit (e.g. accelerated conditions, material characteristics). The circle nodes represent unknown model parameters to be estimated. The triangle nodes represent deterministic functions. The solid and dashed arrows indicate the deterministic and stochastic relationships among the connected nodes, respectively, which are displayed at coefficient layer, response layer and unit layer.

2.2. Material statistical descriptor

Among the mixed-type covariates in the above formulation, the functional covariates $Z_{is}(r), \forall i=1,\ldots,N, s=1,\ldots,S$ represent critical characteristics of material microstructure, such as spatial heterogeneity patterns, which are known to have indispensable impacts on the performance degradation [27]. To evaluate the spatial heterogeneity of material microstructure at microscopic level, advanced sensing devices or technologies, such as transmission electron microscopy (TEM), are often available for the characterization of material microstructure at finer scale. Fig. 3 (a) gives a motivating example of a TEM image of a test unit in the accelerated wear test considered in the paper. The gray part and black part represent two different material phases of that unit, which have different material compositions. As reflected in TEM image with two distinct colors, the material of such test unit is two-phase and often exhibits spatial heterogeneity and non-uniformity. As compared to two-phase material, a test unit with single-phase material has few color contrast or even single color in its TEM image and often exhibits homogeneous spatial patterns. To further extract and quantify spatial heterogeneity patterns of TEM images, the functional microstructure descriptors [4], such as radial distribution function (RDF) and two-point correlation (TPC) function, are popular choices of correlation-based statistical measures. In contrast to scalar covariates which summarize the aggregate-level information, functional covariates can capture spatial heterogeneity patterns more comprehensively. We will elaborate the details of RDF and TPC as follows.

RDF is a useful statistical measure of describing how particle density varies as a function of distance from a reference particle [28]. Considering a test unit consisting of M particles in a volume V, RDF can be calculated as $Z(\boldsymbol{r})=\frac{1}{\kappa}\left\langle\sum_{m\neq 0}\delta(\boldsymbol{r}-\boldsymbol{r}_m)\right\rangle$ where $\kappa=\frac{M}{V}$ is the average number density of particles and $\langle\cdot\rangle$ is the ensemble averaging operator. $\boldsymbol{r}_m, \forall m=1,\ldots,M$ refers to particle coordinates and $\delta(\cdot)$ is the Dirac delta function. Particularly, for a test unit with equivalent particles $1,\ldots,M-1$, RDF calculation can be further simplified as $Z(\boldsymbol{r})=V\frac{M-1}{M}\langle\delta(\boldsymbol{r}-\boldsymbol{r}_1)\rangle$. As shown in Fig. 3(b), RDF is a distance dependent measure and determines how many particles are within a distance between r and $r+\mathrm{d}r$ away from the reference particle. If more particles in a test unit are uniformly distributed, the number of particles within a specified distance from the reference particle over the entire support will be similar. This can further be reflected by the RDF values with less sharp changes involved in the RDF curve. On the other side, if the particles concentrate on certain area of a test unit, the RDF values of different radius over the entire support will become significantly different.

In addition to the RDF measure, TPC is another important functional measure of describing material microstructure characteristics and typically useful for two-phase material [29]. Considering a test unit with two-phase material, we define an indicator function $I^{(h)}(x), h=1,2$ as $I^{(h)}(x)=\begin{cases}1, & x\in V_h\\ 0, & x\in\bar{V}_h\end{cases}$ where V_h and $\bar{V}_h$ refer to the region occupied by phase h and the other phase, respectively. TPC then represents the probability of two randomly chosen points q_1 and q_2 being both in phase h, i.e., $Z^{(h)}(q_1,q_2)=\left\langle I^{(h)}(q_1)I^{(h)}(q_2)\right\rangle$, where $\langle\cdot\rangle$ is the ensemble averaging operator over the entire support of a test unit. Particularly, when the material of a test unit is statistically homogeneous, TPC can be calculated as $Z^{(h)}(q_1,q_2)=\left\langle I^{(h)}(q_1)I^{(h)}(q_1+r)\right\rangle$, where r is a specified spatial distance, as depicted in Fig. 3 (c). When the particles of same phase are uniformly distributed over a test unit, TPC values of different distances over the entire support tend to be similar.

Both of the RDF and TPC descriptors are distance dependent functions and can be used to describe the spatial heterogeneity patterns of material microstructure effectively. By explicitly incorporating such functional descriptors as functional covariates, as presented in Section 2.1, the modeling complexity of the proposed approach increases as compared to the conventional degradation models. Such increased modeling complexity also raises several challenges of model estimation, which will be elaborated and addressed with details in the subsequent section.

2.3. Model estimation

Considering a population of N deteriorating units are tested and m_i degradation measurements are collected on each unit $i, i=1,\ldots,N$,

the available data can be expressed as $\boldsymbol{D} = \{y_{ij}, t_{ij}, \boldsymbol{x}_i, Z_{is}(\cdot), \forall i = 1,\dots,N, j = 1,\dots,m_i, s = 1,\dots,S\}$. Let $\boldsymbol{\gamma}_i$ be a vector of length $L+1$ representing the latent variables of test unit i with $\boldsymbol{\gamma}_i \sim N(0, \boldsymbol{\Sigma}_\gamma)$, where $\boldsymbol{\Sigma}_\gamma = \left[\sigma^2_{\gamma 0}, \sigma^2_{\gamma 1}, \dots, \sigma^2_{\gamma L}\right]\mathbf{I}$. We further denote a set of unknown model parameters as $\boldsymbol{\Theta} = \{\nu_l, \boldsymbol{\beta}_l, \alpha_{ls}(\cdot), \rho_{lps}(\cdot), \sigma^2_\epsilon, \forall l = 0,\dots,L, p = 1,\dots,P, s = 1,\dots,S\}$. Suppose the support space of the extracted functional covariates from previous section is defined as $\mathbb{R}_s = [0, R]$ where R is the range of spatial distance, the marginal likelihood function of the proposed model in Eq. (1) can be expressed as

$$
\begin{aligned}
&L\left(\boldsymbol{\Theta}, \{\boldsymbol{\gamma}_i\}_{i=1}^N \mid \boldsymbol{D}\right) = \prod_{i=1}^{N} \int \prod_{j=1}^{m_i} p\left(y_{ij} \mid \boldsymbol{\Theta}, \boldsymbol{\gamma}_i\right) p(\boldsymbol{\gamma}_i) \mathrm{d}\boldsymbol{\gamma}_i \\
&\propto \prod_{i=1}^{N} \int \cdots \int \left|\sigma^2_\epsilon \mathbf{I}_{m_i}\right|^{-\frac{1}{2}} \exp\{-\frac{1}{2\sigma^2_\epsilon} \sum_{j=1}^{m_i} (y_{ij} - \sum_{l=0}^{L} [\nu_l + \boldsymbol{\beta}_l^{\mathrm{T}} \boldsymbol{x}_i \\
&+ \sum_{s=1}^{S} \int_0^R \alpha_{ls}(r) Z_{is}(r) \mathrm{d}r \\
&+ \sum_{p=1}^{P} x_{ip} (\sum_{s=1}^{S} \int_0^R \rho_{lps}(r) Z_{is}(r) \mathrm{d}r) + \gamma_{li}] \phi_l(t_{ij}))^2\} \, \mathrm{d}\gamma_{0i} \dots \mathrm{d}\gamma_{Li}
\end{aligned} \tag{2}
$$

where $|\cdot|$ refers to the matrix determinant operator and $\mathbf{I}_{m_i}$ is an $m_i \times m_i$ identity matrix. As shown in the above likelihood function, the intrinsic infinite dimensionality of functional data [30] makes the parameters estimation mathematically intractable. Besides, as shown in Eq. (2), the conventional maximum likelihood estimation method aims to maximize the marginal likelihood function [31]. The corresponding latent variables $\boldsymbol{\gamma}_i$ will be integrated out and cannot be estimated. To handle the infinite dimensionality of functional data and to achieve the joint estimation of model parameters and latent variables, we employ a series of estimation techniques and develop the estimation algorithm. The technical details are elaborated as follows.

The functional data can be treated as realizations of a stochastic process and intrinsically involve the infinite dimensionality issue. To address such issue, we employ approximation method to reduce the dimensionality of functional data and facilitate parameter estimation. Based on Mercer's theorem, the covariance matrix of functional data can be expressed by orthogonal eigenfunctions and ordered nonnegative eigenvalues [30]. With these eigenfunctions, we can apply Karhunen–Loève expansion [32] on the centered functional covariates as well as the coefficient functions, and express them as linear combinations of the complete orthogonal basis functions, i.e., $Z_{is}(r) = \sum_{k=1}^{\infty} c_{isk}\psi_k(r)$, $\alpha_{ls}(r) = \sum_{k=1}^{\infty} b_{lsk}\psi_k(r)$ and $\rho_{lps}(r) = \sum_{k=1}^{\infty} b'_{lpsk}\psi_k(r), \forall i = 1,\dots,N, l = 0,\dots,L, p = 1,\dots,P, s = 1,\dots,S$ where c_{isk}, b_{lsk} and b'_{lpsk} are known as functional principle component scores of functional data [30]. Since the eigenvalues of covariance operator of functional data decrease and finally approximate to 0, it is often sufficient to use a small number of eigenfunctions whose eigenvalues are significantly nonzero to accurately approximate the functional data. The number of finite basis functions can be determined efficiently by the fraction of variance explained (FVE) in practice [33]. With the truncated K basis functions, the centered functional covariates and the coefficient functions can then be approximated by $Z_{is}(r) \approx \sum_{k=1}^{K} c_{isk}\psi_k(r)$, $\alpha_{ls}(r) \approx \sum_{k=1}^{K} b_{lsk}\psi_k(r)$ and $\rho_{lps}(r) \approx \sum_{k=1}^{K} b'_{lpsk}\psi_k(r)$. The model parameters then become $\boldsymbol{\Theta} = \{\nu_l, \boldsymbol{\beta}_l, b_{lsk}, b'_{lpsk}, \sigma^2_\epsilon, \forall l = 0,\dots,L, p = 1,\dots,P, s = 1,\dots,S, k = 1,\dots,K\}$. The joint likelihood can be rewritten as

$$
\begin{aligned}
&L\left(\boldsymbol{\Theta}, \{\boldsymbol{\gamma}_i\}_{i=1}^N \mid \boldsymbol{D}\right) \propto \prod_{i=1}^{N} \int \cdots \int \left|\sigma^2_\epsilon \mathbf{I}_{m_i}\right|^{-\frac{1}{2}} \exp\{-\frac{1}{2\sigma^2_\epsilon} \\
&\times \sum_{j=1}^{m_i} (y_{ij} - \sum_{l=0}^{L} [\nu_l + \boldsymbol{\beta}_l^{\mathrm{T}} \boldsymbol{x}_i \\
&+ R(\sum_{s=1}^{S}\sum_{k=1}^{K} b_{lsk}c_{isk}) + R(\sum_{p=1}^{P} x_{ip} (\sum_{s=1}^{S}\sum_{k=1}^{K} b'_{lpsk}c_{isk}))
\end{aligned}
$$

$$
+ \gamma_{li}] \phi_l(t_{ij}))^2\} \, \mathrm{d}\gamma_{0i} \dots \mathrm{d}\gamma_{Li} \tag{3}
$$

As shown in Eq. (3), the infinite dimensionality issue of functional data is resolved and it becomes mathematically tractable to estimate $\boldsymbol{\Theta}$. We denote $\boldsymbol{\Lambda}_i$ as an $m_i \times (L+1)$ design matrix of latent heterogeneity. The detailed derivation of $\boldsymbol{\Lambda}_i$ can be found in Appendix A. We then denote $\boldsymbol{\Omega}_i$ as an $m_i \times U$ design matrix of observed heterogeneity and denote $\boldsymbol{\zeta}$ as the corresponding coefficient vector of length U where $U = (L+1)(1 + P + SK + PSK)$. $\boldsymbol{\Omega}_i$ can be manifested as $\boldsymbol{\Omega}_i = \begin{pmatrix} \boldsymbol{\Lambda}_i & \boldsymbol{A}_{2i} & \boldsymbol{A}_{3i} & \boldsymbol{A}_{4i} \end{pmatrix}$ where the detailed derivations of $\boldsymbol{A}_{2i}$, $\boldsymbol{A}_{3i}$ and $\boldsymbol{A}_{4i}$ can be also found in Appendix A. The coefficient vector of observed heterogeneity is written as $\boldsymbol{\zeta} = \left[\boldsymbol{\nu}^{\mathrm{T}}, \boldsymbol{\beta}_0^{\mathrm{T}}, \dots, \boldsymbol{\beta}_L^{\mathrm{T}}, \boldsymbol{b}_{01}^{\mathrm{T}}, \dots, \boldsymbol{b}_{LS}^{\mathrm{T}}, \boldsymbol{b}_{011}^{\prime\mathrm{T}}, \dots, \boldsymbol{b}_{LPS}^{\prime\mathrm{T}}\right]^{\mathrm{T}}$ where the derivation details are described in Appendix A. The vector form of the proposed model can then be expressed as $\boldsymbol{y}_i = \boldsymbol{\Omega}_i\boldsymbol{\zeta} + \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_i + \boldsymbol{\epsilon}_i$, $i = 1,\dots,N$. The likelihood function can be expressed in a more compact form as

$$
\begin{aligned}
&L(\boldsymbol{\Theta}, \{\boldsymbol{\gamma}_i\}_{i=1}^N \mid \boldsymbol{D}) \propto \prod_{i=1}^{N} \int \cdots \int \left|\sigma^2_\epsilon \mathbf{I}_{m_i}\right|^{-\frac{1}{2}} \\
&\times \exp(-\frac{1}{2\sigma^2_\epsilon} \|\boldsymbol{y}_i - \boldsymbol{\Omega}_i\boldsymbol{\zeta} - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_i\|^2) \, \mathrm{d}\gamma_{0i} \dots \mathrm{d}\gamma_{Li}
\end{aligned} \tag{4}
$$

As shown in Eq. (4), the latent variables $\boldsymbol{\gamma}_i, \forall i = 1,\dots,N$ will be integrated out and cannot be estimated directly. To address the estimation issue of latent factors in the marginal approach, we employ data augmentation technique [34] and introduce the complete data, i.e., $\boldsymbol{D}^* = \{\boldsymbol{D}, \{\boldsymbol{\gamma}_i\}_{i=1}^N\}$. The model parameters can be specified as $\boldsymbol{\Theta} = \{\boldsymbol{\zeta}, \sigma^2_\epsilon, \boldsymbol{\Sigma}_\gamma\}$. Based on the augmented data $\boldsymbol{D}^*$, the complete data likelihood can be expressed as

$$
\begin{aligned}
&L(\boldsymbol{\Theta} \mid \boldsymbol{D}^*) = \prod_{i=1}^{N} p\left(\boldsymbol{y}_i \mid \boldsymbol{\Theta}, \boldsymbol{\gamma}_i\right) p(\boldsymbol{\gamma}_i \mid \boldsymbol{\Theta}) \\
&\propto \prod_{i=1}^{N} (\left|\sigma^2_\epsilon \mathbf{I}_{m_i}\right|^{-\frac{1}{2}} \left|\boldsymbol{\Sigma}_\gamma\right|^{-\frac{1}{2}} \exp(-\frac{1}{2}\boldsymbol{\gamma}_i{}^{\mathrm{T}} \boldsymbol{\Sigma}_\gamma^{-1} \boldsymbol{\gamma}_i)) \cdot \exp(-\frac{1}{2\sigma^2_\epsilon} \|\boldsymbol{y} - \boldsymbol{\Omega}\boldsymbol{\zeta} - \boldsymbol{\Lambda}\boldsymbol{\gamma}\|^2)
\end{aligned} \tag{5}
$$

where $\|\cdot\|$ is Euclidean norm operator. $\boldsymbol{y} = [\boldsymbol{y}_1^{\mathrm{T}}, \dots, \boldsymbol{y}_N^{\mathrm{T}}]^{\mathrm{T}}$ is a vector of length $\sum_{i=1}^N m_i$ representing the degradation measurements for all test units. $\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Omega}_1^{\mathrm{T}} & \dots & \boldsymbol{\Omega}_N^{\mathrm{T}} \end{pmatrix}^{\mathrm{T}}$ is a matrix of dimension $(\sum_{i=1}^N m_i) \times U$ representing the design matrix of observed heterogeneity for all test units. $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}_1 \bigoplus \boldsymbol{\Lambda}_2 \dots \bigoplus \boldsymbol{\Lambda}_N$ is a matrix of dimension $(\sum_{i=1}^N m_i) \times (L+1)N$ representing the design matrix of latent heterogeneity for all test units. $\boldsymbol{\gamma} = [\boldsymbol{\gamma}_1^{\mathrm{T}}, \boldsymbol{\gamma}_2^{\mathrm{T}}, \dots, \boldsymbol{\gamma}_N^{\mathrm{T}}]^{\mathrm{T}}$ is a vector of length $(L+1)N$ representing the latent variables among all units. The log-likelihood of augmented data can then be written as

$$
\begin{aligned}
&l(\boldsymbol{\Theta} \mid \boldsymbol{D}^*) \propto l_1(\boldsymbol{\zeta}, \sigma^2_\epsilon \mid \boldsymbol{D}^*) + l_2(\boldsymbol{\Sigma}_\gamma \mid \boldsymbol{D}^*) \\
&l_1(\boldsymbol{\zeta}, \sigma^2_\epsilon \mid \boldsymbol{D}^*) = -\frac{\sum_{i=1}^N m_i}{2} \log(\sigma^2_\epsilon) - \frac{1}{2\sigma^2_\epsilon} (\|\boldsymbol{y} - \boldsymbol{\Omega}\boldsymbol{\zeta}\|^2 \\
&\qquad + \sum_{i=1}^{N} \mathrm{Tr}(\boldsymbol{\Lambda}_i^{\mathrm{T}}\boldsymbol{\Lambda}_i\boldsymbol{\gamma}_i\boldsymbol{\gamma}_i^{\mathrm{T}}) - 2(\boldsymbol{y} - \boldsymbol{\Omega}\boldsymbol{\zeta})^{\mathrm{T}}\boldsymbol{\Lambda}\boldsymbol{\gamma}) \\
&l_2(\boldsymbol{\Sigma}_\gamma \mid \boldsymbol{D}^*) = -\frac{N}{2} \log\left|\boldsymbol{\Sigma}_\gamma\right| - \frac{1}{2} \sum_{i=1}^{N} \boldsymbol{\gamma}_i{}^{\mathrm{T}} \boldsymbol{\Sigma}_\gamma^{-1} \boldsymbol{\gamma}_i
\end{aligned} \tag{6}
$$

where $\mathrm{Tr}(\cdot)$ is the trace operator. Given that $\boldsymbol{\gamma}$ is known, $\boldsymbol{\zeta}$ and σ^2_ϵ can be obtained by maximizing $l_1(\boldsymbol{\zeta}, \sigma^2_\epsilon \mid \boldsymbol{D}^*)$ and $\boldsymbol{\Sigma}_\gamma$ can be estimated by maximizing $l_2(\boldsymbol{\Sigma}_\gamma \mid \boldsymbol{D}^*)$. However, since the latent variables are unknown, we employ EM technique [34,35] to estimate both unknown model parameters and latent variables iteratively. At iteration τ, the Expectation step yields the conditional expectation of $l(\boldsymbol{\Theta} \mid \boldsymbol{D}^*)$, i.e., $Q(\boldsymbol{\Theta}, \boldsymbol{\Theta}^{(\tau-1)}) = \mathbb{E}_{S(\boldsymbol{\gamma})|\boldsymbol{D},\boldsymbol{\Theta}^{(\tau-1)}}[l(\boldsymbol{\Theta} \mid \boldsymbol{D}^*)]$, where $S(\boldsymbol{\gamma}) = \left(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \dots, \boldsymbol{\gamma}_N, \boldsymbol{\gamma}_1\boldsymbol{\gamma}_1^{\mathrm{T}}, \dots, \boldsymbol{\gamma}_N\boldsymbol{\gamma}_N^{\mathrm{T}}\right)$ is a set of individual statistics and $\boldsymbol{\Theta}^{(\tau-1)}$ is a collection of all obtained model parameters at iteration $\tau - 1$. The Q-function can be explicitly

written as

$$Q(\boldsymbol{\Theta}, \boldsymbol{\Theta}^{(\tau-1)}) \propto -\frac{\sum_{i=1}^{N} m_i}{2} \log(\sigma_\epsilon^2) - \frac{1}{2\sigma_\epsilon^2}(\|\boldsymbol{y} - \boldsymbol{\Omega}\boldsymbol{\zeta}\|^2 + \sum_{i=1}^{N} \mathrm{Tr}(\boldsymbol{\Lambda}_i^{\mathrm{T}}\boldsymbol{\Lambda}_i \mathbb{E}[\boldsymbol{\gamma}_i\boldsymbol{\gamma}_i^{\mathrm{T}} \mid \boldsymbol{D}, \boldsymbol{\Theta}^{(\tau-1)}]) - 2(\boldsymbol{y} - \boldsymbol{\Omega}\boldsymbol{\zeta})^{\mathrm{T}}\boldsymbol{\Lambda}\mathbb{E}[\boldsymbol{\gamma} \mid \boldsymbol{D}, \boldsymbol{\Theta}^{(\tau-1)}]) - \frac{N}{2}\log\left|\boldsymbol{\Sigma}_\gamma\right| - \frac{1}{2}\sum_{i=1}^{N} \mathbb{E}[\boldsymbol{\gamma}_i{}^{\mathrm{T}}\boldsymbol{\Sigma}_\gamma^{-1}\boldsymbol{\gamma}_i \mid \boldsymbol{D}, \boldsymbol{\Theta}^{(\tau-1)}] \tag{7}$$

where $\mathrm{Tr}(\cdot)$ is the trace operator. The corresponding conditional expectation $\mathbb{E}[\boldsymbol{\gamma}_i \mid \boldsymbol{D}, \boldsymbol{\Theta}^{(\tau-1)}]$ and $\mathbb{E}[\boldsymbol{\gamma}_i\boldsymbol{\gamma}_i^{\mathrm{T}} \mid \boldsymbol{D}, \boldsymbol{\Theta}^{(\tau-1)}]$ can be explicitly obtained as

$$\begin{aligned}
&\mathbb{E}[\boldsymbol{\gamma}_i \mid \boldsymbol{D}, \boldsymbol{\Theta}^{(\tau-1)}] = \boldsymbol{\mu}_i^{(\tau-1)} = \frac{1}{\sigma_\epsilon^{2(\tau-1)}}\boldsymbol{V}_i^{(\tau-1)}\boldsymbol{\Lambda}_i^{\mathrm{T}}(\boldsymbol{y}_i - \boldsymbol{\Omega}_i\boldsymbol{\zeta}^{(\tau-1)}), \forall i = 1, \ldots, N \\
&\mathbb{E}[\boldsymbol{\gamma}_i\boldsymbol{\gamma}_i^{\mathrm{T}} \mid \boldsymbol{D}, \boldsymbol{\Theta}^{(\tau-1)}] = \boldsymbol{V}_i^{(\tau-1)} + \boldsymbol{\mu}_i^{(\tau-1)}\boldsymbol{\mu}_i^{(\tau-1)\mathrm{T}}, i = 1, \ldots, N
\end{aligned} \tag{8}$$

where $\boldsymbol{V}_i^{(\tau-1)} = (\boldsymbol{\Sigma}_\gamma^{-1(\tau-1)} + \frac{1}{\sigma_\epsilon^{2(\tau-1)}}\boldsymbol{\Lambda}_i^{\mathrm{T}}\boldsymbol{\Lambda}_i)^{-1}$. The derivation details of conditional expectation can be found in Appendix B. We can further simplify the conditional expectation as $\mathbb{E}[\boldsymbol{\gamma} \mid \boldsymbol{D}, \boldsymbol{\Theta}^{(\tau-1)}] = \boldsymbol{\mu}^{(\tau-1)} = [\boldsymbol{\mu}_1^{(\tau-1)\mathrm{T}}, \boldsymbol{\mu}_2^{(\tau-1)\mathrm{T}}, \ldots, \boldsymbol{\mu}_N^{(\tau-1)\mathrm{T}}]^{\mathrm{T}}$. With the calculated Q-function at current iteration, the maximization step achieves the maximization of $Q(\boldsymbol{\Theta}, \boldsymbol{\Theta}^{(\tau-1)})$. The model parameters at iteration τ can then be updated by $\boldsymbol{\Theta}^{(\tau)} = \arg\max_{\boldsymbol{\Theta}} Q(\boldsymbol{\Theta}, \boldsymbol{\Theta}^{(\tau-1)})$. It is mathematically tractable to maximize $Q(\boldsymbol{\Theta}, \boldsymbol{\Theta}^{(\tau-1)})$ and the closed-form of estimated model parameters at iteration τ can be expressed as

$$\hat{\boldsymbol{\zeta}}^{(\tau)} = \left(\boldsymbol{\Omega}^{\mathrm{T}}\boldsymbol{\Omega}\right)^{-1}\boldsymbol{\Omega}^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{\Lambda}\boldsymbol{\mu}^{(\tau-1)}) \tag{9}$$

$$\hat{\boldsymbol{\Sigma}}_\gamma^{(\tau)} = \frac{1}{N}\sum_{i=1}^{N}((\hat{\boldsymbol{\Sigma}}_\gamma^{(\tau-1)-1} + \frac{1}{\hat{\sigma}_\epsilon^{2(\tau-1)}}\boldsymbol{\Lambda}_i^{\mathrm{T}}\boldsymbol{\Lambda}_i)^{-1} + \boldsymbol{\mu}_i^{(\tau-1)}\boldsymbol{\mu}_i^{(\tau-1)\mathrm{T}}) \tag{10}$$

$$\hat{\sigma}_\epsilon^{2(\tau)} = \frac{1}{\sum_{i=1}^{N} m_i}(\left\|\boldsymbol{y} - \boldsymbol{\Omega}\hat{\boldsymbol{\zeta}}^{(\tau)}\right\|^2 - 2\sum_{i=1}^{N}(\boldsymbol{y}_i - \boldsymbol{\Omega}_i\hat{\boldsymbol{\zeta}}^{(\tau)})^{\mathrm{T}}\boldsymbol{\Lambda}_i\boldsymbol{\mu}_i^{(\tau-1)} + \sum_{i=1}^{N}\mathrm{Tr}(\boldsymbol{\Lambda}_i^{\mathrm{T}}\boldsymbol{\Lambda}_i((\hat{\boldsymbol{\Sigma}}_\gamma^{(\tau)-1} + \frac{1}{\hat{\sigma}_\epsilon^{2(\tau-1)}}\boldsymbol{\Lambda}_i^{\mathrm{T}}\boldsymbol{\Lambda}_i)^{-1} + \boldsymbol{\mu}_i^{(\tau-1)}\boldsymbol{\mu}_i^{(\tau-1)\mathrm{T}}))) \tag{11}$$

The derivation details of the above estimation procedure can be found in Appendix C.

Algorithm 1 Parameters estimation procedure of the proposed model

Initialization: $\boldsymbol{\Theta}^{(0)} = \{\boldsymbol{\zeta}^{(0)}, \sigma_\epsilon^{2(0)}, \boldsymbol{\Sigma}_\gamma^{(0)}\}$

procedure UpdateEstim

for $\iota \leftarrow 1, \ldots, \tau_{\max}$ **do**

Compute $\mathbb{E}[\boldsymbol{\gamma}_i \mid \boldsymbol{D}, \boldsymbol{\Theta}^{(\iota-1)}]$ and $\mathbb{E}[\boldsymbol{\gamma}_i\boldsymbol{\gamma}_i^{\mathrm{T}} \mid \boldsymbol{D}, \boldsymbol{\Theta}^{(\iota-1)}]$ based on Eq. (8)

Derive parameter estimates sequentially by maximizing $Q(\boldsymbol{\Theta}, \boldsymbol{\Theta}^{(\iota-1)})$

1. Obtain $\hat{\boldsymbol{\zeta}}^{(\iota)}$ based on Eq. (9)
2. Obtain $\hat{\boldsymbol{\Sigma}}_\gamma^{(\iota)}$based on Eq. (10)
3. Obtain $\hat{\sigma}_\epsilon^{2(\iota)}$ based on Eq. (11)

end for

end procedure

The estimation algorithm of the proposed model is summarized in Algorithm 1. At each iteration, the conditional expectations are calculated first and the model parameters are then updated iteratively via maximizing the Q-function. The updating procedure of parameter estimates will continue until the following stopping criterion is achieved, i.e., $\tau_{\max} = \arg\min\{\tau : \left\|\boldsymbol{\Theta}^{(\tau+1)} - \boldsymbol{\Theta}^{(\tau)}\right\| \le \epsilon_c\}$ where ϵ_c is a predefined tolerance level.

Further, we derive the interval estimates of model parameters $\boldsymbol{\zeta}$ via bootstrap method [36]. Based on the estimated parameters, we can simulate M bootstrap samples of performance degradation by Monte Carlo simulation, where M is sufficiently large number. For the mth bootstrap sample, $\forall m = 1 \ldots M$, we can employ Algorithm 1 to estimate model parameters, denoted as $\hat{\boldsymbol{\zeta}}_m$. The $100(1-\alpha)\%$ bootstrap confidence interval for single parameter $\hat{\varsigma}, \forall\hat{\varsigma} \in \hat{\boldsymbol{\zeta}}$, can be constructed as $\left(\hat{\varsigma}_{\frac{\alpha}{2}}, \hat{\varsigma}_{1-\frac{\alpha}{2}}\right)$ where $\hat{\varsigma}_q$ is the qth empirical quantile of $\left(\hat{\varsigma}_1, \ldots, \hat{\varsigma}_M\right)$ and α is the significance level.

Based on the above estimated model parameters, we further calculate important reliability characteristics, such as the remaining useful life (RUL) and its reliability function. The RUL of test unit i in degradation process is defined as the time duration from the current time point t_c till the time point when the performance degradation of test unit i reaches failure threshold H. The RUL of unit i can be expressed as

$$RUL_i = \inf\{t \in \mathbb{R}_{\ge 0} : y_i(t + t_c) \ge H \mid y_i(t_c) < H\} \tag{12}$$

where $y_i(t + t_c)$ is the performance degradation of test unit i at time $t + t_c$. The estimated reliability function of RUL of test unit i can then be derived as

$$\begin{aligned}
&\hat{S}_i = \frac{Pr(RUL_i > t)}{Pr(RUL_i \ge 0)} = \frac{\Phi(\hat{h}_i(t + t_c))}{\Phi(\hat{h}_i(t_c))} \\
&\hat{h}_i(d) = \frac{H - \hat{u}_i(d)}{(\hat{\sigma}_\epsilon^2 + \sum_{l=0}^{L}\hat{\sigma}_{\gamma l}^2\phi_l^2(d))^{\frac{1}{2}}}
\end{aligned} \tag{13}$$

where $\Phi(\cdot)$ is cumulative distribution function of standard normal distribution. $\hat{\sigma}_\epsilon$ is estimated standard deviation of error term ϵ_{ij} and $\hat{\sigma}_{\gamma l}^2$ is estimated variance component of latent variables γ_{li}. $\hat{u}_i(d)$ is the estimated mean degradation level of test unit i at time d in presence of observed covariates, i.e., $\hat{u}_i(d) \mid \hat{\boldsymbol{\Theta}} = \sum_{l=0}^{L}[\hat{v}_l + \hat{\boldsymbol{\beta}}_l^{\mathrm{T}}\boldsymbol{x}_i + R(\sum_{s=1}^{S}\sum_{k=1}^{K}\hat{b}_{lsk}c_{isk}) + R(\sum_{p=1}^{P}x_{ip}(\sum_{s=1}\sum_{k=1}\hat{b'}_{lpsk}c_{isk}))]\phi_l(d)$.

3. Real case study

3.1. Experimental data description

To illustrate the proposed modeling framework and to further evaluate its prediction performance as well as model interpretation capability, we provide a real case study to analyze the tribological degradation data of copper alloys. We will utilize the proposed degradation model to quantify the influences of both external and internal observed factors as well as latent heterogeneity within each individual test unit. Specifically, accelerated wear tests of Cu-Ni-Sn alloys are carried out at elevated load conditions using the Koehler K93500 pin-on-disk tester [27], as shown in Fig. 4. The test units consist of both as-received and annealed material specimens of Cu-Ni-Sn alloys. As compared to the as-received test units, the microstructure and physical properties of annealed test units are often altered considerably through the annealing process, and thus the corresponding tribological performance degradation may also differ. The as-received unit has greater hardness due to significantly distinct crystal structure while the annealed test unit has altered mechanical properties (e.g. reduced hardness) via the annealing process at higher temperature. For each test unit, multiple tribological degradation measurements, namely the height loss measurements in micrometer (um), are collected over time (in seconds) by a linear variable displacement transducer (LVDT) during an accelerated wear test [27]. Fig. 5 shows an example of heterogeneous performance degradation of a subset of test units with two different material types (e.g. as-received and annealed) under different

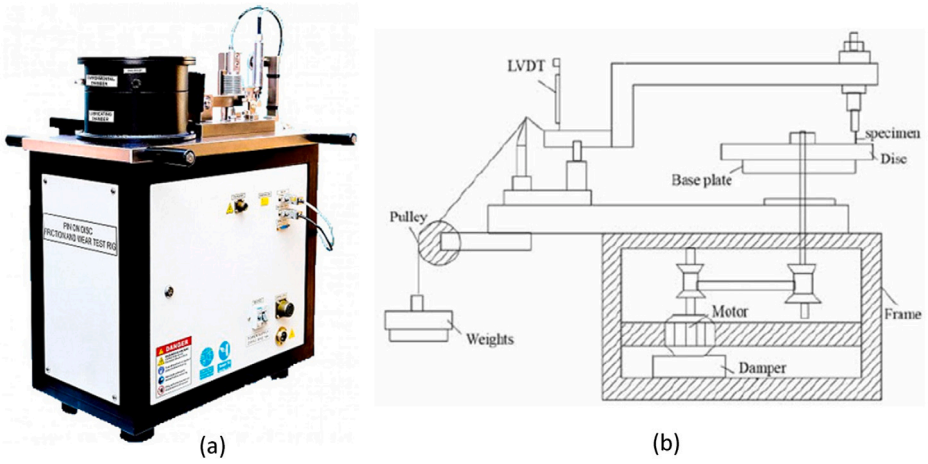

Fig. 4. The testing environment of the investigated accelerated wear test, (a) the testing equipment, (b) the diagram of wear test.

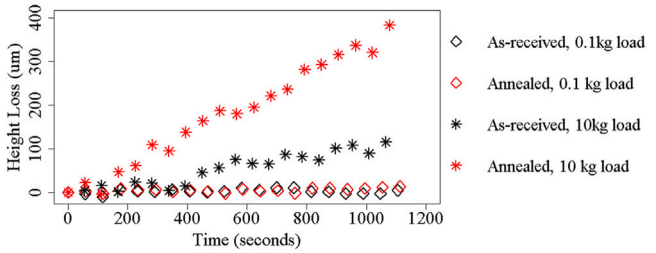

Fig. 5. Performance degradation of a subset of test units with different material types under different load conditions.

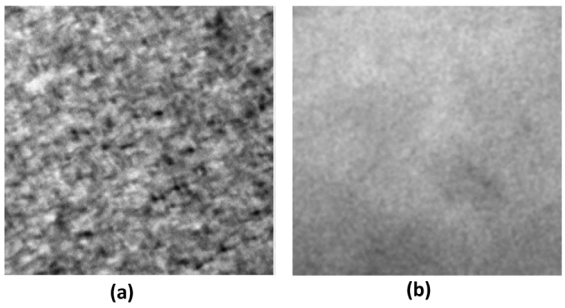

Fig. 6. TEM micrographs of two different material types: (a) as-received copper alloy, (b) annealed copper alloy.

load conditions (e.g. from 0.1 kg to 10 kg). As shown in Fig. 5, the height loss measurements have small values for both as-received and annealed test units when applied load is 0.1 kg, which implies that the degradation process is not significant when the stress factor is small. On the other side, when load becomes significantly larger (e.g. 10 kg), the degradation processes of both types of test units become significant and the two test units with different material types exhibit different degradation patterns. As shown in Fig. 5, both the internal factors (e.g. material types) and the external factors (e.g. load conditions) play important roles in the degradation pattern. There is a need to explicitly quantify the influences of both external load condition and internal material information and to explore their potential interaction.

To capture material microstructure characteristics, we consider transmission electron microscopy (TEM) images of the microstructures of both as-received and annealed test units. The TEM images embrace useful information about material microstructure at finer scale. Particularly, we use grayscale TEM images at a nanometer (nm) length scale (e.g. 20 nm) or less to characterize the microstructure properties of different types of copper alloys. Fig. 6 shows two TEM images of as-received and annealed test units, respectively, which have significant visual difference. The texture pattern of the annealed test unit, as illustrated in Fig. 6(b), has less spatial heterogeneity than the as-received test unit. This indicates a more uniformly homogeneous microstructure of the annealed test units. In the following section, we will show the analytical procedure of quantifying the spatial heterogeneity of single test unit at microscopic level and examining the microstructure difference of test units with different material types.

3.2. Functional covariates extraction

As shown in Fig. 6, the microstructures of the test units with different material types exhibit different spatial heterogeneity patterns. To capture the spatial heterogeneity with rich spatial information, two functional microstructure descriptors, namely TPC and RDF, are considered. As described in Section 2.2, the functional covariate $Z(r)$ is used to represent the material microstructure information where r is the distance measured at spatial scale in pixel. The distance measure for TPC function ranges from 0 to 200 pixels and the distance measure for RDF is between 0 and 250 pixels. Fig. 7 shows the corresponding functional covariates extracted based on TPC and RDF descriptors. The TPC (or RDF) value of an annealed test unit is uniformly larger than the value of same functional descriptor of an as-received unit at various spatial scales (measured in pixels). This implies a uniformly more homogeneous material microstructure of the annealed unit. As the spatial distance r increases, the microstructures of both test units tend to become less homogeneous (with a smaller TPC or RDF value). When the spatial distance r is larger than 25 pixels, the TPC (or RDF) value of as-received test unit tends to approach 0, indicating that the cluster size in its TEM image is typically no larger than 25 pixels. Further, TPC is more sensitive than RDF in capturing spatial heterogeneity of annealed unit at finer scale.

3.3. Performance comparison with alternative modeling approaches

We first apply centering to the raw degradation data such that their baseline degradation measurement (at $t = 0$) has been centered around 0, i.e., $\eta_{0i} = 0, \forall i = 1, \ldots, 12$. As shown in Fig. 5, the degradation data exhibits approximately linear degradation path and thus we consider the first order polynomial basis function for real data analysis, i.e., $\phi_1(t_{ij}) = t_{ij}, \forall i = 1, \ldots, 12, j = 1, \ldots, 20$. We then decompose the basis coefficient η_{1i} into several parts, as described in Section 2.1. We use the extracted functional covariate $Z_{i1}(r), \forall i = 1 \ldots 12$, as described in previous section, to represent the unit-specific microstructure pattern. In addition, we consider the scalar covariate x_{i1} to represent the load

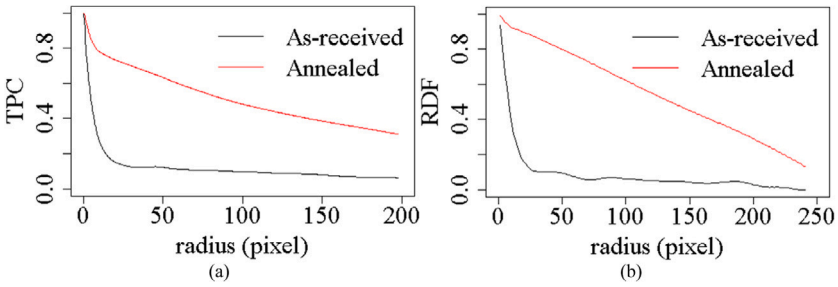

Fig. 7. Functional covariates extracted based on different functional microstructure descriptors: (a) TPC, (b) RDF.

Table 1
Model specification summary of different degradation modeling approaches.

Model	Observed factors				Latent heterogeneity	Order of polynomial basis	
	Material microstructure		Load condition	Interaction		first order	first and second orders
	Functional	Scalar	Scalar				
Model 1 [9]			✓			✓	
Model 2 [12]	✓					✓	
Model 3	✓		✓			✓	
Model 4	✓		✓	✓		✓	
Model 5	✓		✓	✓		✓	✓
Model 6 [11]		✓	✓	✓		✓	
Model 7	✓		✓	✓	✓	✓	

condition exerted on test unit i. By incorporating both $Z_{i1}(r)$ and x_{i1} as well as their interaction term into the proposed model, we further introduce a unit-specific latent variable γ_{1i} to capture the latent heterogeneity within each test unit i. We then employ the proposed estimation algorithm to jointly estimate the influences of mixed-type covariates and latent heterogeneity via setting the tolerance level ϵ_c as 1e-6 [35]. The estimation algorithm can achieve satisfactory convergence within the specified maximum number of iterations (e.g. 50 iterations) via multiple runs with different initial values. For basis function approximation in handling the infinite dimensionality of functional covariates during estimation, the optimal numbers of orthogonal basis functions for TPC and RDF are selected as 2 and 3, respectively, based on the criterion of FVE.

With the selected number of basis functions and the estimated models, we further evaluate the models with different functional descriptors (e.g. TPC and RDF) and compare their prediction performances. We partition the data into training (first 80% of degradation observations of each test unit) and test data sets. We employ 5-fold cross-validation (CV) to evaluate and compare the prediction performance of the two models. The total CV error of all test units using model with TPC descriptor is smaller than the total CV error of RDF-based model. Thus, we select TPC as microstructure descriptor to evaluate the performance of proposed modeling framework and further to compare it with other alternative models.

To investigate the importance of incorporating mixed-type covariates and their potential interaction, and also the importance of considering both observed heterogeneity and latent heterogeneity, we compare the proposed model with several models based on simplified model specifications. Many of existing degradation modeling approaches [9, 12] also explicitly or implicitly consider some of these model specifications. The alternative model specifications include: (i) Model 1, which only considers scalar covariate of load condition [9]; (ii) Model 2, which only considers functional covariate of material microstructure [12]; (iii) Model 3, which considers mixed-type covariates but fails to consider their interaction, and (iv) Model 4, which considers both mixed-type covariates and their interaction but fails to consider latent heterogeneity. Among the first four model specifications, Model 4 is the most complex one. Based on Model 4, we further adjust its modeling complexity from the following three aspects, namely (i) adding higher order basis function in Model 5; (ii) replacing functional covariate with scalar covariate to represent material microstructure in Model 6 [11]; and (iii) incorporating latent variable to represent latent heterogeneity in Model 7 (i.e., the proposed model). It is noticed that, for Model 6, we consider the scalar covariate of fractal dimension D_α to capture material microstructure, which is a popular choice of feature extraction method and often considered in microstructure image analysis [37]. Table 1 summarizes the differences between the proposed approach and six alternative models with different model specifications.

To comprehensively evaluate and compare the modeling performance of the above models, we consider the following different evaluation criteria, namely (i) model goodness-of-fit criteria, e.g. R-squared and data log-likelihood; (ii) model selection criteria, e.g. Akaike Information Criterion (AIC) [38] and Bayesian Information Criterion (BIC) [39], and (iii) model prediction criteria, e.g. mean squared error (MSE) of both training and test data, denoted as $\bar{\Delta}_{\text{train}}$ and $\bar{\Delta}_{\text{test}}$. The model selection criteria of AIC and BIC can be unified as

$$C(\boldsymbol{\Theta}) = -2\ln L(\hat{\boldsymbol{\Theta}}, \{\hat{\gamma}_i\}_{i=1}^N \mid \boldsymbol{D}) + pk \tag{14}$$

where $L(\hat{\boldsymbol{\Theta}}, \{\hat{\gamma}_i\}_{i=1}^N \mid \boldsymbol{D})$ is the maximized joint likelihood and pk is the penalty term to leverage the model complexity. In particular, p is total number of estimated model parameters, $k = 2$ (for AIC) and $k = \ln N$ (for BIC).

Based on the above criteria, Table 2 summarizes the results of performance comparison. As compared to the other models, the proposed model achieves comparable goodness-of-fit performance, the smallest AIC (or BIC) value and the smallest MSE values. Several additional implications and discussions can be obtained as follows. First, the better goodness-of-fit and prediction performances of Model 4 over Models 1–3 indicate the importance of incorporating both internal and external factors (as well as their potential interaction) during the degradation modeling. Second, the overfitting issue of Model 5 with extra quadratic term implies that the complex curvilinear model specification will not provide additional modeling benefits to the degradation data analysis in this study. Third, the better performance of Model 4 over Model 6 (which only considers single-type covariates) reflects the richer representation of microstructure information using functional covariates

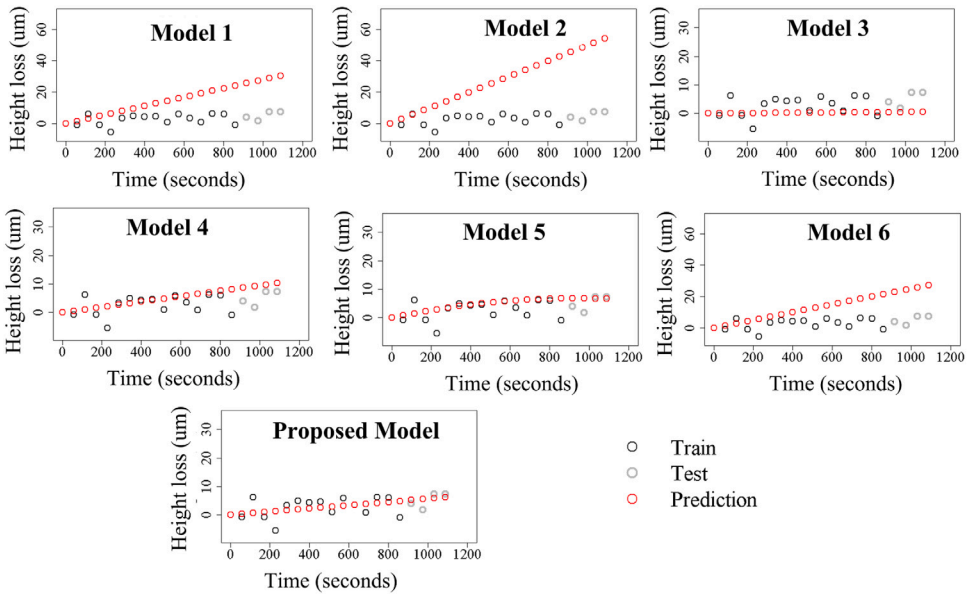

Fig. 8. Prediction performance comparison among different models.

Table 2
Model comparison results.

Model	Results					Model specification
	R^2	Log-likelihood	AIC	BIC	($\bar{\Delta}_{\text{train}}$, $\bar{\Delta}_{\text{test}}$)	
Model 1	0.73	−937.4	1880.8	1890.6	(1019.2, 3779.6)	$\boldsymbol{x}$
Model 2	0.57	−982.3	1972.6	1985.6	(1626.8, 6646.2)	$Z(r)$
Model 3	0.93	−811.1	1632.1	1648.4	(273.3, 1067.6)	$\boldsymbol{x} + Z(r)$
Model 4	0.97	−712.1	1438.3	1461.1	(97.5, 248.9)	$\boldsymbol{x} \times Z(r)$
Model 5	0.97	−712.1	1440.1	1466.2	(97.5, 253.6)	$\phi_1(t_{ij}) + \phi_2(t_{ij})$
Model 6	0.73	−936.3	1882.7	1898.9	(1007.9, 3748.8)	Z
Model 7	0.97	−698.7	1413.3	1439.4	(72.4, 156.3)	$\boldsymbol{x} \times Z(r) + \boldsymbol{\gamma}$

than conventional scalar covariates and further demonstrates the benefits of incorporating mixed-type covariates to improve the modeling accuracy. Last, by comparing the proposed model (which considers latent heterogeneity) with Model 4 (which fails to consider the influence of unobserved factors), both model selection results and prediction results indicate the importance of considering both the influences of observed factors and latent heterogeneity in developing the degradation model. The predicted values of performance degradation based on different models (e.g. Models 1–7) and the observed degradation data of a single test unit are also displayed in Fig. 8. As compared to the alternative modeling approaches (e.g. Models 1–6), the predicted values of performance degradation based on the proposed model are closer to the actual observations of degradation data over time.

3.4. Model interpretation

Based on the above performance evaluation and comparison, the satisfactory prediction performance of proposed modeling framework is demonstrated. We further investigate the model interpretability of the proposed work. The effects of different types of covariates on tribological degradation are discussed as below. Based on the proposed estimation algorithm, the point estimates and 95% confidence intervals via bootstrap sampling (with $M = 10\,000$) are obtained and summarized in Table 3. As shown in Table 3, both parameters $\hat{\nu}_1$ and $\hat{\beta}_{11}$ are significant at a significant level of 0.05. $\hat{\nu}_1$ captures the tribological degradation rate of studied alloy in absence of the influences of mixed-type covariates. A positive value of $\hat{\nu}_1$ indicates an increasing trend of performance degradation (i.e., material height loss) of copper alloys over time. $\hat{\beta}_{11}$ quantifies the marginal effects of scalar covariate,

Table 3
Model parameter estimation results.

Parameter	Point estimate	95% Confidence interval
ν_1	0.0163	(0.0054, 0.0273)
β_{11}	0.0202	(0.0178, 0.0227)
b_{111}	0.000025	(3.455e−6, 0.000047)
b_{112}	−0.00047	(−0.00163, 0.00069)
b'_{1111}	0.00002	(0.000011, 0.000029)
b'_{1112}	0.000326	(−0.000333, 0.000986)

i.e., load condition, on performance degradation of test units. A positive value of $\hat{\beta}_{11}$ indicates that load condition is an effective stress factor in accelerating the tribological degradation process of copper alloys.

Further, we investigate the marginal effect of functional covariates (i.e., TPC curves) on performance degradation. As shown in Eq. (1), $\int_0^R \alpha_{11}(r) Z_{i1}(r) \mathrm{d}r$ captures the contributing marginal effects of functional covariate $Z_{i1}(r)$ on the slope of degradation path, which can be written as $R(b_{111}c_{i11} + b_{112}c_{i12})$ via basis function approximation (in Section 2.3). A non-zero value of parameter b_{111} at a significance level of 0.05 indicates that there exists a marginal effect of material microstructures (captured by functional covariate) among different material types. Fig. 9 shows the estimated contributing marginal effects, $\int_0^R \alpha_{11}(r) Z_{i1}(r) \mathrm{d}r$, among different test units with two different types of copper alloys. As shown in Fig. 9, the microstructure effects of as-received test units are negative while the microstructure effects of annealed test units are positive. This indicates that a less homogeneous microstructure of as-received test units (as compared to annealed units) yields greater hardness and ultimately reduces the corresponding tribological degradation rate.

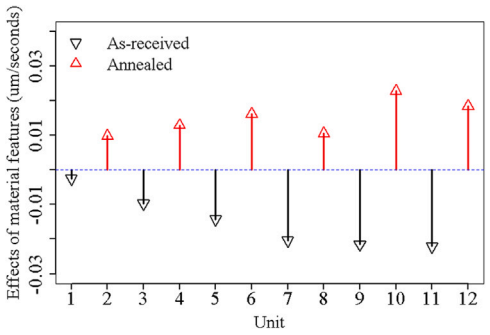

Fig. 9. Effects of internal material characteristics among all test units.

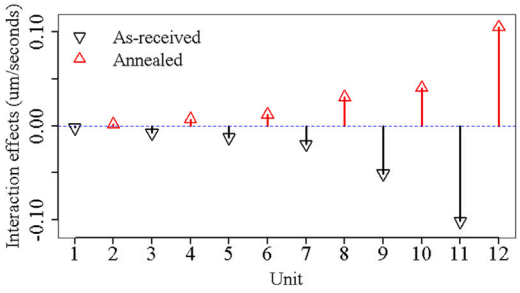

Fig. 10. Interaction effect of external load condition and internal material microstructure among all test units.

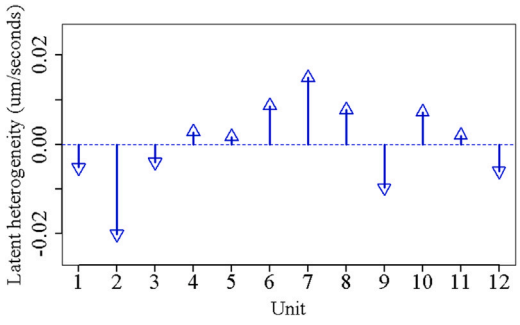

Fig. 11. Quantification of individual latent heterogeneity.

Moreover, the non-zero value of parameter b'_{1111} in Table 3 at a significance level of 0.05 also indicates that there exists an interaction effect between material microstructure (captured by functional covariate) and load condition (captured by scalar covariate) on the slope of degradation path. Specifically, as shown in Eq. (1), such interaction effect between scalar and functional covariates can be characterized as $x_{i1}\int_0^R \rho_{111}(r)Z_{i1}(r)\mathrm{d}r$, which can further be approximated as $Rx_{i1}(b'_{1111}c_{i11} + b'_{1112}c_{i12})$ (as manifested in Section 2.3). Fig. 10 shows the estimated interaction effect, $x_{i1}\int_0^R \rho_{111}(r)Z_{i1}(r)\mathrm{d}r$, among different test units. As shown in Fig. 10, the interaction effect of material microstructure and load condition for as-received test units are negative while the interaction effect for annealed test units are positive. This implies that the performance degradation of a test unit with less homogeneous microstructure is less sensitive to the accelerated load condition. All of the above rich model interpretations will allow the reliability engineers at product design and development stage to better identify the most appropriate material types as well as accelerated conditions to improve the efficiency of accelerated testing as well as to improve the reliability performance of test units.

In addition to the quantification of the influences of the observed factors, the proposed work is able to quantify the latent heterogeneity,

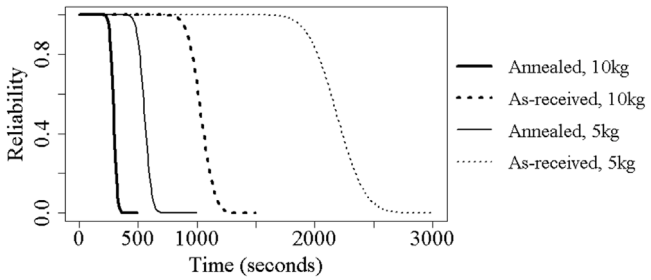

Fig. 12. Reliability curves of four test units with different material types under different load conditions.

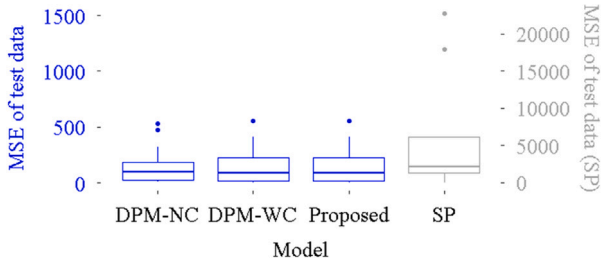

Fig. 13. MSE of test data comparison among different models with latent heterogeneity.

captured by γ_{1i}, within each test unit. As shown in Fig. 11, the positive value of latent heterogeneity indicates a positive effect on the slope of degradation path and vice versus. Such latent heterogeneity is essentially caused by unobserved or unknown factors shared within each test unit (e.g. the potential effect of accumulated wear debris on specimen contact surface). Such information can allow the reliability engineers to target specific test units which exhibit large latent heterogeneity and to investigate the potential contributing factors within each test unit via further data collection and analysis.

Further, we calculate the estimated reliability function based on estimated model parameters, as described in Section 2.3. Assuming a predefined failure threshold of each unit as $H = 100$ micrometer (um) [27], and different load conditions are applied to test units with different material types. As shown in Fig. 12, the reliability curves reflect the reliability characteristics of RULs for test units under different accelerated load conditions with different material types at $t_c = 0$. Among all of the annealed test units, when the external accelerated condition becomes smaller, the test unit has better reliability performance. Among the test units under the same load condition, the annealed test unit tends to accelerate the degradation process and has worse reliability performance. Besides, since the annealed material is more sensitive to the accelerated conditions, the reliability performance differences among the annealed test units under different accelerated load conditions are more significant as compared to the as-received test units. Overall, when the applied load condition becomes smaller and the material type is as-received, the test unit has better reliability performance.

3.5. Performance comparison among degradation models with latent heterogeneity

To further demonstrate the performance of the proposed approach, we compare the proposed model with several existing degradation models with latent heterogeneity. The benchmark models in comparison study include the stochastic process-based model with random parameter (SP-based model) [20], the degradation path model without incorporating any covariates (DPM-NC model) [1] and the degradation

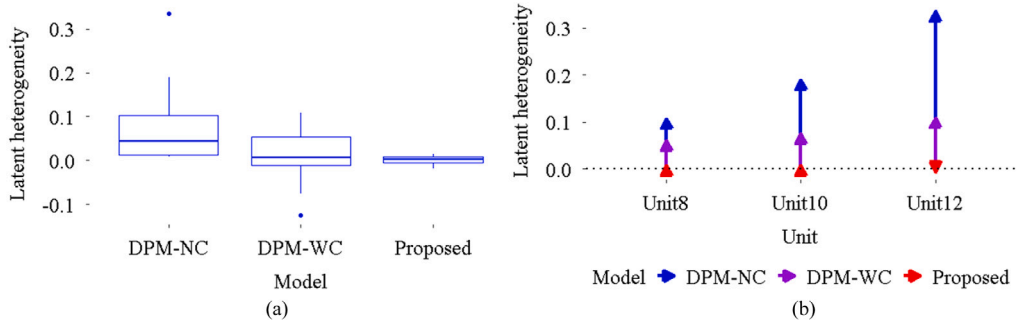

Fig. 14. Individual latent heterogeneity comparison based on different models with latent heterogeneity, (a) among all units, (b) among three annealed units.

path model with incorporating scalar covariates of acceleration conditions (DPM-WC model) [24]. It is noticed that the proposed model belongs to the category of degradation path model by incorporating both scalar covariates of accelerated conditions and functional covariates of material information. Fig. 13 shows the prediction performance (measured in MSE of test data) among all models. The average MSE of test data based on the SP model is much larger than those based on the degradation path models, i.e., DPM-NC model, DPM-WC model and the proposed model. One potential reason is that the SP-based model considers the random functions to capture the large uncertainty of the stochastic deterioration process and uses the random parameters to capture latent heterogeneity. The modeling complexity of the SP-based model is higher than the actual degradation data which embraces limited uncertainty. Thus, the SP-based model results in model overfitting with higher MSE of test data. Among different degradation path models, their prediction performances are comparable, as shown in Fig. 13. The degradation path-based models are all parametric and have a smaller number of model parameters as compared to the SP-based model, which is essentially a nonparametric approach.

Based on the above results, the proposed model demonstrates similar prediction performance as the existing degradation path models with latent heterogeneity. Further, we compare the estimated individual latent heterogeneity among these models and investigate the modeling benefits of the proposed approach. The DPM-NC model mainly introduces the latent heterogeneity without incorporating any covariates and thus has the largest estimated latent heterogeneity among the three models, as shown in Fig. 14(a). The DPM-WC model further incorporates load condition to explain part of individual heterogeneity and thus leads to smaller estimated latent heterogeneity as compared to DPM-NC model. The proposed model achieves the largest reduction of estimated individual latent heterogeneity with the incorporation of mixed-type covariates and their potential interaction, as shown in Fig. 14(a). A large portion of individual heterogeneity is essentially explained by the observed covariates as individual observed heterogeneity and thus the individual latent heterogeneity is reduced considerably. For annealed test units, the external accelerated conditions and the interaction effect have significant contribution to explain the individual observed heterogeneity.

Further, when the load conditions become larger, there is more significantly reduced amount of individual latent heterogeneity of annealed units. To illustrate this finding, we show the estimated individual latent heterogeneity of three annealed test units as an example in Fig. 14(b). The load conditions applied to these annealed units become significantly large (e.g. ≥ 2 kg) and increase from the selected unit 8 to unit 12. We then compare the estimated quantities of individual latent heterogeneity among DPM-NC model (blue lines), DPM-WC model (purple lines) and proposed model (red lines). With incorporating the external accelerated conditions, the DPM-WC model yields more reduction of individual latent heterogeneity as compared to DPM-NC model when the load conditions become larger. With further incorporating the interaction effect between load condition and material characteristics, the proposed model achieves larger reduction of latent heterogeneity as compared to DPM-WC model when the applied load conditions increase and the synergistic effect becomes more significant. Overall, the proposed model can reduce the latent heterogeneity via incorporating interpretable mixed-type covariates and their interaction while it still maintains the satisfactory prediction performance as compared to the existing degradation path models.

4. Conclusion

In this paper, we propose a degradation modeling approach with both generic model formulation and effective estimation algorithm to characterize heterogeneous degradation data with covariates. The proposed model incorporates both mixed-type (i.e., scalar and functional) covariates and latent heterogeneity to improve the prediction accuracy. The scalar and functional covariates represent the external accelerated conditions and the internal material microstructure characteristics of test units, respectively. The developed estimation algorithm further allows the joint quantification of influences of mixed-type covariates as well as latent heterogeneity. In particular, the basis function approximation technique is employed to address the infinite dimensionality issue of model estimation when functional covariate is involved. Data augmentation technique is further employed to simultaneously estimate both the influences of observed mixed-type covariates as well as latent heterogeneity captured by the unit-specific latent variable. Based on a real case study of analyzing the degradation data in an accelerated wear test, the proposed work demonstrates its satisfactory prediction performance over several models based on simplified model specifications, such as models which neglect functional covariates and/or latent heterogeneity. The proposed work also demonstrates its appealing model interpretability as compared to several existing degradation models with latent heterogeneity. The proposed work allows the reliability engineers to better identify the most influencing factors for improving reliability performance of product units and to design more efficient accelerated tests in response to the varied material characteristics of product units.

Several assumptions of the proposed work and the future directions are summarized as follows. In this paper, we utilize baseline material microstructure information of test units before the reliability testing is performed. Since the microstructure of test units may change over time during the reliability testing, it will be important to incorporate in-situ monitoring data of material microstructure in the future to further improve the modeling accuracy. In addition, we focus on modeling the accelerated degradation testing data with covariates in this paper. The external accelerated conditions are well controlled in the laboratory setting. Thus, there is no measurement error or randomness involved in the covariates. In the future, we will extend the proposed modeling framework for analyzing the degradation data with noisy or stochastic covariates, such as uncertain environment conditions in field operation stage.

CRediT authorship contribution statement

Xuxue Sun: Methodology, Software, Writing – original draft, Visualization, Formal analysis, Validation, Writing – review & editing. **Wenjun Cai:** Investigation, Resources, Data curation. **Mingyang Li:** Conceptualization, Writing – review & editing, Supervision.

Acknowledgments

This work was supported in part by National Science Foundation under Grant No. 1825761 and Grant No. 1856196.

Appendix A. Derivations of $\boldsymbol{\Lambda}_i$, $\boldsymbol{A}_{2i}$, $\boldsymbol{A}_{3i}$ and $\boldsymbol{A}_{4i}$

Based on Eq. (3), the degradation measurements can be fully expressed as

$y_{ij} = \sum_{l=0}^{L}[\nu_l + \boldsymbol{\beta}_l^{\mathrm{T}}\boldsymbol{x}_i + R(\sum_{s=1}^{S}\sum_{k=1}^{K} b_{lsk}c_{isk}) + R(\sum_{p=1}^{P} x_{ip}(\sum_{s=1}^{S}\sum_{k=1}^{K} b'_{lpsk}c_{isk})) + \gamma_{li}]\phi_l(t_{ij}) + \epsilon_{ij}, i = 1,\dots,N, j = 1,\dots,m_i$

where y_{ij} is degradation measurement of test unit i measured at time t_{ij} and ϵ_{ij} is the error term. We denote the vector form as $\boldsymbol{y}_i = [y_{i1},\dots,y_{im_i}]^{\mathrm{T}}$ and $\boldsymbol{\epsilon}_i = [\epsilon_{i1},\dots,\epsilon_{im_i}]^{\mathrm{T}}$. ν_l is the population-level average performance degradation at lth decomposition, $\forall l = 0,\dots,L$. Let $\boldsymbol{\nu} = [\nu_0 \dots \nu_L]^{\mathrm{T}}$ be a vector of population-level average performance degradation at all decomposition levels. x_{ip} is the pth scalar covariate of unit i and $\boldsymbol{x}_i = [x_{i1},\dots,x_{iP}]$ is a vector of all P observed scalar covariates. $\boldsymbol{\beta}_l = [\beta_{l1} \dots \beta_{lP}]^{\mathrm{T}}$ is a vector of the coefficients of total P scalar covariates at lth decomposition, $\forall l = 0,\dots,L$. R is the range of spatial distance and c_{isk} is the kth basis coefficient for sth functional covariate of unit i. b_{lsk} and b'_{lpsk} are kth basis coefficients for coefficient function of sth functional covariate and the interaction terms, respectively. We denote $\boldsymbol{b}_{ls} = [b_{ls1},\dots,b_{lsK}]^{\mathrm{T}}$ and $\boldsymbol{b}'_{lps} = [b'_{lps1},\dots,b'_{lpsK}]^{\mathrm{T}}, \forall l = 0,\dots,L, p = 1,\dots,P, s = 1,\dots,S$. γ_{li} is the latent factor of unit i at lth decomposition and let $\boldsymbol{\gamma}_i = [\gamma_{0i},\dots,\gamma_{Li}]^{\mathrm{T}}$ be a vector of latent variables of unit i at all decomposition levels. $\phi_l(\cdot)$ is basis function at lth decomposition. The design matrix of latent heterogeneity can be expressed as

$$\boldsymbol{\Lambda}_i = \left(w_{uv}^{1i}\right) \in \mathbb{R}^{m_i\times(L+1)}$$

where $w_{uv}^{1i} = \phi_v(t_{iu}), \forall u = 1,\dots,m_i, v = 0,\dots,L$ is the element at uth row and vth column. Further, we let $\boldsymbol{A}_{2i}$, $\boldsymbol{A}_{3i}$ and $\boldsymbol{A}_{4i}$ be the block matrices, i.e.,

$$\boldsymbol{A}_{2i} = \begin{pmatrix}\boldsymbol{A}_{20i} & \boldsymbol{A}_{21i} & \dots & \boldsymbol{A}_{2Li}\end{pmatrix} \in \mathbb{R}^{m_i\times(L+1)P}$$

where $\boldsymbol{A}_{2li} = \left(w_{uv}^{2li}\right) \in \mathbb{R}^{m_i\times P}, \forall l = 0,\dots,L$ is the submatrix and the element at uth row and vth column is expressed as $w_{uv}^{2li} = x_{iv}\phi_l(t_{iu}), \forall u = 1,\dots,m_i, v = 1,\dots,P$

$$\boldsymbol{A}_{3i} = \begin{pmatrix}\boldsymbol{A}_{30i} & \boldsymbol{A}_{31i} & \dots & \boldsymbol{A}_{3Li}\end{pmatrix} \in \mathbb{R}^{m_i\times(L+1)SK}$$

where each submatrix $\boldsymbol{A}_{3li} \in \mathbb{R}^{m_i\times SK}$ can be expressed as $\boldsymbol{A}_{3li} = \begin{pmatrix}\boldsymbol{A}_{3li1} & \dots & \boldsymbol{A}_{3liS}\end{pmatrix}$, $\forall l = 0,\dots,L$. Each submatrix $\boldsymbol{A}_{3lis}$ can further be written as $\boldsymbol{A}_{3lis} = \left(w_{uv}^{3lis}\right) \in \mathbb{R}^{m_i\times K}, \forall s = 1,\dots,S$ where $w_{uv}^{3lis} = Rc_{isv}\phi_l(t_{iu}), \forall u = 1,\dots,m_i, v = 1,\dots,K$ is the element at uth row and vth column of the submatrix.

$$\boldsymbol{A}_{4i} = \begin{pmatrix}\boldsymbol{A}_{40i} & \boldsymbol{A}_{41i} & \dots & \boldsymbol{A}_{4Li}\end{pmatrix} \in \mathbb{R}^{m_i\times(L+1)PSK}$$

where $\boldsymbol{A}_{4li} = \begin{pmatrix}\boldsymbol{A}_{4li1} & \dots & \boldsymbol{A}_{4liP}\end{pmatrix} \in \mathbb{R}^{m_i\times PSK}, \forall l = 0,\dots,L$ is the submatrix. Each submatrix $\boldsymbol{A}_{4lip}, \forall p = 1,\dots,P$ can be further expressed as $\boldsymbol{A}_{4lip} = \begin{pmatrix}\boldsymbol{A}_{4lip1} & \dots & \boldsymbol{A}_{4lipS}\end{pmatrix} \in \mathbb{R}^{m_i\times SK}$. Finally, the submatrix $\boldsymbol{A}_{4lips}, \forall s = 1,\dots,S$ can be written as $\boldsymbol{A}_{4lips} = \left(w_{uv}^{4lips}\right) \in \mathbb{R}^{m_i\times K}$ where $w_{uv}^{4lips} = Rx_{ip}c_{isv}\phi_l(t_{iu}), \forall u = 1,\dots,m_i, v = 1,\dots,K$ is the element at uth row and vth column.

With the above block matrices, the design matrix of observed heterogeneity can then be manifested as

$$\boldsymbol{\Omega}_i = \begin{pmatrix}\boldsymbol{\Lambda}_i & \boldsymbol{A}_{2i} & \boldsymbol{A}_{3i} & \boldsymbol{A}_{4i}\end{pmatrix} \in \mathbb{R}^{m_i\times U} \text{ where}$$

$$U = (L+1)(1+P+SK+PSK)$$

Further, we denote $\boldsymbol{\zeta} = \left[\boldsymbol{\nu}^{\mathrm{T}}, \boldsymbol{\beta}_1^{\mathrm{T}},\dots,\boldsymbol{\beta}_L^{\mathrm{T}}, \boldsymbol{b}_{01}^{\mathrm{T}}\dots\boldsymbol{b}_{LS}^{\mathrm{T}}, \boldsymbol{b}'^{\mathrm{T}}_{011}\dots\boldsymbol{b}'^{\mathrm{T}}_{LPS}\right]^{\mathrm{T}}$ as a vector of coefficients for the design matrix of observed heterogeneity. With the above notations, we can simplify the proposed model into the vector form, i.e., $\boldsymbol{y}_i = \boldsymbol{\Omega}_i\boldsymbol{\zeta} + \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_i + \boldsymbol{\epsilon}_i,\ i = 1,\dots,N$.

Appendix B. Derivation of Eq. (8)

As shown in the proposed model Eq. (1), we have $\epsilon_{ij} \sim N(0,\sigma_\epsilon^2)$. The marginal density can then be expressed as $\boldsymbol{y}_i \sim N(\boldsymbol{\Omega}_i\boldsymbol{\zeta}, \boldsymbol{\Lambda}_i\boldsymbol{\Sigma}_\gamma\boldsymbol{\Lambda}_i^{\mathrm{T}} + \sigma_\epsilon^2\mathbf{I}_{m_i})$. Given that $\boldsymbol{\gamma}_i$ is known, the conditional distribution becomes $\boldsymbol{y}_i \mid \boldsymbol{\gamma}_i \sim N(\boldsymbol{\Omega}_i\boldsymbol{\zeta} + \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_i, \sigma_\epsilon^2\mathbf{I}_{m_i})$. The latent variables $\boldsymbol{\gamma}_i$ is assumed to be normally distributed, i.e., $\boldsymbol{\gamma}_i \sim N(0,\boldsymbol{\Sigma}_\gamma)$. Based on Bayes' rule, the conditional density $p(\boldsymbol{\gamma}_i \mid \boldsymbol{y}_i,\boldsymbol{\Theta})$ can be calculated as

$$p(\boldsymbol{\gamma}_i \mid \boldsymbol{y}_i,\boldsymbol{\Theta}) = \frac{p(\boldsymbol{y}_i \mid \boldsymbol{\gamma}_i,\boldsymbol{\Theta})p(\boldsymbol{\gamma}_i \mid \boldsymbol{\Theta})}{p(\boldsymbol{y}_i \mid \boldsymbol{\Theta})} = \frac{C_{11}\frac{1}{\sigma_\epsilon}\left|\Sigma_\gamma\right|^{-\frac{1}{2}}\exp\left[-\frac{1}{2\sigma_\epsilon^2}\left\|\boldsymbol{y}_i - \boldsymbol{\Omega}_i\boldsymbol{\zeta} - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_i\right\|^2 - \frac{1}{2}\boldsymbol{\gamma}_i^{\mathrm{T}}\boldsymbol{\Sigma}_\gamma^{-1}\boldsymbol{\gamma}_i\right]}{C_{12}\left|\boldsymbol{\Lambda}_i\boldsymbol{\Sigma}_\gamma\boldsymbol{\Lambda}_i^{\mathrm{T}} + \sigma_\epsilon^2\mathbf{I}_{m_i}\right|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}(\boldsymbol{y}_i - \boldsymbol{\Omega}_i\boldsymbol{\zeta})^{\mathrm{T}}(\boldsymbol{\Lambda}_i\boldsymbol{\Sigma}_\gamma\boldsymbol{\Lambda}_i^{\mathrm{T}} + \sigma_\epsilon^2\mathbf{I}_{m_i})^{-1}(\boldsymbol{y}_i - \boldsymbol{\Omega}_i\boldsymbol{\zeta})\right]}$$

where C_{11} and C_{12} are normalizing constants. With the expansion of Euclidean norm, i.e., $\left\|\boldsymbol{y}_i - \boldsymbol{\Omega}_i\boldsymbol{\zeta} - \boldsymbol{\Lambda}_i\boldsymbol{\gamma}_i\right\|^2 = \left\|\boldsymbol{y}_i - \boldsymbol{\Omega}_i\boldsymbol{\zeta}\right\|^2 + \left\|\boldsymbol{\Lambda}_i\boldsymbol{\gamma}_i\right\|^2 - 2(\boldsymbol{y}_i - \boldsymbol{\Omega}_i\boldsymbol{\zeta})^{\mathrm{T}}\boldsymbol{\Lambda}_i\boldsymbol{\gamma}_i$, the conditional density can be further represented as

$$\begin{aligned}
p(\boldsymbol{\gamma}_i \mid \boldsymbol{y}_i,\boldsymbol{\Theta}) &\propto \frac{\left|\sigma_\epsilon^2\boldsymbol{\Sigma}_\gamma\right|^{-\frac{1}{2}}}{\left|\boldsymbol{\Lambda}_i\boldsymbol{\Sigma}_\gamma\boldsymbol{\Lambda}_i^{\mathrm{T}} + \sigma_\epsilon^2\mathbf{I}_{m_i}\right|^{-\frac{1}{2}}}\exp[-\frac{1}{2\sigma_\epsilon^2}((\boldsymbol{y}_i - \boldsymbol{\Omega}_i\boldsymbol{\zeta})^{\mathrm{T}}(\boldsymbol{y}_i - \boldsymbol{\Omega}_i\boldsymbol{\zeta}) + \boldsymbol{\gamma}_i^{\mathrm{T}}\boldsymbol{\Lambda}_i^{\mathrm{T}}\boldsymbol{\Lambda}_i\boldsymbol{\gamma}_i \\
&- 2(\boldsymbol{y}_i - \boldsymbol{\Omega}_i\boldsymbol{\zeta})^{\mathrm{T}}\boldsymbol{\Lambda}_i\boldsymbol{\gamma}_i) - \frac{1}{2}\boldsymbol{\gamma}_i^{\mathrm{T}}\boldsymbol{\Sigma}_\gamma^{-1}\boldsymbol{\gamma}_i + \frac{1}{2}(\boldsymbol{y}_i - \boldsymbol{\Omega}_i\boldsymbol{\zeta})^{\mathrm{T}}(\boldsymbol{\Lambda}_i\boldsymbol{\Sigma}_\gamma\boldsymbol{\Lambda}_i^{\mathrm{T}} + \sigma_\epsilon^2\mathbf{I}_{m_i})^{-1}(\boldsymbol{y}_i - \boldsymbol{\Omega}_i\boldsymbol{\zeta})] \\
&\propto \left|(\boldsymbol{\Sigma}_\gamma^{-1} + \frac{\boldsymbol{\Lambda}_i^{\mathrm{T}}\boldsymbol{\Lambda}_i}{\sigma_\epsilon^2})^{-1}\right|^{-\frac{1}{2}}\exp[-\frac{1}{2}(\boldsymbol{y}_i - \boldsymbol{\Omega}_i\boldsymbol{\zeta})^{\mathrm{T}}(\boldsymbol{\Lambda}_i\boldsymbol{\Sigma}_\gamma\boldsymbol{\Lambda}_i^{\mathrm{T}} + \sigma_\epsilon^2\mathbf{I}_{m_i})^{-1}(\frac{\boldsymbol{\Lambda}_i\boldsymbol{\Sigma}_\gamma\boldsymbol{\Lambda}_i^{\mathrm{T}}}{\sigma_\epsilon^2} + \mathbf{I}_{m_i}) \\
&\cdot(\boldsymbol{y}_i - \boldsymbol{\Omega}_i\boldsymbol{\zeta}) + \frac{1}{2}(\boldsymbol{y}_i - \boldsymbol{\Omega}_i\boldsymbol{\zeta})^{\mathrm{T}}(\boldsymbol{\Lambda}_i\boldsymbol{\Sigma}_\gamma\boldsymbol{\Lambda}_i^{\mathrm{T}} + \sigma_\epsilon^2\mathbf{I}_{m_i})^{-1}(\boldsymbol{y}_i - \boldsymbol{\Omega}_i\boldsymbol{\zeta}) \\
&- \frac{1}{2}\boldsymbol{\gamma}_i^{\mathrm{T}}\frac{\boldsymbol{\Lambda}_i^{\mathrm{T}}\boldsymbol{\Lambda}_i}{\sigma_\epsilon^2}\boldsymbol{\gamma}_i \\
&+ \frac{1}{\sigma_\epsilon^2}(\boldsymbol{y}_i - \boldsymbol{\Omega}_i\boldsymbol{\zeta})^{\mathrm{T}}\boldsymbol{\Lambda}_i\boldsymbol{\gamma}_i - \frac{1}{2}\boldsymbol{\gamma}_i^{\mathrm{T}}\boldsymbol{\Sigma}_\gamma^{-1}\boldsymbol{\gamma}_i] \\
&\propto \left|(\boldsymbol{\Sigma}_\gamma^{-1} + \frac{\boldsymbol{\Lambda}_i^{\mathrm{T}}\boldsymbol{\Lambda}_i}{\sigma_\epsilon^2})^{-1}\right|^{-\frac{1}{2}}\exp[-\frac{1}{2}(\boldsymbol{y}_i - \boldsymbol{\Omega}_i\boldsymbol{\zeta})^{\mathrm{T}}(\boldsymbol{\Lambda}_i\boldsymbol{\Sigma}_\gamma\boldsymbol{\Lambda}_i^{\mathrm{T}} + \sigma_\epsilon^2\mathbf{I}_{m_i})^{-1}\frac{\boldsymbol{\Lambda}_i\boldsymbol{\Sigma}_\gamma\boldsymbol{\Lambda}_i^{\mathrm{T}}}{\sigma_\epsilon^2} \\
&\times(\boldsymbol{y}_i - \boldsymbol{\Omega}_i\boldsymbol{\zeta}) \\
&- \frac{1}{2}\boldsymbol{\gamma}_i^{\mathrm{T}}(\boldsymbol{\Sigma}_\gamma^{-1} + \frac{\boldsymbol{\Lambda}_i^{\mathrm{T}}\boldsymbol{\Lambda}_i}{\sigma_\epsilon^2})\boldsymbol{\gamma}_i + \frac{(\boldsymbol{y}_i - \boldsymbol{\Omega}_i\boldsymbol{\zeta})^{\mathrm{T}}\boldsymbol{\Lambda}_i}{\sigma_\epsilon^2}\boldsymbol{\gamma}_i] \\
&\propto \left|\boldsymbol{V}_i\right|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}(\boldsymbol{\gamma}_i - \boldsymbol{\mu}_i)^{\mathrm{T}}\boldsymbol{V}_i^{-1}(\boldsymbol{\gamma}_i - \boldsymbol{\mu}_i)\right]
\end{aligned}$$

where $\boldsymbol{\mu}_i = \frac{1}{\sigma_\epsilon^2}\boldsymbol{V}_i\boldsymbol{\Lambda}_i^{\mathrm{T}}(\boldsymbol{y}_i - \boldsymbol{\Omega}_i\boldsymbol{\zeta})$ and $\boldsymbol{V}_i = (\boldsymbol{\Sigma}_\gamma^{-1} + \frac{\boldsymbol{\Lambda}_i^{\mathrm{T}}\boldsymbol{\Lambda}_i}{\sigma_\epsilon^2})^{-1}$. Thus, $\boldsymbol{\gamma}_i \mid \boldsymbol{y}_i,\boldsymbol{\Theta}$ has gaussian density with mean $\boldsymbol{\mu}_i$ and variance $\boldsymbol{V}_i$. In the E-step of EM estimation framework, we want to compute the expectation of $\boldsymbol{\gamma}_i$ given the observed data and current parameter estimates. With the above derived conditional density, the conditional expectation quantities can then be obtained as

$$\mathbb{E}[\boldsymbol{\gamma}_i \mid \boldsymbol{D},\boldsymbol{\Theta}] = \boldsymbol{\mu}_i = \frac{1}{\sigma_\epsilon^2}\boldsymbol{V}_i\boldsymbol{\Lambda}_i^{\mathrm{T}}(\boldsymbol{y}_i - \boldsymbol{\Omega}_i\boldsymbol{\zeta})$$

$$\begin{aligned}
\mathbb{E}[\boldsymbol{\gamma}_i\boldsymbol{\gamma}_i^{\mathrm{T}} \mid \boldsymbol{D},\boldsymbol{\Theta}] &= \mathbb{V}(\boldsymbol{\gamma}_i \mid \boldsymbol{D},\boldsymbol{\Theta}) + \mathbb{E}[\boldsymbol{\gamma}_i \mid \boldsymbol{D},\boldsymbol{\Theta}]\left(\mathbb{E}[\boldsymbol{\gamma}_i \mid \boldsymbol{D},\boldsymbol{\Theta}]\right)^{\mathrm{T}} \\
&= \boldsymbol{V}_i + \boldsymbol{\mu}_i\boldsymbol{\mu}_i^{\mathrm{T}}, \forall i = 1,\dots,N
\end{aligned}$$

Further, the conditional expectation for all test units can be simplified as

$$\mathbb{E}[\boldsymbol{\gamma} \mid \boldsymbol{D},\boldsymbol{\Theta}] = \boldsymbol{\mu} = \begin{pmatrix}\mathbb{E}[\boldsymbol{\gamma}_1 \mid \boldsymbol{D},\boldsymbol{\Theta}] \\ \mathbb{E}[\boldsymbol{\gamma}_2 \mid \boldsymbol{D},\boldsymbol{\Theta}] \\ \dots \\ \mathbb{E}[\boldsymbol{\gamma}_N \mid \boldsymbol{D},\boldsymbol{\Theta}]\end{pmatrix} = \left(\boldsymbol{\mu}_1^{\mathrm{T}}\boldsymbol{\mu}_2^{\mathrm{T}}\dots\boldsymbol{\mu}_N^{\mathrm{T}}\right)^{\mathrm{T}}$$

Appendix C. Derivations of Eqs. (9), (10) and (11)

In the M step of EM estimation framework, the parameter estimates $\hat{\boldsymbol{\Theta}}$ can be obtained by maximizing Q function $Q(\boldsymbol{\Theta}, \boldsymbol{\Theta}^{(\tau-1)})$, as demonstrated in Eq. (7). Specifically, the parameter estimate $\hat{\boldsymbol{\zeta}}$ at iteration τ can be obtained by $\hat{\boldsymbol{\zeta}}^{(\tau)} = \arg\max_{\boldsymbol{\zeta}} Q(\boldsymbol{\Theta}, \boldsymbol{\Theta}^{(\tau-1)})$. This can be further explicitly derived as

$$\frac{\partial Q}{\partial \boldsymbol{\zeta}} = -\frac{1}{2\sigma_\epsilon^2}\left(\frac{\partial \boldsymbol{\zeta}^{\mathrm{T}}\boldsymbol{\Omega}^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\zeta}}{\partial \boldsymbol{\zeta}} - \frac{\partial \boldsymbol{\zeta}^{\mathrm{T}}\boldsymbol{\Omega}^{\mathrm{T}}\boldsymbol{y}}{\partial \boldsymbol{\zeta}} - \frac{\partial \boldsymbol{y}^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\zeta}}{\partial \boldsymbol{\zeta}} + \frac{\partial \mathbb{E}[\boldsymbol{\gamma}^{\mathrm{T}} \mid \boldsymbol{D}, \boldsymbol{\Theta}^{(\tau-1)}]\boldsymbol{\Lambda}^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\zeta}}{\partial \boldsymbol{\zeta}} + \frac{\partial \boldsymbol{\zeta}^{\mathrm{T}}\boldsymbol{\Omega}^{\mathrm{T}}\boldsymbol{\Lambda}\mathbb{E}[\boldsymbol{\gamma}^{\mathrm{T}} \mid \boldsymbol{D}, \boldsymbol{\Theta}^{(\tau-1)}]}{\partial \boldsymbol{\zeta}}\right) = 0$$

The above equation can be simplified as

$$2\boldsymbol{\Omega}^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{\zeta} + 2\boldsymbol{\Omega}^{\mathrm{T}}\boldsymbol{\Lambda}\mathbb{E}[\boldsymbol{\gamma} \mid \boldsymbol{D}, \boldsymbol{\Theta}^{(\tau-1)}] - 2\boldsymbol{\Omega}^{\mathrm{T}}\boldsymbol{y} = 0$$

By solving the above equation, the closed-form of parameter estimate for $\hat{\boldsymbol{\zeta}}$ can be obtained as

$$\hat{\boldsymbol{\zeta}}^{(\tau)} = (\boldsymbol{\Omega}^{\mathrm{T}}\boldsymbol{\Omega})^{-1}\boldsymbol{\Omega}^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{\Lambda}\mathbb{E}[\boldsymbol{\gamma} \mid \boldsymbol{D}, \boldsymbol{\Theta}^{(\tau-1)}]) = (\boldsymbol{\Omega}^{\mathrm{T}}\boldsymbol{\Omega})^{-1}\boldsymbol{\Omega}^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{\Lambda}\boldsymbol{\mu}^{(\tau-1)})$$

Similarly, the parameter $\hat{\boldsymbol{\Sigma}}_\gamma$ at iteration τ can be obtained by maximizing Q function, i.e., $\hat{\boldsymbol{\Sigma}}_\gamma^{(\tau)} = \arg\max_{\boldsymbol{\Sigma}_\gamma} Q(\boldsymbol{\Theta}, \boldsymbol{\Theta}^{(\tau-1)})$. This can be achieved by solving the following equation

$$\frac{\partial Q}{\partial \boldsymbol{\Sigma}_\gamma} = -\frac{N}{2}\frac{\partial \log\left|\boldsymbol{\Sigma}_\gamma\right|}{\partial \boldsymbol{\Sigma}_\gamma} - \frac{1}{2}\sum_{i=1}^{N}\frac{\partial \mathbb{E}[\boldsymbol{\gamma}_i{}^{\mathrm{T}}\boldsymbol{\Sigma}_\gamma^{-1}\boldsymbol{\gamma}_i \mid \boldsymbol{D}, \boldsymbol{\Theta}^{(\tau-1)}]}{\partial \boldsymbol{\Sigma}_\gamma} = 0$$

By solving the above equation, we can obtain the closed-form of parameter estimate for $\hat{\boldsymbol{\Sigma}}_\gamma$ as

$$\begin{aligned}\hat{\boldsymbol{\Sigma}}_\gamma^{(\tau)} &= \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[\boldsymbol{\gamma}_i\boldsymbol{\gamma}_i^{\mathrm{T}} \mid \boldsymbol{D}, \boldsymbol{\Theta}^{(\tau-1)}] = \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{V}_i^{(\tau-1)} + \boldsymbol{\mu}_i^{(\tau-1)}\boldsymbol{\mu}_i^{(\tau-1)\mathrm{T}}) \\ &= \frac{1}{N}\sum_{i=1}^{N}((\hat{\boldsymbol{\Sigma}}_\gamma^{(\tau-1)-1} + \frac{1}{\hat{\sigma}_\epsilon^{2(\tau-1)}}\boldsymbol{\Lambda}_i^{\mathrm{T}}\boldsymbol{\Lambda}_i)^{-1} + \boldsymbol{\mu}_i^{(\tau-1)}\boldsymbol{\mu}_i^{(\tau-1)\mathrm{T}})\end{aligned}$$

Based on the updated estimates of $\hat{\boldsymbol{\zeta}}^{(\tau)}$ and $\hat{\boldsymbol{\Sigma}}_\gamma^{(\tau)}$ from above equations, the parameter estimate $\hat{\sigma}_\epsilon^2$ at iteration τ can then be obtained by $\hat{\sigma}_\epsilon^{2(\tau)} = \arg\max_{\sigma_\epsilon^2} Q(\boldsymbol{\Theta}, \boldsymbol{\Theta}^{(\tau-1)})$. Further, this can be achieved by solving the following equation

$$\begin{aligned}\frac{\partial Q}{\partial \sigma_\epsilon^2} = &-\frac{\sum_{i=1}^{N} m_i}{2}\frac{\partial \log(\sigma_\epsilon^2)}{\partial \sigma_\epsilon^2} - \frac{\partial \frac{1}{2\sigma_\epsilon^2}}{\partial \sigma_\epsilon^2}\left(\left\|\boldsymbol{y} - \boldsymbol{\Omega}\hat{\boldsymbol{\zeta}}^{(\tau)}\right\|^2\right. \\ &+ \sum_{i=1}^{N}\mathrm{Tr}(\boldsymbol{\Lambda}_i^{\mathrm{T}}\boldsymbol{\Lambda}_i\mathbb{E}[\boldsymbol{\gamma}_i\boldsymbol{\gamma}_i^{\mathrm{T}} \mid \boldsymbol{D}, \boldsymbol{\Theta}^{(\tau-1)}]) \\ &\left. - 2(\boldsymbol{y} - \boldsymbol{\Omega}\hat{\boldsymbol{\zeta}}^{(\tau)})^{\mathrm{T}}\boldsymbol{\Lambda}\mathbb{E}[\boldsymbol{\gamma} \mid \boldsymbol{D}, \boldsymbol{\Theta}^{(\tau-1)}]\right) = 0\end{aligned}$$

By solving the above equation, the closed-form of parameter estimate for $\hat{\sigma}_\epsilon^2$ becomes

$$\begin{aligned}\hat{\sigma}_\epsilon^{2(\tau)} &= \frac{1}{\sum_{i=1}^{N} m_i}\left(\left\|\boldsymbol{y} - \boldsymbol{\Omega}\hat{\boldsymbol{\zeta}}^{(\tau)}\right\|^2 - 2\sum_{i=1}^{N}(\boldsymbol{y}_i - \boldsymbol{\Omega}_i\hat{\boldsymbol{\zeta}}^{(\tau)})^{\mathrm{T}}\boldsymbol{\Lambda}_i\mathbb{E}[\boldsymbol{\gamma}_i \mid \boldsymbol{D}, \boldsymbol{\Theta}^{(\tau-1)}]\right. \\ &\left. + \sum_{i=1}^{N}\mathrm{Tr}(\boldsymbol{\Lambda}_i^{\mathrm{T}}\boldsymbol{\Lambda}_i\mathbb{E}[\boldsymbol{\gamma}_i\boldsymbol{\gamma}_i^{\mathrm{T}} \mid \boldsymbol{D}, \boldsymbol{\Theta}^{(\tau-1)}])\right) \\ &= \frac{1}{\sum_{i=1}^{N} m_i}\left(\left\|\boldsymbol{y} - \boldsymbol{\Omega}\hat{\boldsymbol{\zeta}}^{(\tau)}\right\|^2 - 2\sum_{i=1}^{N}(\boldsymbol{y}_i - \boldsymbol{\Omega}_i\hat{\boldsymbol{\zeta}}^{(\tau)})^{\mathrm{T}}\boldsymbol{\Lambda}_i\boldsymbol{\mu}_i^{(\tau-1)}\right. \\ &\left. + \sum_{i=1}^{N}\mathrm{Tr}(\boldsymbol{\Lambda}_i^{\mathrm{T}}\boldsymbol{\Lambda}_i((\hat{\boldsymbol{\Sigma}}_\gamma^{(\tau)-1} + \frac{1}{\hat{\sigma}_\epsilon^{2(\tau-1)}}\boldsymbol{\Lambda}_i^{\mathrm{T}}\boldsymbol{\Lambda}_i)^{-1} + \boldsymbol{\mu}_i^{(\tau-1)}\boldsymbol{\mu}_i^{(\tau-1)\mathrm{T}}))\right)\end{aligned}$$

Based on the above derivations, the expectation step and the maximization step of the estimation procedure will be repeated alternately as the overall algorithm iterates.

References

[1] Meeker WQ, Escobar LA. Statistical methods for reliability data. John Wiley & Sons; 2014.
[2] Gorjian N, Ma L, Mittinty M, Yarlagadda P, Sun Y. A review on degradation models in reliability analysis. In: Engineering asset lifecycle management. Springer; 2010, p. 369–84.
[3] Goodhew PJ, Humphreys J. Electron microscopy and analysis. CRC Press; 2000.
[4] Torquato S. Statistical description of microstructures. Annu Rev Mater Res 2002;32(1):77–111.
[5] Ye Z-S, Xie M. Stochastic modelling and analysis of degradation for highly reliable products. Appl Stoch Models Bus Ind 2015;31(1):16–32.
[6] Gebraeel N. Sensory-updated residual life distributions for components with exponential degradation patterns. IEEE Trans Autom Sci Eng 2006;3(4):382–93.
[7] Kharoufeh JP, Solo CJ, Ulukus MY. Semi-Markov models for degradation-based reliability. IIE Trans 2010;42(8):599–612.
[8] Shu M-H, Hsu B-M, Kapur KC. Dynamic performance measures for tools with multi-state wear processes and their applications for tool design and selection. Int J Prod Res 2010;48(16):4725–44.
[9] Zhao X, He S, Xie M. Utilizing experimental degradation data for warranty cost optimization under imperfect repair. Reliab Eng Syst Saf 2018;177:108–19.
[10] Fang G, Pan R, Hong Y. Copula-based reliability analysis of degrading systems with dependent failures. Reliab Eng Syst Saf 2020;193:106618.
[11] Park JS, Lee SM, Joo BS, Jang H. The effect of material properties on the stick–slip behavior of polymers: A case study with PMMA, PC, PTFE, and PVC. Wear 2017;378:11–6.
[12] Si W, Yang Q, Wu X. Material degradation modeling and failure prediction using microstructure images. Technometrics 2019;61(2):246–58.
[13] Zhou S, Xu A. Exponential dispersion process for degradation analysis. IEEE Trans Reliab 2019;68(2):398–409.
[14] Zhai Q, Chen P, Hong L, Shen L. A random-effects Wiener degradation model based on accelerated failure time. Reliab Eng Syst Saf 2018;180:94–103.
[15] Pan D, Lu S, Liu Y, Yang W, Liu J-B. Degradation data analysis using a wiener degradation model with three-source uncertainties. IEEE Access 2019;7:37896–907.
[16] Wen Y, Wu J, Das D, Tseng T-LB. Degradation modeling and RUL prediction using Wiener process subject to multiple change points and unit heterogeneity. Reliab Eng Syst Saf 2018;176:113–24.
[17] Xiao M, Zhang Y, Li Y, Wang W. Degradation modeling based on Wiener process considering multi-source heterogeneity. IEEE Access 2020;8:160982–94.
[18] Hao S, Yang J, Berenguer C. Degradation analysis based on an extended inverse Gaussian process model with skew-normal random effects and measurement errors. Reliab Eng Syst Saf 2019;189:261–70.
[19] Zhao X, Gaudoin O, Doyen L, Xie M. Optimal inspection and replacement policy based on experimental degradation data with covariates. IISE Trans. 2019;51(3):322–36.
[20] Hermann S, Ickstadt K, Müller CH. Bayesian prediction for a jump diffusion process–with application to crack growth in fatigue experiments. Reliab Eng Syst Saf 2018;179:83–96.
[21] Veloso GA, Loschi RH. Dynamic linear degradation model: Dealing with heterogeneity in degradation paths. Reliab Eng Syst Saf 2021;210:107446.
[22] Yuan T, Bae SJ, Zhu X. A Bayesian approach to degradation-based burn-in optimization for display products exhibiting two-phase degradation patterns. Reliab Eng Syst Saf 2016;155:55–63.
[23] Xu Z, Hong Y, Jin R. Nonlinear general path models for degradation data with dynamic covariates. Appl Stoch Models Bus Ind 2016;32(2):153–67.
[24] Pan R, Crispin T. A hierarchical modeling approach to accelerated degradation testing data analysis: A case study. Qual Reliab Eng Int 2011;27(2):229–37.
[25] Schilling RJ, Carroll JJ, Al-Ajlouni AF. Approximation of nonlinear systems with radial basis function neural networks. IEEE Trans Neural Netw 2001;12(1):1–15.
[26] Huang JZ, Wu CO, Zhou L. Polynomial spline estimation and inference for varying coefficient models with longitudinal data. Statist Sinica 2004;763–88.
[27] Singh J, Cai W, Bellon P. Dry sliding of Cu–15 wt% Ni–8 wt% Sn bronze: Wear behaviour and microstructures. Wear 2007;263(1–6):830–41.
[28] Goel S, Luo X, Reuben RL. Molecular dynamics simulation model for the quantitative assessment of tool wear during single point diamond turning of cubic silicon carbide. Comput Mater Sci 2012;51(1):402–8.
[29] Jiao Y, Stillinger F, Torquato S. Modeling heterogeneous materials via two-point correlation functions: Basic principles. Phys Rev E 2007;76(3):031110.
[30] Ramsay JO, Silverman BW. Applied functional data analysis: Methods and case studies. Springer; 2007.
[31] Wood SN. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. J R Stat Soc Ser B Stat Methodol 2011;73(1):3–36.
[32] Hsing T, Eubank R. Theoretical foundations of functional data analysis, with an introduction to linear operators. 997, John Wiley & Sons; 2015.
[33] Chen K, Lei J. Localized functional principal component analysis. J Amer Statist Assoc 2015;110(511):1266–75.
[34] Wei GC, Tanner MA. A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. J Amer Statist Assoc 1990;85(411):699–704.
[35] Wu CJ, et al. On the convergence properties of the EM algorithm. Ann Statist 1983;11(1):95–103.
[36] Efron B. Better bootstrap confidence intervals. J Amer Statist Assoc 1987;82(397):171–85.
[37] Chaudhuri BB, Sarkar N. Texture segmentation using fractal dimension. IEEE Trans Pattern Anal Mach Intell 1995;17(1):72–7.
[38] Shibata R. Selection of the order of an autoregressive model by Akaike's information criterion. Biometrika 1976;63(1):117–26.
[39] Watanabe S. A widely applicable Bayesian information criterion. J Mach Learn Res 2013;14(Mar):867–97.