

A Functional Time Warping Approach to Modeling and Monitoring Truncated Degradation Signals

Rensheng R. Zhou, Nicoleta Serban, Nagi Gebraeel

H. Milton Stewart School of Industrial Systems and Engineering
Georgia Institute of Technology

rzhou8@gatech.edu, nserban@isye.gatech.edu, nagi@isye.gatech.edu

Hans-Georg Müller

Department of Statistics

University of California, Davis

hgmuller@ucdavis.edu

Abstract: *Degradation signals* are sensor-based signals that are correlated with degradation processes of engineering components. In this paper, we present a flexible modeling framework for characterizing degradation signals that can only be observed up to a pre-specified failure threshold. The underlying assumption of this framework is that the engineering components degrade according to a similar trend, referred to as the *common shape function*, but at different degradation rates. Under this assumption, the degradation signals of different components are synchronized using a random time warping process, which transforms the common trend function into degradation processes that progress at varying rates. Our primary objective is to obtain real-time predictions for the residual lifetime of components deployed in the field. In the initial step, the historical degradation signals are used to recover the distribution of the degradation processes under the assumptions of the proposed time warping model. Next, the distribution of the degradation process is updated using the signal(s) of partially degraded component(s). The updated model is then utilized to predict the residual lifetime distributions of these components. We test the performance of our methodology using vibration-based degradation signals from a rotating machinery experiment and simulated degradation signals. Additional information and codes are available as supplementary material online.

Key Words: Functional data analysis; Degradation Modeling; Residual life prediction.

1 INTRODUCTION

In recent years, sensor technologies have been widely used to acquire real-time information about the degradation processes of engineering components. The resulting signals are usually correlated with the degradation processes and they are known as *degradation signals* (Nelson 1990). In degradation data analysis, a component’s *failure* is typically defined to be the time point when its degradation signal first reaches a critical level known as the *failure threshold*. In many engineering applications, a component is shut down or replaced as soon as the level of its degradation signal reaches the failure threshold. Therefore, no further observations are acquired beyond this threshold. We refer to the signals that are observed in these applications as (*amplitude*) *truncated degradation signals*. Truncated degradation signals are common in many engineering applications (e.g., Gebraeel et al., 2005; Gebraeel and Pan, 2008; Lu and Meeker, 1993; Virkler et al., 1979). The crack propagation data in Virkler et al., 1979 and the bearing degradation data studied in Gebraeel et al., 2005 are two examples of such signals (Figure 1).

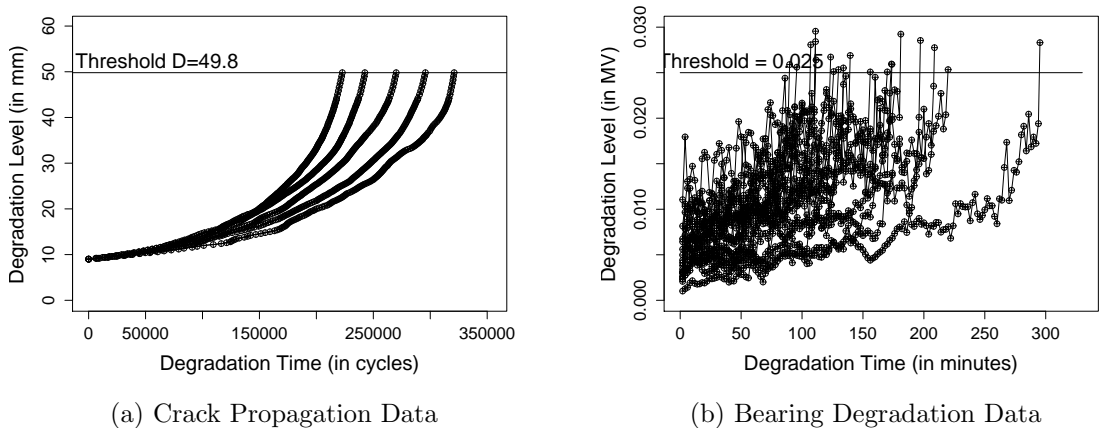


Figure 1: Examples of Truncated Degradation Signals

Most existing degradation models focus on estimating population-related reliability characteristics, such as failure time distributions. In contrast, the overarching objective of this research is to predict the residual life distribution (the distribution of time to reach the failure threshold, and it will be abbreviated as “RLD” in the remaining paper) of a *fielded component*, i.e. a component that is still operating in the field, by using the partial observations of its degradation signal (Medjaher et al., 2012). Online predictions of RLD for fielded

components are often used for high-level decision-making, such as device replacement and maintenance scheduling among others.

Our prediction approach is a two-step procedure. In the first stage, we develop a modeling framework that is suitable for characterizing truncated degradation signals. We train the model using historical degradation signals, referred to as *training* degradation signals hereafter. At this stage, we estimate the distribution of the degradation process, which in turn, allows us to recover an approximate (prior) distribution for the residual lifetime. In the second stage, we predict the RLD of a fielded component by updating the distribution of the degradation process obtained at the first stage using the partial observations of its degradation signal. This is an online procedure because the updating is performed as more degradation observations of the fielded component are collected.

In the existing literature, most degradation models assume that the parametric representation of the degradation process can be derived based on the available physical or chemical theories associated with the degradation process. However, such theories may be unknown in many engineering applications, and therefore, the parametric representation can only be inferred empirically or visually from degradation observations. Parametric modeling without underlying physical or chemical theories is restrictive and sensitive to model mis-specification (Zhou et al., 2011). This is particularly noteworthy in situations when the available data are sparse and/or when the underlying degradation process does not have a well-specified parametric form.

Most existing approaches, which do not constrain the degradation process to a parametric form, however, are not directly applicable for modeling truncated degradation signals since they generally assume that the observed random signals share the same time domain. This assumption does not hold for truncated degradation signals since their support varies from one component to another. To overcome this challenge, the authors of Müller and Zhang (2005) propose a time-varying functional model and apply it to estimation and prediction of the reproductive activity of medflies. To estimate the distribution of the reproductive activity process at one time point, the authors only use the trajectories of medflies that are still alive up to the current time. This is to ensure that the trajectories used to estimate the model share the same time domain. One drawback of this approach is that the training

data set is not fully exploited since the trajectories of medflies that have already died by the current time are discarded. This inefficient use of data will induce decreasing estimation accuracy at later time points when there are fewer living medflies. We demonstrate this in a simulation study presented in Section 6.

In this paper, we propose to model truncated degradation signals using a flexible model that does not assume a specific parametric form for the degradation process and it does not assume that the signals share the same time domain. Our method draws its roots from functional time warping techniques. The primary goal of time warping, also known as curve registration/synchronization, is to align each curve by warping the time domains such that the warped curves maximally coincide with a template or common shape function (Liu and Müller, 2004), or is to optimize pairwise alignments (Tang and Müller, 2008; Tang and Müller, 2009). Various time warping approaches have been proposed to date. Some important examples include landmark registration (e.g., Kneip and Gasser, 1992), dynamic time warping (e.g., Wang and Gasser, 1997; Wang and Gasser, 1999), functional convex synchronization (e.g., Liu and Müller, 2004), nonparametric maximum-likelihood warping (e.g., Gervini and Gasser, 2005) and self-modeling registration (e.g., Gervini and Gasser, 2004; Telesca and Inoue, 2008).

We pursue a self-modeling registration approach because it allows estimation of the common shape function by borrowing information across signals. In our method, the underlying assumption is that components degrade according to a similar trend, referred to as the *common shape function*, but at different degradation rates. The degradation signals of the components are synchronized using a random time warping process, which transforms the common trend function into degradation processes that progress at varying rates. Our methodology is partially inspired by Telesca and Inoue (2008), which proposed a Bayesian hierarchical curve registration (BHCR) model. Although the original BHCR model was devised only for signals that share the same time domain, we relax this assumption and demonstrate that our modified time warping model is well-suited for modeling truncated signals.

The methodology presented in this paper applies to a wide range of engineering applications. For example, gear boxes and generators of wind turbines have maximum permissible

vibration levels beyond which a turbine must be shut down for maintenance. The International Standards Organization (ISO) has devised a series of standard vibration severity levels for various categories of rotating machinery (ISO 7919) and allowable vibration severity levels for non-rotating machinery (ISO 10816). These standards represent acceptable and unacceptable operation levels at which machinery shutdown is necessary. Examples of some of the applications covered under these standards include gas and steam turbines, hydraulic power generating and pumping plants, rotodynamic pumps, among many others. These machinery components often do not operate as standalone components but rather are part of operating systems. For example, pumps are typically the most critical components of cooling systems in nuclear power plants, gear boxes are inherent components of wind turbines, generators can be found on-board aircraft, ships, helicopters and manufacturing machinery.

Vibration standards represent only one sector of industrial applications where failure is based on failure or operation severity thresholds. There are indeed numerous other types of applications that have different standards and that are based on different characteristics. For example, in power networks, transformers represent some of the most critical components of the networks. The Duval Triangle provides acceptable and unacceptable levels of gases concentrations (C_2H_2 , CH_4 , C_2H_4) in the oil surrounding the transformer coils beyond which a transformer must be shutdown. In summary, there are many applications where components are required to be shutdown or removed for service once a performance or degradation attribute crosses some pre-specified threshold.

In the case study of this paper, we utilize vibration monitoring data from a rolling element bearing testing setup which is used to perform accelerated degradation tests on thrust ball bearings. We evaluate the efficiency of our proposed degradation model using these vibration-based degradation signals. We compare our proposed approach with the parametric model in Gebraeel et al. (2005) as a benchmark.

The remainder of the paper is organized as follows. We first describe the general problem and its challenges in Section 2. We present the time warping model applied to prediction of the residual life of fielded components in Section 3. The performance of our model is investigated using bearing degradation signals in Section 5 and simulated degradation signals in Section 6. We conclude this paper in Section 7. Some technical details and additional

simulation studies are deferred to the Supplementary Materials.

2 PROBLEM DESCRIPTION AND CHALLENGES

We denote a component's degradation process by $X(t)$, where t is the chronological time. In our framework, a component's underlying failure time T_u is defined as the time when its degradation process $X(\cdot)$ reaches a failure threshold D . Depending on the applications, the failure threshold may be above or below the starting point. Without loss of generality, we focus on the former case. Under this assumption, T_u is defined as follows:

$$T_u = \inf_{t \geq 0} [X(t) \geq D]. \quad (1)$$

Ideally, $X(\cdot)$ could be observed continuously over time and with no measurement errors. In practice, a degradation process can only be monitored at inspection (or sampling) time points, and is commonly contaminated with measurement errors, denoted by $\epsilon(\cdot)$. We represent the observed degradation signal by $Y(\mathbf{t}) = X(\mathbf{t}) + \epsilon(\mathbf{t})$, where \mathbf{t} is a vector of discrete time points. Without loss of generality, we assume that the degradation signal is collected at $\mathbf{t} = (d, 2d, 3d, \dots)$, where d represents the observation time interval. We note that degradation signals do not need to be observed regularly.

In applications with truncated signals, degradation processes are monitored up to the observed failure time T_o , which is defined as the time when the observed degradation signal $Y(\cdot)$ first reaches the failure threshold D :

$$T_o = \min_{t=md, m \in \mathbb{Z}^+} [Y(t) \geq D]. \quad (2)$$

Therefore, T_u is not observed, but instead its approximation T_o is available.

To illustrate the definitions of T_u and T_o , we provide an example in Figure 2. In this figure, the solid line represents the unobserved degradation process $X(\cdot)$ and the solid points represent the discrete degradation observations of $Y(\mathbf{t})$. The degradation signal is truncated at the failure threshold $D = 12$. According to the definitions in (1-2), $T_u = 3.6$ and $T_o = 4$.

In this paper, our primary objective is to predict the underlying residual life of a fielded component, denoted by RL_u^* ($RL_u^* = T_u^* - t^*$), using the partial observations $Y^*(\mathbf{t})$ of its degradation signal, observed up to a certain time point t^* . To this purpose, we model T_u as

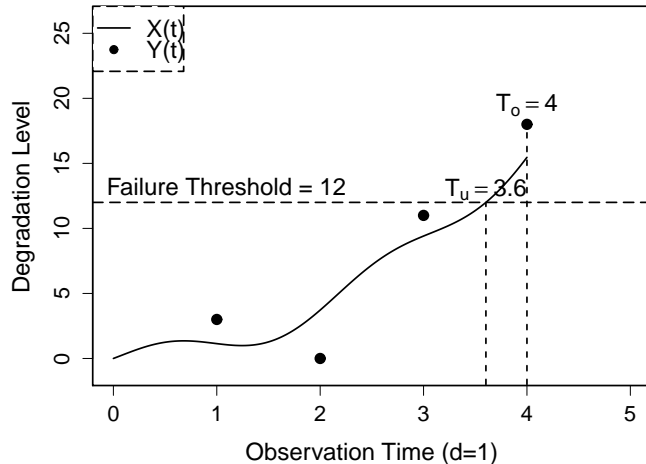


Figure 2: An illustration of the difference between the actual unobserved crossing time T_u and the approximation T_o , the first observation time after crossing the failure threshold.

a function of the component’s observations and a set of (unknown) parameters. Particularly, we specify a statistical model G , such that $T_u = G(Y(\mathbf{t}), \boldsymbol{\beta})$, and then estimate the model parameters $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\beta}}$. This will provide a prediction for RL_u^* as follows:

$$\hat{RL}_u^* = \hat{T}_u^* - t^* = G(Y^*(\mathbf{t}), \hat{\boldsymbol{\beta}}) - t^*$$

Most methods require observing T_u in the training stage. As mentioned above, one can approximate T_u given the observed truncated signal. A direct approximation would be $\tilde{T}_u = T_o$. There are many other approximation methods, for instance, the previous time point before T_o , the interpolation time point at the failure threshold between the observation at “ T_o ” and that of the previous time point, and so on. (Our simulation studies indicate that the model performance is insensitive to the choice among these approximation methods.) The bias due to the approximation is dependent upon both the underlying process and the parameters relevant to the observation process, specifically, the observation time interval d and the amplitude of the measurement error $\epsilon(\cdot)$. Theoretical insights about the approximation error and its influence on the residual life predictions can be provided by the authors upon readers’ requests.

One approach for deriving the statistical model $G(Y(\mathbf{t}), \boldsymbol{\beta})$ is using regression analysis

in which the underlying failure time is the response variable, and the predictors consist of the observed degradation signals. The regression model parameters may be estimated using training degradation signals. Given $Y^*(\mathbf{t})$, the observed signal of a fielded component, one can then predict its underlying failure time by using the trained regression model. For example, in Müller and Zhang (2005), the authors propose a time-varying functional regression method for predicting the remaining life of medflies. In their approach, they apply a functional principal component analysis to the signals up to the current time, and therefore, estimate a time-varying model. The observed functional signals are then represented by the time-varying functional principal component scores, which are continuously updated as time progresses. The time-varying scores become the predictor variables in the functional regression model.

In this paper, we first estimate a model for $X(\cdot)$ and use the definition in (1) along with the model for $X(\cdot)$ to obtain predictions for T_u . This first step provides a prior model distribution for $X(\cdot)$. Similar to our previous research (Gebrael et al., 2005; Zhou et al., 2011; Zhou et al., 2012), we update the prior model distribution of $X(\cdot)$ using the partial observations $Y^*(\mathbf{t})$ of a fielded component, denote this updated model distribution $X^*(\cdot)$, and use it further to predict T_u^* . This paper introduces a novel approach for the model distribution of $X(\cdot)$ given that we observe truncated degradation signals. In Section 3, we present the model decomposition and assumptions for $X(\cdot)$. One advantage of our modeling approach is that it provides a direct prediction formula for T_u^* . We present this in Section 4.

3 MODELING AND MONITORING TRUNCATED SIGNALS

In this section, we present a time warping framework for modeling the underlying degradation process $X(\cdot)$ based on truncated degradation signals. The associated model parameters are estimated using a Bayesian approach. This is the first step in our approach as discussed in the previous section.

3.1 Model Framework and Assumptions

The training data set consists of a collection of degradation signals $Y_i(t_{ij})$, for $i = 1, \dots, I$ (I is the number of components) and $j = 1, \dots, m_i$ (m_i is the number of observation time points for signal i).

The central assumption of the underlying degradation process is $X_i(\cdot) = Z(h_i(\cdot))$, where $Z(\cdot)$ is a shape function that characterizes the common degradation trend across components and $h_i(\cdot)$ is a synchronizing function that characterizes the rate at which the i th component degrades. Under this assumption, the model decomposition is

$$Y_i(t) = X_i(t) + \varepsilon_i(t) = Z(h_i(t)) + \varepsilon_i(t), \quad t \in [0, T_{i,u}] \quad (3)$$

$$Z(s) = \mathbf{B}_m'(s)\boldsymbol{\beta}, \quad s \in [0, 1] \quad (4)$$

$$\mu_i(s) \equiv h_i^{-1}(s) = \mathbf{B}_\mu'(s)\boldsymbol{\phi}_i, \quad s \in [0, 1] \quad (5)$$

More details and assumptions about this model are as follows:

- Throughout this paper, s will represent synchronized time and t will represent physical observation time, which is discrete in practice. $t \in [0, T_{i,u}]$, where $T_{i,u}$ represents the underlying failure time of the i th unit.

- The measurement error term $\varepsilon_i(t)$ is assumed to be independent and normally distributed with zero mean and variance σ_ε^2 . It is also assumed to be independent of all the other random terms in the model.

- The shape function $Z(s)$ is assumed to be unknown, and therefore, decomposed using a set of basis functions $\mathbf{B}_m(s)$. Throughout this paper, we use cubic B-spline basis functions of dimension d_β . We assume $Z(s)$ is smooth and model it using penalized regression splines, where the penalty, which accounts for the differences of adjacent B-spline coefficients, is defined to ensure adequate smoothness of the fitted curves (Eilers and Marx, 1996). Under the Bayesian framework (Lang and Brezger, 2004), this corresponds to specifying a first-order random walk shrinkage prior on the shape coefficients $\boldsymbol{\beta} = (\beta_1, \dots, \beta_{d_\beta})$ such that

$$\beta_k = \beta_{k-1} + e_k, \quad e_k \sim \text{Normal}(0, \lambda_\beta)$$

Under this formulation, $\beta \sim Normal(0, \Sigma_\beta)$ (Lang and Brezger, 2004), with

$$\Sigma_\beta^{-1} = \Omega_\beta / \lambda_\beta; \Omega_\beta = \begin{pmatrix} 2 & -1 & 0 & & & 0 \\ -1 & 2 & -1 & \ddots & & \\ 0 & -1 & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & -1 & 0 \\ & & & -1 & 2 & -1 \\ 0 & & & 0 & -1 & 1 \end{pmatrix}_{d_\beta \times d_\beta}$$

where Ω_β is a precision matrix and λ_β controls the smoothness level of $Z(s)$.

- The time warping function $\mu_i(s)$ is the inverse of the synchronizing function $h_i(t)$. Since we assume $\mu_i(s)$ has an unknown shape, we decompose $\mu_i(s)$ using a set of basis functions $\mathbf{B}_\mu(s)$ of the dimension d_ϕ . The functions $\mu_i(s)$ and $h_i(t)$ are subject to the following two assumptions:

- * (monotonicity) $\mu_i(s)$ and $h_i(t)$ are both strictly monotone increasing, and therefore invertible;

Remark: The monotonicity constraint ensures a one-to-one correspondence between synchronized time s and physical observation time t . In our definition, $\mu_i(s) \equiv h_i^{-1}(s)$, or equivalently, $h_i(t) \equiv \mu_i^{-1}(t)$. To ensure the monotonicity of the time warping functions, we constrain $\phi_i = (\phi_{i,1}, \dots, \phi_{i,d_\phi})$ such that

$$\phi_{i,1} < \phi_{i,2} < \dots < \phi_{i,d_\phi}. \tag{6}$$

Under the above condition for ϕ_i , the derivative of $\mu_i(s)$ is positive according to the basic properties of B-splines. Therefore, $\mu_i(s)$ is a monotone increasing function. Note that this condition is sufficient but not necessary. Details can be found in Brumback and Lindstrom, 2004.

- * (image) $0 \leq h_i(t) \leq 1$, $h_i(0) = 0$, $h_i(T_{i,u}) = 1$.

Remark: The standardized time interval $[0, 1]$ has an intuitive interpretation. Every component starts to degrade from the beginning of its lifetime, represented by the synchronized time $s = 0$, to the end of it, represented by the synchronized

time $s = 1$. Each component corresponds to a specific time warping function since each component degrades at its own rate. To ensure that the image assumption holds, we constrain ϕ_i such that

$$\phi_{i,1} = 0, \quad \phi_{i,d_\phi} = T_{i,u}. \quad (7)$$

- We assume that the time warping functions $\mu_i(s)$ are smooth and model these functions using penalized regression splines by placing a first-order random walk shrinkage prior on the coefficients. Specifically, we assume $\phi_i \sim N(b_0 \mathbf{\Upsilon}, \mathbf{\Sigma}_\phi)$ with constraints specified in (6) and (7). Here $\mathbf{\Upsilon}$ is the vector associated with the identity time-transformation function, i.e., $\mathbf{B}_\mu' \mathbf{\Upsilon} = t$. $\mathbf{\Sigma}_\phi^{-1} = \mathbf{\Omega}_\phi / \lambda_\phi$, where $\mathbf{\Omega}_\phi$ is a precision matrix of the dimension d_ϕ that has the same structure as that of $\mathbf{\Omega}_\beta$ and λ_ϕ is a scalar that controls the smoothness level.

In the simpler case when $\mu_i(s) = b_0 s$, the time warping process involves only stretching or compressing the time domain at a constant rate. In other words, the time warping process retains the original shape of the common shape function. Here b_0 can be interpreted as an average acceleration factor, which describes the overall mean degradation rate.

We provide one example in Figure 3 to illustrate this modeling framework. In this figure, the long dashed line represents a common shape function, which is quadratic with respect to the synchronized time $s \in [0, 1]$, i.e., $Z(s) = s^2$. The dot-dashed lines represent the underlying degradation processes of two different components. For component 1, the time warping function is linear: $\mu_1(s) = 5s$, therefore, $h_1(t) = \frac{t}{5}$. According to our model, this leads to $X_1(t) = Z(h_1(t)) = (\frac{t}{5})^2$ and the corresponding underlying failure time $T_{1,u} = 5$. For component 2, the time warping function is quadratic: $\mu_2(s) = 3s^2$, therefore, $h_2(t) = \sqrt{\frac{t}{3}}$. According to our model, this leads to $X_2(t) = Z(h_2(t)) = \frac{t}{3}$ and the corresponding underlying failure time $T_{2,u} = 3$.

3.2 Model Estimation

To simplify the derivation of our model estimation procedure, we divide the model parameters into $\boldsymbol{\theta} = (\sigma_\varepsilon^2, \boldsymbol{\beta}, b_0, \lambda_\phi)$ and $\boldsymbol{\phi} = (\phi_1, \dots, \phi_I)$. One approach to estimating these parameters is by maximizing the marginal log-likelihood of the pooled degradation data

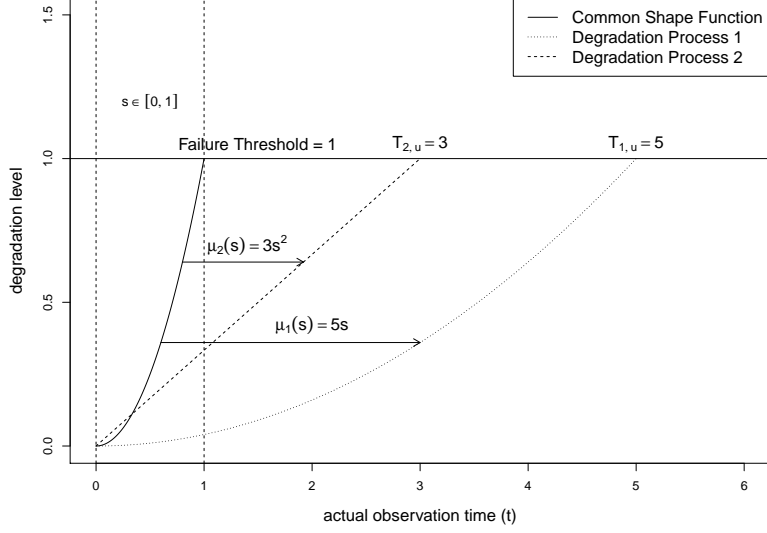


Figure 3: Illustration of two degradation signals following our central time-warping model assumption.

$\mathbf{Y} = (\mathbf{Y}_1, \dots, \mathbf{Y}_I)$ decomposed as:

$$l_Y = \sum_{i=1}^I l_{Y_i} = \sum_{i=1}^I \log \int f(\mathbf{Y}_i | \phi_i; \boldsymbol{\theta}) f(\phi_i | \boldsymbol{\theta}) d\phi_i = \sum_{i=1}^I \log \int f(\mathbf{Y}_i | \phi_i; \boldsymbol{\beta}, \sigma_\varepsilon^2) f(\phi_i | \lambda_\phi, b_0) d\phi_i .$$

Due to the model complexity, maximizing the above marginal log-likelihood is intractable. In this paper, we resort to a Bayesian approach for model estimation. For computational convenience, we choose conjugate priors for all parameters that are needed for estimation (In our notation, if $X \sim \text{Gamma}(k, r)$, then $E(X) = k/r$):

$$\begin{aligned} 1/\sigma_\varepsilon^2 &\sim \text{Gamma}(k_{\sigma_\varepsilon}, \gamma_{\sigma_\varepsilon}) \text{ and } 1/\sigma_{b_0}^2 \sim \text{Gamma}(k_{\sigma_{b_0}}, r_{\sigma_{b_0}}) \\ b_0 &\sim \text{Normal}(m_{b_0}, \sigma_{b_0}^2) \\ 1/\lambda_\beta &\sim \text{Gamma}(k_{\lambda_\beta}, r_{\lambda_\beta}) \text{ and } 1/\lambda_\phi \sim \text{Gamma}(k_{\lambda_\phi}, r_{\lambda_\phi}) \end{aligned}$$

Given the observed data Y , the posterior distribution of $(\boldsymbol{\theta}, \boldsymbol{\phi})$ is:

$$\begin{aligned} f(\boldsymbol{\theta}, \boldsymbol{\phi} | \mathbf{Y}) &\propto f(\mathbf{Y} | \boldsymbol{\theta}, \boldsymbol{\phi}) f(\boldsymbol{\theta}, \boldsymbol{\phi}) \\ &= f(\mathbf{Y} | \boldsymbol{\phi}; \boldsymbol{\beta}, \sigma_\varepsilon^2) f(\boldsymbol{\phi} | b_0; \lambda_\phi) f(\boldsymbol{\beta} | \lambda_\beta) \\ &\times f(\sigma_\varepsilon^2 | k_{\sigma_\varepsilon}, r_{\sigma_\varepsilon}) f(b_0 | m_{b_0}, \sigma_{b_0}^2) f(\sigma_{b_0}^2 | k_{\sigma_{b_0}}, r_{\sigma_{b_0}}) f(\lambda_\beta | k_{\lambda_\beta}, r_{\lambda_\beta}) f(\lambda_\phi | k_{\lambda_\phi}, r_{\lambda_\phi}) \end{aligned}$$

Since the joint posterior density is analytically intractable, we propose to use an MCMC-based posterior inference procedure. More specifically, we implement a Metropolis-within-Gibbs algorithm. In this algorithm, we use Metropolis-Hastings sampling to sample the time-transformation coefficients ϕ_i , $i = 1, \dots, I$ and Gibbs sampling to sample all other parameters from their respective closed form full conditionals. A detailed description of the estimation algorithm and the full conditionals are given in supplementary sections A and B.

The foregoing implementation requires the specification of the tuning parameters d_β and d_ϕ , which reduces to a model selection problem. One could implement cross-validation to select d_β and d_ϕ to minimize a predefined measure for the prediction errors. However, this approach is computationally prohibitive. In this paper, we propose to use the Bayes factor (Kass and Raftery, 1995) for determining d_β and d_ϕ , as it is consistent with our Bayesian estimation framework and computationally more efficient. More details about using the Bayes Factor for model selection are presented in supplementary section C.

4 RESIDUAL LIFE PREDICTION

The focus of this paper is on predicting the RLD of a fielded component using partial observations of its degradation signal. In this section, we discuss how to derive this prediction using the time warping model for the underlying degradation process $X(\cdot)$.

Assume that we have collected a partial degradation signal $Y^*(\mathbf{t})$ for a fielded component, where \mathbf{t} is a vector of discrete time points within the interval $[0; t^*]$. To predict the RLD for this new component, one could simply use the time warping model for $X(\cdot)$ without taking into account the additional observations $Y^*(\mathbf{t})$. However, the RLD prediction would be more accurate if instead we update the distribution of $X(\cdot)$ given the partial degradation signal $Y^*(\mathbf{t})$, denote it $X^*(\cdot)$, and use (1) to predict T_u^* .

An appealing advantage of our proposed time warping approach is that we have a very simple formula for predicting the residual life of a fielded component due to the image assumption:

$$RL_u^* = T_u^* - t^* = \mu^*(1) - t^* = \mathbf{B}_\mu'(1)\phi^* - t^*. \quad (8)$$

Therefore, we do not need to obtain the updated distribution model $X^*(t)$ but only the updated basis coefficient of its time warping function ϕ^* . We resort to MCMC to estimate the distribution of ϕ^* . Specifically, we use Metropolis-Hastings sampling to sample ϕ^* from its posterior distribution,

$$f(\phi^*|\mathbf{Y}^*; \hat{\boldsymbol{\theta}}) \propto f(\mathbf{Y}^*|\phi^*; \hat{\boldsymbol{\theta}})f(\phi^*|\hat{\boldsymbol{\theta}}) = f(\mathbf{Y}^*|\phi^*; \hat{\sigma}_\varepsilon^2, \hat{\boldsymbol{\beta}})f(\phi^*|b_0, \hat{\lambda}_\phi).$$

Details about the prediction algorithm are provided in the prediction part of supplementary section A. Using this method, we obtain a sample from the posterior distribution of RL_u^* . Further, based on this sample, we can obtain an estimate of the residual lifetime by taking the posterior mode or mean. Credible intervals can be used to provide a range of plausible values for the residual lifetime.

One practical question in the estimation and prediction procedures is assessing convergence of the proposed MCMC algorithms. Many diagnostic tests have been developed to date and some examples are the Gelman and Rubin test, the Geweke test, the Heidelberger and Welch test and so on. These diagnosis tests can be easily performed in R packages such as “boa” and “coda”. Generally, it is recommended to employ more than one test when assessing the quality of samples from an MCMC algorithm. More details about these tests can be found in Cowles and Carlin, 1996 and Brooks and Roberts, 1998.

5 CASE STUDY OF ROTATING MACHINERY DEGRADATION

The use of rotating machinery is widespread in industry. A large percentage of maintenance activities associated with this type of machinery generally involves inspection and replacement of bearings. Failure in bearings is generally due to a fatigue process. The degradation

process usually starts with spalling (the creation of minute pits) on the surface of the steel raceways. As the rolling elements pass over the spalls, they cause an increase in the vibration levels of the machine. Bearing-specific frequencies can then be observed using vibration monitoring techniques. These frequencies are a function of the rotational speed, type of bearings, number of rolling elements, and bearing geometry.

In our experiment, 34 bearings were tested until failure. For each bearing, vibration data was collected every 2 minutes, about 100 observations per signal. The experimental set up is described in detail by Gebraeel et al. (2005). We compare our proposed approach with the parametric model in Gebraeel et al. (2005) as a benchmark, which is proposed to model the same data set. In this model, an exponential functional form is identified to represent the degradation trend and the error term is assumed to follow a Brownian Motion process (we denote this model by “exp-brow” hereafter). Examples of bearing degradation signals are provided in Figure 1(b).

To evaluate the performance of our proposed degradation model, we use these vibration-based degradation signals and repeat the following study for 50 times. We randomly select 20 signals as training signals and the other 14 degradation signals are used as testing signals for evaluation of the absolute prediction errors. The training signals represent a database of training degradation signals for estimating the model. On the other hand, the validation signals represent signals that are observed from fielded bearings. Figure 4 shows the estimated common shape function; the degradation trend is neither linear nor exponential as assumed by the parametric model in Gebraeel et al. (2005). Table 1 summarizes the model parameter estimates.

Parameters	\hat{d}_β	\hat{d}_ϕ	$\hat{\beta}$	\hat{b}_0	$\hat{\lambda}_\phi$	$\hat{\sigma}_\varepsilon$
Estimates	4	4	[0, 0.00749, 0.00969, 0.025]	152.8	365	0.00407

Table 1: Estimates of Model Parameters

For each test bearing, we predict its RLD by using the partial observations of its degradation signal. To assess the prediction accuracy, we use the following error criterion:

$$\text{Absolute Prediction Error} = \frac{|\text{Predicted Life} - \text{Actual Life}|}{\text{Actual Life}} \tag{9}$$

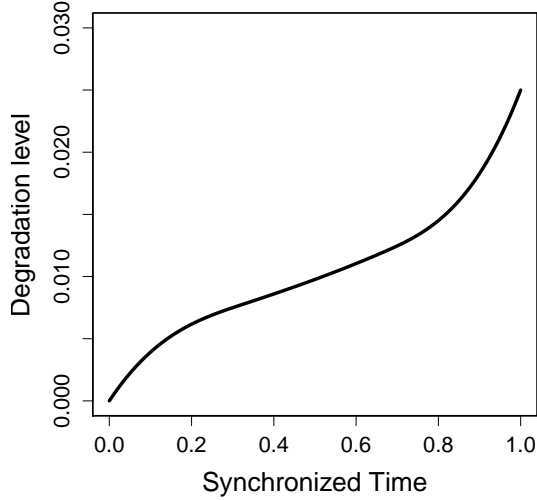


Figure 4: Estimated Common Shape Function

We compute the absolute prediction errors described in (9) at varying percentiles of a component’s entire life: 10%, 30%, \dots , and 90% (90% implies that 90% of the component’s life has passed). The prediction results of using “time warping” and “exp-brow” are summarized in Figure 5. The solid dots and lines represent the median and (± 2) standard deviations of the absolute prediction errors of “time warping”. We first observe that both the median and standard deviations of the absolute prediction errors decrease as the lifetime percentile increases. For example, the median absolute prediction error corresponding to the 10% percentile is around 30% while the median absolute prediction error corresponding to the 90% percentile is only around 7%. The proposed method thus shows clear improvements as new fielded data is available. The improvements are due to our proposed Bayesian method, which allows updating the distribution of the time warping coefficient by using real-time data.

In Figure 5, the dashed dots and lines represent the results of “exp-brow”. From this figure, we observe that the median absolute prediction errors of “time warping” are consistently smaller than those of “exp-brow” at all lifetime percentiles. This indicates that our proposed time warping approach predicts the RLD of fielded bearings with higher accuracy. At all lifetime percentiles except 50%, the standard deviations of the absolute prediction errors are smaller for those of “time warping” than for “exp-brow”. This indicates that our

proposed time warping approach is more stable in predicting the RLD of fielded bearings. The superior performance of our proposed time warping approach demonstrates that our model is relatively more robust to model mis-specifications.

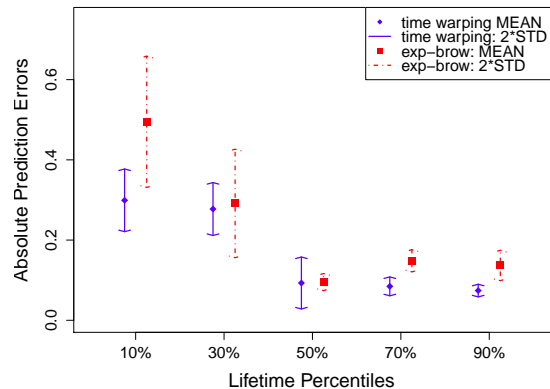


Figure 5: Absolute RLD prediction errors for “time warping” and “exp-brow”

6 SIMULATION STUDIES

We conduct three simulation studies for different purposes. In Section 6.1, we present a detailed simulation study, in which both the estimation and prediction results are reported and discussed. In the simulation study in Section 6.2, we focus on comparing the prediction performance of our proposed approach with the one proposed in Müller and Zhang (2005). The third simulation study included in supplementary section D is a sensitivity analysis, in which we assess the prediction performance under departures from the model assumptions and various settings of the parameters.

6.1 Simulation Study I

In this section, we assess the performance of our proposed functional time warping model by using simulated degradation signals. We repeat the following simulation 50 times. In each simulation, we randomly generate 70 signals according to the time warping model in (3). The common shape function $Z(\cdot)$ is generated as a linear combination of cubic B-

spline basis of the dimension $d_\beta = 5$. The basis coefficient vector for the common shape function is $\beta = (0, 40, 50, 60, 150)$. The time warping functions $\mu_i(\cdot)$ are also simulated as a linear combination of cubic B-spline basis functions, but of the dimension $d_\phi = 4$. The basis coefficient for each time warping function follows from a normal distribution defined over the constrained space as given by conditions (6) and (7). More specifically, $\phi_i \sim N(b_0 \mathbf{Y}, \mathbf{\Omega}_\phi / \lambda_\phi)$ with $b_0 = 100$ and $\lambda_\phi = 100$. The signals are generated with measurement errors and we set $\sigma_\varepsilon = 5$. For each component, we observe one degradation signal that consists of about 50 to 150 observations. The failure threshold is specified as $D = 100$. In Figure 6, we provide examples of the simulated degradation signals. The dots represent the degradation observations and the dashed lines represent the fitted underlying degradation process based on our proposed estimation algorithm.

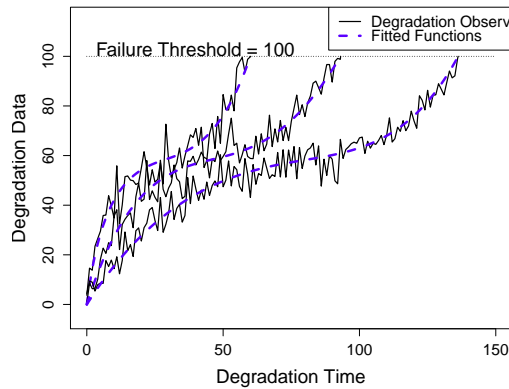


Figure 6: Simulation Study I: examples of simulated degradation signals with their estimated underlying functions.

In each simulation, we randomly select 50 signals as training degradation signals to estimate the model parameters and the remaining 20 signals are used as testing signals to evaluate how our model performs in terms of residual life prediction. To determine the tuning parameters d_β and d_ϕ , we compute the Bayes Factor score under different combinations of settings. In this paper, we use a cubic b-spline basis of order 4. This implies that the smallest possible value for d_β and d_ϕ is 4, since the number of basis cannot be smaller than the basis order. Due to this reason, we do not consider 3 or even smaller values for the basis dimensions. One example of the Bayes Factor Score (after scaling) versus d_β and d_ϕ is provided in Figure

7. The optimal values of d_β and d_ϕ are determined by maximizing the proposed Bayes Factor score. In this example, they are $(d_\beta, d_\phi) = (7, 4)$. In each simulation, we first repeat this model selection process and then proceed the estimation and prediction procedures by using the selected tuning parameters. Figure 8(a) displays the estimate of the common shape function $Z(\cdot)$ and Figure 8(b) displays the estimate of the underlying degradation trend for one component. In both figures, the solid lines represent the true functions and the dashed lines represent the estimated functions. Both the estimated common shape function and the estimated individual degradation trend function are estimated accurately. One can assess the residuals for the adequacy of the model assumptions. For instance, we can use Pearson chi-square test for assessing the normality assumptions of residuals or simply visualize the residuals using various graphical techniques. For this simulation study, the residuals do not show any significant pattern and, therefore, a normal distribution is adequate.

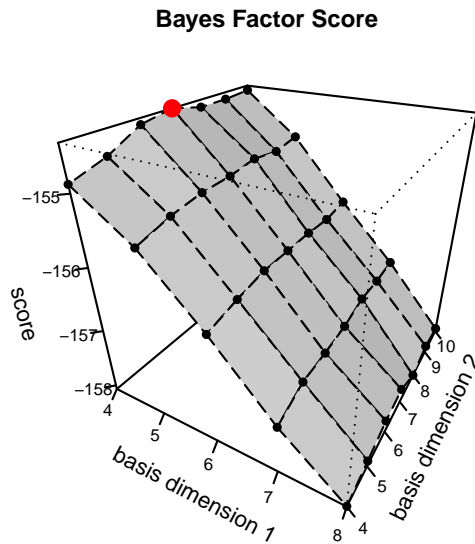


Figure 7: Simulation Study I: Bayes Factor Score v.s. Tuning Parameters (d_β and d_ϕ)

We update the RLD predictions as more and more observations are collected. This process is illustrated in Figure 9. In this example, we compute the absolute prediction errors described in (9) at varying percentiles of a component's entire life: 20%, 50%, and 80%.

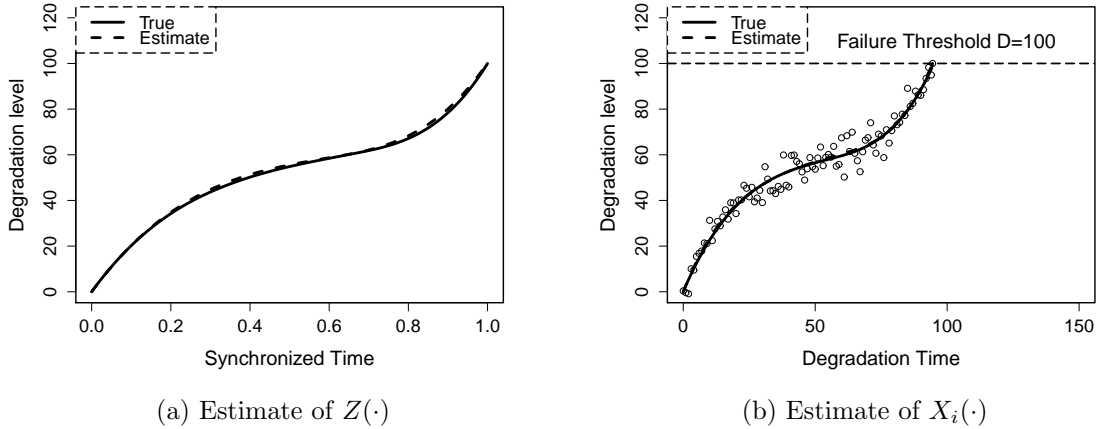


Figure 8: Simulation Study I: estimation of common shape function and individual underlying degradation function

The degradation observations are displayed in the left figure and the corresponding lifetime predictions are displayed in the right figure.

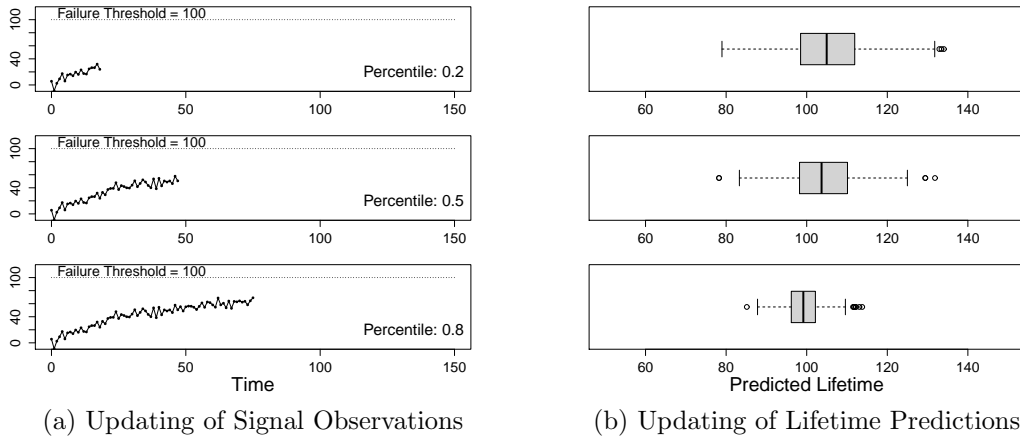


Figure 9: Simulation Study I: illustration of the updating process

The prediction results for all test units are summarized and reported in Figure 10, in which the x axis represents the time percentiles and the y axis represents the absolute prediction errors of the residual life. We compute the absolute prediction errors at varying percentiles of a component's entire life: 10%, 30%, ..., 90%. The median absolute prediction errors corresponding to each percentiles are all less than 10%. This indicates that our proposed model can predict the residual life accurately based on the partial observations of

its degradation signal. Another observation is that the median and variance of the absolute prediction errors decrease with increasing life percentiles. For example, the median absolute prediction error corresponding to the 10% percentile is around 8% while the median absolute prediction error corresponding to the 90% percentile is only around 2%. This observation is consistent with our bearing case study, which indicates the benefits of our proposed Bayesian updating framework.

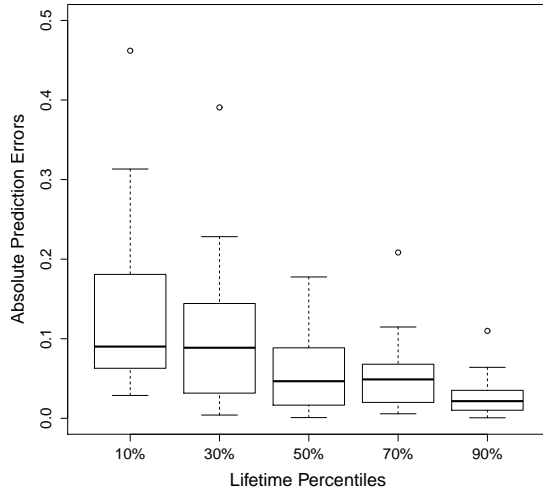


Figure 10: Simulation Study I: absolute RLD prediction errors at varying lifetime percentiles

6.2 Simulation Study II

In this section, we focus on comparing the performance of our proposed approach, “time warping” hereafter, and the approach proposed in Müller and Zhang (2005), “regression” hereafter. We will conduct two simulation studies according to two different model settings:

- Model Setting 1: the degradation signals are simulated based on our proposed model in (3). To simulate $Z(\cdot)$, we set $d_\beta = 5$ and $\beta = (0, 4, 6, 12, 30)$. The time warping functions $\mu_i(\cdot)$ are simulated with $d_\phi = 4$ and $\phi_i \sim N(20\Upsilon, \Omega_\phi/5)$ defined over the constrained space as given by conditions (6) and (7). The error standard deviation is $\sigma_\varepsilon = 0.75$. The average number of observations per signal is about 20. We set the failure threshold as $D = 20$.

- Model Setting 2: the degradation signals are generated according to the following formula: $Y_i(t) = \mu(t) + \xi_i\phi(t) + \epsilon_i(t)$ with $\mu(t) = 100t^2e^{t^2/9}$, $\xi_i \sim Uniform(-60, 60)$, $\phi(t) = t^2$, $0 \leq t \leq 1$, and $\sigma_\epsilon = 0.1$. The average number of observations per signal is about 15. The failure threshold is set as $D = 2$.

We repeat the following simulation process 50 times. In each simulation, we randomly generate 35 signals from which we randomly select 25 signals as training signals for estimation and the remaining 10 signals are selected as testing signals for validation. We use the error criterion described in (9) to assess the accuracy of the residual life predictions for the two methods, “time warping” and “regression”.

The results under the model setting 1 are summarized in Figure 11 (a). In this plot, we show the median and (± 2) standard deviations of the absolute prediction errors at varying lifetime percentiles for “time warping” (solid) and “regression” (dashed). The median absolute prediction errors of “time warping” are consistently smaller than those of “regression” at all lifetime percentiles. This indicates that our proposed method predicts the RLD with higher accuracy. Moreover, the standard deviations of the absolute prediction errors for “time warping” are consistently smaller than those of “regression”. This indicates that our proposed time warping method is more stable in predicting the RLD. Importantly, the difference in the absolute prediction errors between these two approaches is more significant at higher life percentiles. For example, the difference of the median absolute prediction errors is around 0.1% at the 10% percentile, while the difference is around 3.5% at the 90% percentile.

We are also interested in investigating the robustness of the performance difference between the two approaches to models that do not follow the decomposition in (3). To this purpose, we consider the model setting 2, which is equally challenging for both approaches. We summarize the prediction results in Figure 11 (b). Similar to simulation setting 1, “time warping” has smaller median and standard deviation of absolute prediction errors than those of “regression” although the difference is not as significant as that in the simulation setting 1.

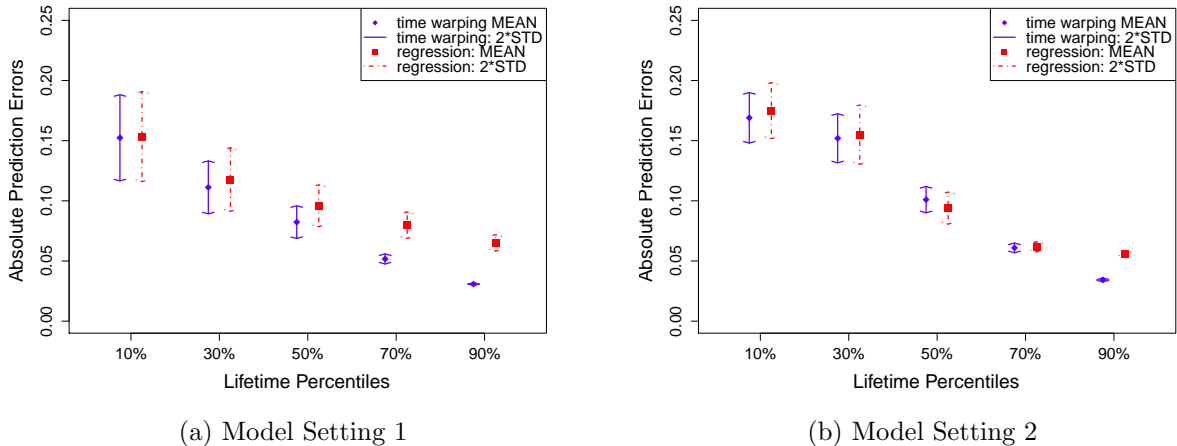


Figure 11: Simulation Study II: absolute RLD prediction errors for “time warping” and “regression”

7 SUMMARY

In this paper, we propose a functional time warping model for characterizing truncated degradation signals. Using this model, we can predict the RLD of a fielded component using partial observations of its truncated degradation signal. Our proposed model is flexible in that the degradation trend function is assumed to be unknown and estimated using a data-driven approach. When the available data are sparse and/or when the degradation trend does not have a simple form, our approach can be much more accurate and stable than the parametric benchmarks in predicting the RLD of fielded components. This has been demonstrated in our bearing case study.

In our modeling framework, one assumption is that all signals follow a model with an identical degradation trend. This assumption is in fact more general, especially in cases when the underlying degradation processes are monotonic. When $X_i(\cdot)$ is monotonic, there exists one monotonic function $Z(\cdot)$ (the common shape function in our context) and “transfer functions” $h_i(\cdot)$ (the “time warping functions” in our context), such that $X_i(\cdot) = Z(h_i(\cdot))$. This general result allows us to model a group of signals that may have quite different shapes. For example, in Figure 3, the degradation trend of component 1 is quadratic while that of component 2 is linear.

Importantly, we can extend our model framework by allowing the shape function to be

different among components. For instance, we may assume that $Z_i(t) = c_i + a_i Z(t)$, in which $Z(t)$ represents the common shape function, c_i and a_i are the additional random parameters that account for the heterogeneity of individual shape functions. This formulation is more general than the one proposed in this paper. However, under this framework, the image assumption does not hold, and therefore we can not predict the RLD of fielded components using formula (8). To assess the sensitivity of our model performance to this assumption, we have provided some preliminary studies in supplementary section D.

In this paper, the RLD predictions are updated as an online procedure. One possible concern of our approach is thus related to the computational speed of the prediction algorithm. On one hand, the computational time depends on many factors including the model complexity and the convergence rate of the MCMC estimation algorithm. For instance, the computational time for the updating process for all test units in the presented simulation study is at the level of a few seconds. On the other hand, how often the RLD predictions need to be updated varies from one application to another. In applications such as wind turbines and power generators, systems/components operate at stable degradation parameters for years. In these cases, updating the RLD predictions once every several weeks suffices for monitoring purposes. For other component types, the updating may be needed at smaller interval of times but not as small to require a computationally fast estimation algorithm.

In certain applications, the diagnostic variables being monitored (e.g., the example of dissolved gasses given in Duval, 2002; Saha, 2003; Wang and Vandermaar 2002) may be correlated. Therefore, one need to monitor different types of diagnostic variables simultaneously. This provides potential opportunities for future extensions to this line of work.

Supplementary materials

Additional details and derivations (sections A-D): We provide the details about our algorithms for model estimation, selection and prediction. Some additional simulation study results are also included. (pdf file)

Computer code: We provide the R codes for implementing all the algorithms described in this article. Additionally, we include a README file for instructional purposes. (zip)

file)

Acknowledgements

We thank Donatello Telesca for providing their codes. We are also thankful to the editor, the associate editor and the reviewers for their helpful comments and suggestions. Rensheng Zhou has been supported by the National Science Foundation under grant CMMI-0738647.

References

- [1] Brooks, S.P., and Roberts, G.O. (1998), “Convergence Assessment Techniques for Markov Chain Monte Carlo” *Statistics and Computing* 8(4), 319-335.
- [2] Brumback, C.L., and Lindstrom, J.M. (2004), “Self-modeling With Flexible, Random Time Transformations” *Biometrics* 60, 461-470.
- [3] Cowles, M.K. and Carlin, B.P. (1996), “Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review” *Journal of the American Statistical Association* 91, 883-904.
- [4] Duval, M. (2002), “A review of faults detectable by gas-in-oil analysis in transformers”, *IEEE Electrical Insulation Magazine* 18(3), 8-17.
- [5] Eilers, P., and Marx, B. (1996), “Flexible Smoothing Using BSplines and Penalized Likelihood” *Statistical Science* 11, 1200-1224.
- [6] Gebraeel, N., Lawley, M., Li, R. and Ryan, J. (2005), “Residual-life distributions from component degradation signals: A Bayesian approach” *IIE Transactions* 37, 543-557.
- [7] Gebraeel, N. and Pan, J. (2008), “Prognostic degradation models for computing and updating residual life distributions in a time-varying environment”, *IEEE Transactions on Reliability* 57 (4), 539-550.

- [8] Gervini, D. and Gasser, T. (2004), “Self-modeling Warping Functions” *Journal of the Royal Statistical Society (Series B)* 66(2), 959-971.
- [9] Gervini, D. and Gasser T. (2005), “Nonparametric maximum likelihood estimation of the structural mean of a sample of curves” *Biometrika* 92 (4), 801-820.
- [10] Kass, R. and Raftery, A. (1995), “Bayes factors” *Journal of the American Statistical Association* 90, 773-795.
- [11] Kneip, A. and Gasser T.(1992), “Statistical tools to analyze data representing a sample of curves” *The Annals of Statistics* 20 (3), 1266-1305.
- [12] Lang, S., and Brezger, A. (2004), “Bayesian P-Splines” *Journal of Computational and Graphical Statistics* 13, 183-212.
- [13] Liu, X. and Müller, H.G. (2004), “Functional convex averaging and synchronization for time-warped random curves” *Journal of the American Statistical Association* 99, 687-699.
- [14] Lu, J.C. and Meeker, W. (1993), “Using degradation measures to estimate a time-to-failure distribution”, *Technometrics* 35, 161-174.
- [15] Medjaher, K, Tobon-Mejia, D.A. and Zerhouni, N. (2012), “Remaining Useful Life Estimation of Critical Components With Application to Bearings” *IEEE TRANSACTIONS ON RELIABILITY* 61(2), 292-302.
- [16] Müller, H.G. and Zhang, Y. (2005), “Time-varying functional regression for predicting remaining lifetime distributions from longitudinal trajectories”, *Biometrics* 61, 1064-1075.
- [17] Nelson, W. (1990), “Accelerated Testing Statistical Models, Test Plans and Data Analysis” John Wiley & Sons, New York.
- [18] Saha, T. K. (2003), “Review of modern diagnostic techniques for assessing insulation condition in aged transformers”, *IEEE Transactions on Dielectrics and Electrical Insulation* 10(5), 903-917.

- [19] Tang, R. and Müller, H.G. (2008), “Pairwise curve synchronization for functional data” *Biometrika* 95, 875–889.
- [20] Tang, R. and Müller, H.G. (2009), “Time-synchronized clustering of gene expression trajectories” *Biostatistics* 10, 32–45.
- [21] Telesca, D. and Inoue, L.Y.T. (2008), “Bayesian hierarchical curve registration” *Journal of the American Statistical Association* 103 (481), 328-339.
- [22] Virkler, D.A., Hillberry, B.M. and Goel, P.K. (1979), “The statistical nature of fatigue crack propagation”, *Transactions of the ASME - Journal of Engineering Materials and Technology* 101, 148-153.
- [23] Wang, K. and Gasser, T. (1997), “Alignment of curves by dynamic time warping” *The Annals of Statistics* 25, 1251-1276.
- [24] Wang, K. and Gasser, T. (1999), “Synchronizing sample curves nonparametrically” *The Annals of Statistics* 27, 439-460.
- [25] Wang, M. and Vandermaar, A.J. (2002), “Review of condition assessment of power transformers in service” *IEEE Electrical Insulation Magazine* 18(5), 8-17.
- [26] Zhou, R.R., Serban, N., and Gebraeel, N. (2011), “Degradation modeling Applied to Residual Lifetime Prediction using Functional Data Analysis”, *The Annals of Applied Statistics*, 5(2B) 1586-1610.
- [27] Zhou, R.R., Gebraeel, N., and Serban, N. (2012), “Degradation modeling and monitoring of truncated degradation signals”, *IIE Transactions*, 44(9) 793-803.