Reliability Engineering and System Safety 133 (2015) 266–274

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

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

An adaptive functional regression-based prognostic model for applications with missing data

Xiaolei Fang, Rensheng Zhou, Nagi Gebraeel *

Georgia Institute of Technology, Atlanta, GA 30332, USA

ARTICLE INFO

Article history:
Received 27 January 2014
Received in revised form
19 August 2014
Accepted 31 August 2014
Available online 16 September 2014

Keywords:
Condition monitoring
Prognostics
Functional principal components analysis
Functional regression analysis
Remaining useful life

ABSTRACT

Most prognostic degradation models rely on a relatively accurate and comprehensive database of historical degradation signals. Typically, these signals are used to identify suitable degradation trends that are useful for predicting lifetime. In many real-world applications, these degradation signals are usually incomplete, i.e., contain missing observations. Often the amount of missing data compromises the ability to identify a suitable parametric degradation model. This paper addresses this problem by developing a semi-parametric approach that can be used to predict the remaining lifetime of partially degraded systems. First, key signal features are identified by applying Functional Principal Components Analysis (FPCA) to the available historical data. Next, an adaptive functional regression model is used to model the extracted signal features and the corresponding times-to-failure. The model is then used to predict remaining lifetimes and to update these predictions using real-time signals observed from fielded components. Results show that the proposed approach is relatively robust to significant levels of missing data. The performance of the model is evaluated and shown to provide significantly accurate predictions of residual lifetime using two case studies.

1. Introduction

Prognostic degradation models focus on characterizing how degradation signals (condition-based sensor signals) evolve over time. Typically, the goal is to estimate lifetime, or in our case, predict and update remaining lifetime in real-time. A large number of degradation models have been proposed in the literature, such as the ones presented in [1–4,15–17,11,5,12,6–10,13,14,18]. The key objective of these models is to predict failure or explicitly estimate remaining useful lifetimes. However, the effectiveness of these models relies primarily on the fidelity of estimating their parameters. Parameter estimation is often driven by the characteristics and the quality of historical data (also known as, training data). For example, most models assume that historical degradation signals are completely and, for all practical purposes, continuously observed from an "as good as new" state up to the point of failure. In reality, however, continuous or frequent observations of degradation are not always possible or economical. Examples of such scenarios include monitoring cracks on gas turbine blades that require shutting down the turbine or assessing the concentration of dissolved gases in transformers. Other examples may involve sensor failure or disconnection. Thus, in practice, it is more likely that degradation signals are observed randomly or at intermittent points in time resulting in sparse or fragmented signal observations as illustrated in Fig. 1.

If parametric models are used to model signals with such high levels of missing data, it is likely that the available data will not be enough to accurately identify a suitable trend or general path for the degradation signals. In this paper, we utilize a semi-parametric approach to develop a prognostic degradation model for sparse and fragmented signals. First, Functional Principal Component Analysis (FPCA) is used to identify key features of the incomplete signals. FPCA is a nonparametric Functional Data Analysis (FDA) technique that identifies the important sources of pattern variation among functional data. One important benefit of FPCA is that it provides a low-dimensional and parsimonious representation of each curve by reducing it to a set of functional principal components scores (i.e., FPC-scores), which were referred to earlier as signal features. The FPC-scores estimated using signals with missing observations are likely to be similar to those that would have been estimated if all observation were present. Of course, as the level of missing data increases so does the difference between these scores. Once the signal features are extracted, an adaptive functional regression model is used to model the relationship between the FPC-scores and historical Times-to-Failure (TTFs). Functional regression is an extension of ordinary regression that

* Corresponding author. Tel.: +1 404 894 0054; fax: +1 404 385 6178.
E-mail address: nagi@isye.gatech.edu (N. Gebraeel).

http://dx.doi.org/10.1016/j.ress.2014.08.013
0951-8320/

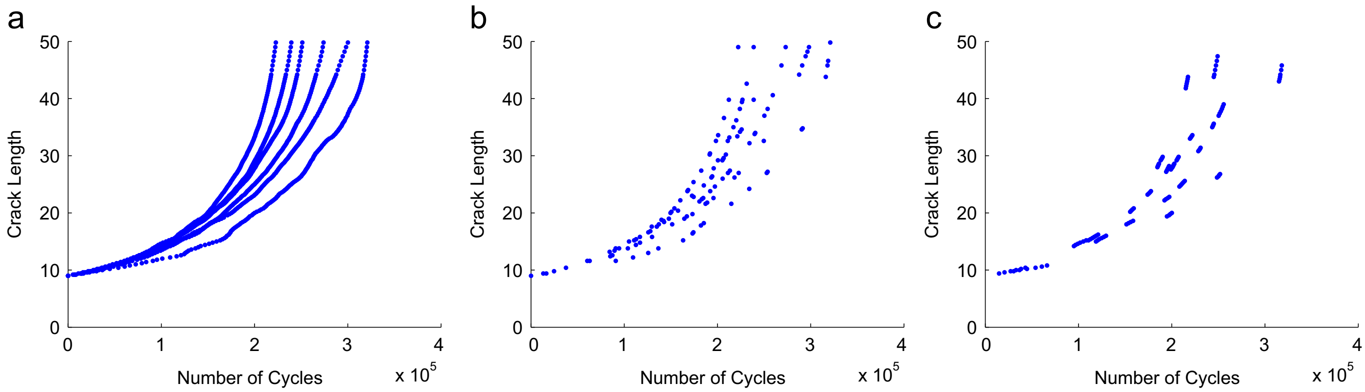

Fig. 1. Examples of complete, sparse and fragmented degradation signals from crack growth data. (a) Complete signals. (b) Sparse signals. (c) Fragmented signals.

accounts for the case where predictors are random functions and responses are scalars or functions [19]. A popular approach for implementing functional regression with scalar responses is to represent predictor functions with FPC-scores, and fit an ordinary regression model on these scores and the response scalars (see for example [19–23]).

Our proposed framework also provides a means to incorporate in situ signals observed from partially degraded components in the field in order to update the model. The updated model is then used to revise the predicted remaining lifetimes.

The remainder of the paper is organized as follows: Section 2 reviews some of the relevant nonparametric modeling approaches. Section 3 discusses the development of the degradation model. In Section 4, we introduce the functional regression model that is used for predicting residual lifetime. An adaptive approach for updating the residual life predictions is discussed in Section 5. We then evaluate the performance of our methodology using real-world crack growth degradation signals in Section 6, and real-world vibration-based bearing degradation signals in Section 7. Finally, conclusions and future research are presented in Section 8.

2. Literature review

Many degradation modeling approaches have been developed in the literature. Some examples involve the use of neural networks and fuzzy logic models [4–7], Kalman and particle filtering techniques [8–10], statistical degradation models [11–14], and various types of stochastic processes, such as the Wiener and Gamma processes [15–18]. However, most of the existing models focus only on complete degradation signals. Some research efforts have focused on modeling sparse and irregularly observed degradation data (also known as *longitudinal data* in the bioinformatics and medical literature) using FDA [24–32]. For example, in [24] the authors used FPCA for applications with sparse longitudinal data. They assumed that repeated measurements exhibit an underlying smooth random (subject-specific) trajectory plus measurement errors. They proposed the PACE approach (principal components analysis through conditional expectation) to extend the applicability of FPCA to situations with sparse longitudinal data. In [25], the authors introduced a latent Gaussian process model for sparse longitudinal data measurements observed at irregular intervals. They also used FPCA and illustrated their model on biliary cirrhosis data. Sentürk and Nguyen [26] applied FPCA to model sparse and noise-contaminated longitudinal data. This work was later extended in [27] to address sparse and unsynchronized longitudinal data. In [28], the authors incorporated FPC-scores in a reduced rank mixed effects framework. In [29], the authors proposed a support vector machine approach that used FPCA to perform multi-category classification of sparse data with a multi-category response.

The underlying theme of most of the research mentioned above is that different realizations of the longitudinal data (which are analogous to our signals) are assumed to share the same time domain. In other words, the start and end points are the same for all the realizations. As pointed out in [30–32], one of the major limitations of using FPCA is that it indeed requires that each observed curve shares the same time domain. Unfortunately, this is not the case for most engineering systems as they are often shutdown, taken off-line for repair, or replaced once their degradation signals reach an alarm threshold (see for example ISO 2372 and 10816 for machine vibration). In other words, degradation signal are typically truncated at the threshold beyond which no subsequent observations can be acquired. In such scenarios, using FPCA results in a significantly biased estimate of the mean and covariance functions due to the fact unobserved data beyond the threshold [30]. This aspect is a unique difference between most of the related literature and the work proposed in this paper.

There have been some recent attempts to tackle this problem. For example, [31] proposed a procedure that relied on axis transformation. (Instead of plotting the degradation level on the y-axis with time on the x-axis, the axes were reversed.) However, this approach created another problem especially with noisy signals whereby the same degradation level may correspond to multiple time stamps. From a practical standpoint, this was infeasible. Thus, the approach was limited to strictly monotonic signals with very low noise levels. [32] developed a functional time warping approach that can synchronize the truncated degradation signals, but it focuses only on complete degradation signals. Unlike existing work, our modeling approach is significantly more general in that it can be applied to various types of signals, and has no restrictions on the signal-to-noise ratio.

3. Degradation model development

As mentioned earlier, we consider a problem setting wherein historical degradation signals contain a significant level of missing data. Specifically, we focus on two types of signals, sparsely observed and fragmented signals, as illustrated earlier in Fig. 1. Furthermore, signals are only observed up to a predetermined failure/replacement threshold. We assume that degradation signals are independent realizations of a smooth random function over a bounded time domain $[0, M]$. M is the maximum observation time point, which is assumed to be finite since any industrial application has a finite time-of-failure. (Hypothetically, this is the degradation signal of the component with the longest possible lifetime.) Taking observation error into account, our approach is to model the amplitude of the degradation signal of the ith component, $S_i(t)$, as follows:

$$S_i(t) = \mu(t) + X_i(t) + \epsilon_i(t), \tag{1}$$

where $i = 1, \ldots, n$, and n represents the number of components, $\mu(t)$ represents the underlying trend of the degradation signal that is common to the entire population and is assumed to be deterministic, $X_i(t)$ represent stochastic deviations from the underlying trend due to the inherent variability in the degradation rates of the components. $X_i(t)$ are assumed to follow a stochastic process with mean zero and covariance function $cov((X(t), X(t')) = C(t, t')$. Finally, $\epsilon_i(t)$ are independent and identically distributed observation errors with mean zero and variance σ^2. $X_i(t)$ and $\epsilon_i(t)$ are assumed to be independent.

Using the Mercer's theorem [34], the covariance function $C(t, t')$ can be expanded as shown below,

$$C(t, t') = \sum_{k=1}^{\infty} \lambda_k \phi_k(t) \phi_k(t'), \ \ t, t' \in [0, M], \tag{2}$$

where $\phi_k(t)$ for $k = 1, 2, \ldots$ are the eigenfunctions and $\lambda_1 \geq \lambda_2 \geq \ldots$, are the ordered nonnegative eigenvalues.

Since the eigenfunctions are orthogonal and form a functional basis, $X_i(t)$ can be expressed as follows:

$$X_i(t) = \sum_{k=1}^{\infty} \xi_{ik} \phi_k(t), \tag{3}$$

where ξ_{ik} for $k = 1, 2, \ldots$ are the FPC-scores. These scores are independent random variables with mean $E[\xi_{ik}] = 0$ and variance $E[\xi_{ik}^2] = \lambda_k$. Thus, the amplitude of the degradation signal can now be rewritten as follows:

$$S_i(t) = \mu(t) + \sum_{k=1}^{\infty} \xi_{ik} \phi_k(t) + \epsilon_i(t). \tag{4}$$

Generally, the eigenvalues λ_k for $k = 1, 2, \ldots$ decrease to zero rapidly. Therefore, it is reasonable to assume that an appropriate K can always be chosen to approximate $X_i(t)$. For example, Eq. (4) can be truncated by selecting K to minimize the Akaike information criterion (i.e., AIC) defined by [23]. AIC is a widely used model selection criterion, which deals with the trade-off between the goodness-of-fit and the complexity of the model. In our case, we select the first k eigenfunctions to approximate Eq. (4) and calculate their corresponding AIC values, i.e., $\text{AIC}_k (k = 1, 2, \ldots; \lambda_k > 0)$. Then the k corresponds to the smallest AIC value is chosen to truncate our model. The (truncated) model that will be the basis of our work is expressed below by Eq. (5):

$$S_i(t) = \mu(t) + \sum_{k=1}^{K} \xi_{ik} \phi_k(t) + \epsilon_i(t). \tag{5}$$

The different parts in Eq. (5), i.e., $\mu(t)$, ξ_{ik}, $\phi_k(t)$ and $\epsilon_i(t)$, can be estimated using the historical degradation signals. Details of the estimation can be found in Appendix A.

4. Functional regression analysis

We now discuss the functional regression model used to estimate component lifetime. First, we denote the lifetime of the ith component as Y_i. By setting the scalar Y_i as response and degradation signals $S_i(t)$ as the predictor, we can establish a classic linear functional regression model [19]:

$$Y_i = \int_{[0,M]} \alpha(t) S_i(t) \, dt, \tag{6}$$

where the regression parameter function $\alpha(\cdot)$ is assumed to be smooth and square integrable over the interval $[0, M]$. Since $\phi_k(t)$ form a complete orthonormal basis for $k = 1, 2, \ldots$, $\alpha(t)$ can be expanded in terms of these basis as follows: $\alpha(t) = \sum_{k=1}^{\infty} \beta_k \phi_k(t)$. Since $S_i(t) = \mu(t) + \sum_{k=1}^{\infty} \xi_{ik} \phi_k(t) + \epsilon_i(t)$, we can now express the our functional regression model as follows:

$$\begin{aligned} Y_i &= \int_{[0,M]} \alpha(t)\mu(t) \, dt + \int_{[0,M]} \left\{ \sum_{k=1}^{\infty} \beta_k \phi_k(t) \right\} \left\{ \sum_{k=1}^{\infty} \xi_{ik} \phi_k(t) \right\} dt \\ &\quad + \int_{[0,M]} \alpha(t) \epsilon_i(t) \, dt = \beta_0 + \sum_{k=1}^{\infty} \beta_k \xi_{ik} + \varepsilon_i, \end{aligned} \tag{7}$$

where $\beta_0 = \int_{[0,M]} \alpha(t)\mu(t) \, dt$ is the intercept, $\beta_k = \int_{[0,M]} \alpha(t) \phi_k(t) \, dt$ are regression coefficients, and $\varepsilon_i = \int_{[0,M]} \alpha(t) \epsilon_i(t) \, dt$ is the error term. Recall that we truncate $S_i(t)$ using a finite sum of K terms to approximate the infinite sum in Eq. (5). Using this truncated form, the regression model can be expressed as,

$$Y_i = \beta_0 + \sum_{k=1}^{K} \beta_k \xi_{ik} + \varepsilon_i. \tag{8}$$

The coefficients β_k for $k = 1, \ldots, K$ can be estimated using least square estimation. We assume that there exists a database of historical degradation signals that we refer to as training signals. We also assume that the lifetimes of the components corresponding to the training signals are known a priori. The training signals are first used to estimate the FPC-scores as discussed in Section Appendix A. Next, the regression model is estimated using the FPC-scores and the corresponding lifetimes. Note that more complicated functional regression models, for example, functional quadratic regression model or functional polynomial regression model, can be also used to mode the relationship between the lifetime Y_i and degradation signals $S_i(t)$.

The main objective of this framework is to enable predicting the residual lifetime of components that are still operating in the field. To do this, we propose an updating scheme whereby in situ observations from the fielded components are leveraged to update the model based on the latest degradation states of the fielded components, thus improving the accuracy of remaining life predictions.

5. Estimating and updating remaining lifetimes

Degradation signals observed from fielded components provide a wealth of knowledge about their current degradation. Consequently, an inherent component of our framework is to provide the means to incorporate in situ signals observed from the field and utilize them to provide more accurate predictions of remaining lifetime. To illustrate this updating scheme, recall that there are two sets of degradation signals, training and validation. The training signals are used to estimate the FPC-scores, and build a regression model using the scores and the TTFs of their corresponding signals. On the other hand, the set of validation signals are assumed to represent signals observed from fielded components. Next, consider a validation signal observed up to a specific time point. The FPC-scores of the partial validation signal are estimated using the eigenfunctions calculated from the training set. The FPC-scores are then input to the regression model and a time-to-failure for the validation signal is estimated. As more signals are observed, the FPC-scores are revised and a new TTF is estimated.

As mentioned earlier, to correctly use FPCA, all the functional curves associated with the degradation signals must share the same time domain. However, the training and validation signals used during the updating scheme will not possess this characteristic (probably never will for engineering applications where a failure threshold is used to plane for maintenance or repair). Due to the existence of a failure threshold, the support over which the signals are observed will vary from one component to another. To address this limitation, we borrow an adaptive updating approach that was first presented by [24].

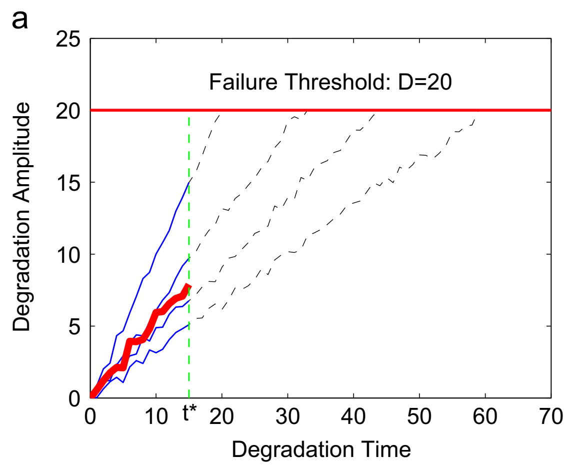

Fig. 2. Method of updating the model as time advances. (a) Updating at time $t^* = 15$. (b) Updating at time $t^* = 25$.

To illustrate this approach, we first consider a validation signal observed up to time t^* (see Fig. 2). We then select only the training signals that have survived up to time t^*. This subset of training signals is used to compute FPC-scores (along with the eigenvalues and eigenfunctions) and estimate the functional regression model. Next, we estimate the FPC-scores of the validation signal (represented by the thicker curve in Fig. 2). By applying the validation FPC-scores to the revised functional regression model, the lifetime of the "validation component" can be estimated. The remaining lifetime can be easily computed by subtracting the current operating time t^*.

As more observations become available, different subsets of the training degradation signals will be selected at different values of t^*. However, the same estimation process will be repeated at every new value of t^*. Fig. 2 shows an example of how training signals are selected at different observation times of the validation signal. The dotted lines represent the entire set of training signals. The continuous part marks the part of the data that is used for estimating the functional regression model. The signal marked with a thick line represents the portion of the validation signal observed at time t^*.

In the following section, we evaluate the performance of our methodology using two different sets of degradation data. The first case study deals with crack growth data from several tests specimens of aluminum alloy and the second uses relatively noisier vibration-based data from a rotating machinery.

6. Case study of crack growth data

In this study, crack growth data first presented by [35] is used to test the proposed model. The data set consists of crack propagation measurements pertaining to 68 identical center-cracked aluminum plates that were tested under identical experimental conditions. The data consists of the number of cycles for discrete levels of crack length from 9.0 mm to 49.8 mm. Each crack growth signal consisted of 164 data points.

We randomly chose 58 signals as our training signals with the remaining 10 signals for validation, i.e., representing signals observed in real-time from fielded components. The performance of our model was tested under three scenarios: (1) signals with complete observations, (2) signals with sparse observations, and (3) fragmented signals. For the complete signals, we use the original crack data since observations were made at a relatively high frequency. In the sparse scenario, we randomly sample 12 observations from each training and validation signal. In this scenario, model estimation and validation are conducted using the sparse signals, i.e., the sparse training signals are used for model estimation and the validation signals are used to evaluate performance. A similar approach is also used in the fragmented scenario except that the fragmented signals are generated by randomly sampling 3 fragments with 4 observations per fragment from each original signal (training and validation). A sample of the data is shown earlier in Fig. 1. The evaluation process is replicated 100 times, and the performance of our model is evaluated by computing the prediction error at predetermined life percentiles. Prediction errors are calculated at different life percentiles using the following error criteria

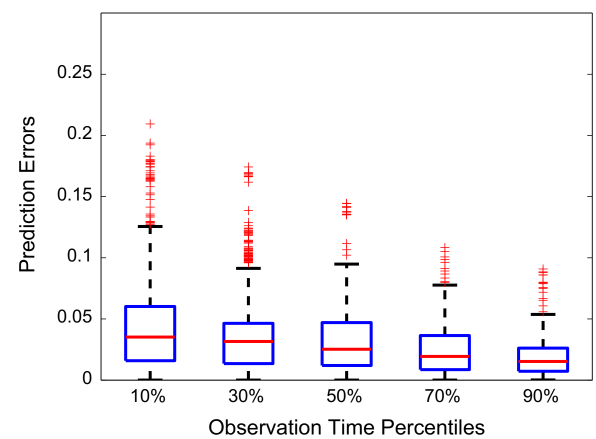

Fig. 3. Prediction errors of the proposed model under the complete degradation signals scenario.

$$\text{error} = \frac{|\text{Estimated Life} - \text{Actual Life}|}{\text{Actual Life}} \times 100\%. \tag{9}$$

6.1. Results and analysis

The prediction errors for the three signal scenarios are evaluated at the following life percentiles: 10% (10% of the component's lifetime has passed at the point the prediction was made), 30%, 50%,70% and 90%. Figs. 3, 4 and 5 depict box plots of the prediction errors for the complete, sparse, and fragmented scenarios, respectively. Note that remaining life predictions are updated based on the validation signals that have been observed at each life percentile.

Overall, the complete signal scenario illustrated Fig. 3 has the lowest mean prediction error. It also has the smallest confidence intervals across all life percentiles when compared to the other two signal scenarios. The box plots also demonstrate the effect of the

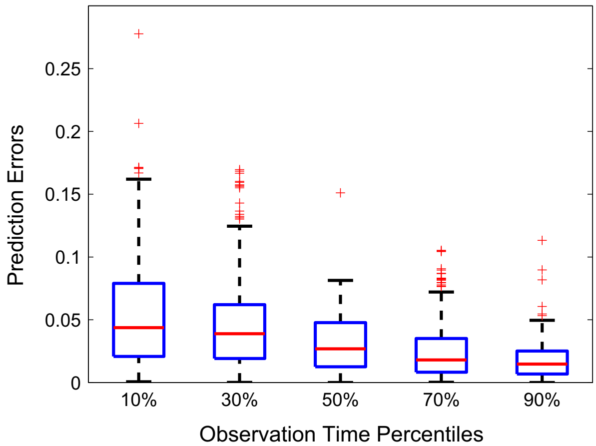

Fig. 4. Prediction errors of the proposed model under the sparse degradation signals scenario.

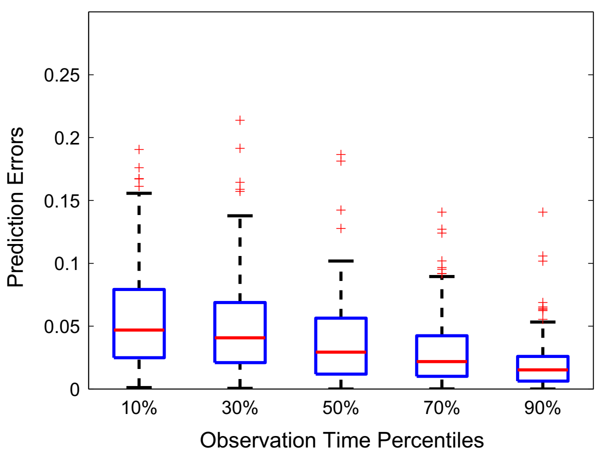

Fig. 5. Prediction errors of the proposed model under the fragmented degradation signals scenario.

updating process which can be clearly seen through the progressive reduction in the mean prediction error and corresponding variance. When comparing the prediction errors across the three signal scenarios, we observe that although significantly less signal observations were used in the sparse and fragmented scenarios, the proposed model still maintains relatively robust performance compared to complete scenario (compare Fig. 3 with Figs. 4 and 5). In fact, there seems to be no significant difference between the three signal scenarios at and beyond the 50*th* life percentile.

Based on these results, we believe that our modeling approach demonstrates significant potential for predicting remaining lifetimes of fielded, especially in applications with significant levels of missing data. This conclusion is of course governed by the characteristics of the signals. For example, results are especially promising in this example because the signals have a high signal-to-noise ratio. In the next case study, we evaluate the performance of our model using degradation signals with a significantly lower signal-to-noise ratio.

7. Case study of rotating machinery degradation

In this case study, the performance of our model is evaluated using vibration-based degradation signals generated from degradation of a rotating machinery. Specifically, the experimental test rig is designed to perform accelerated degradation tests on rolling element thrust bearings. Vibration signatures are used to monitor bearing degradation. A detailed description of the experimental setup, test conditions, and the degradation signals can be found in [36]. The degradation signals used in the study represent the average amplitude of the defective frequency and its first six harmonics over time. A bearing is considered to have failed once the amplitude of its degradation signal crosses a pre-specified threshold of 0.025 V_{rms} (which are mapped from industrial ISO standards). For the purpose of brevity, the reader is referred to [36] for additional details.

The data set consists of degradation signals for 31 bearings that were tested until failure. Observations were acquired every 2 minutes with lifetimes ranging between 12 to 36 hours. For the sparse scenario, a new training is created by randomly choosing 10 observations from each of the original complete signals. A similar process is also used to construct the degradation signals for the fragmented case. Three fragments are randomly chosen from each original signal, each fragment is consisting of 3 observations. Examples of sparse and fragmented bearing degradation signals are shown in plot "a" of Figs. 6 and 7. A leave-one-out cross validation is performed by choosing one signal and using the remaining 30 for training and model estimation. Prediction errors are estimated using the same expression in Eq. (9).

The above procedure starting with creating the database up to the cross validation and evaluating the prediction errors is repeated 20 times, resulting in 620 validation tests. This is performed for the complete, sparse, and fragmented signals. To benchmark the performance of our model, the sparse and fragmented signal scenarios are compared to two existing models (1) 'classical FPCA', and the (2) 'axis-transformation FPCA' model. With respect to the complete signal scenario, our model is benchmarked against the parametric exponential stochastic model [13] that was developed for this specific data set.

7.1. Results and analysis for degradation signals with missing data

We now discuss the benchmark models used for the sparse and fragmented signals. The first benchmark model, which we define as 'classical FPCA', involves using FPCA in the classical sense. In other words, FPCA is used to fit a nonparametric model to data. The model is then used to estimate failure times and/or remaining lifetimes by extrapolation, given a predefined failure threshold. This is similar to the manner in which FPCA was used in [33,25,28,24]. The second benchmark is the 'axis-transformation FPCA' model proposed by [31], which was discussed in Section 2.

Prediction errors for the three models were evaluated using Eq. (9) and are summarized in Figs. 6 and 7. Results show that on average the prediction errors when using our functional regression model are relatively lower than the other two benchmark models. This is true for both the sparse and fragmented scenarios. As mentioned earlier, one of the main shortcomings of 'classical FPCA' is that it requires all signals to share the same time domain, a characteristic that is not necessarily satisfied especially that many equipment are shutdown for repair or maintenance once their degradation signals reach a pre-specified threshold.

Prediction errors for the three models are evaluated in a similar manner as discussed in the previous case study using Eq. (9). By observing the box plots of Figs. 6 and 7, it can be seen that our approach performs better than 'classical FPCA'. Although FPCA has been used proven to work well in situations involving sparse data, one of its main shortcomings, as mentioned earlier, is the fact that it assumes all signals share the same time domain. From a practical viewpoint, this attribute does not hold in many engineering applications where components may be shutdown for repair/maintenance at different times once their degradation signals reach a pre-specified alarm or replacement threshold. By comparing the mean and variance of the prediction errors at the 10th, 30th, 50th, and 70th life percentiles, we see that both are significantly smaller

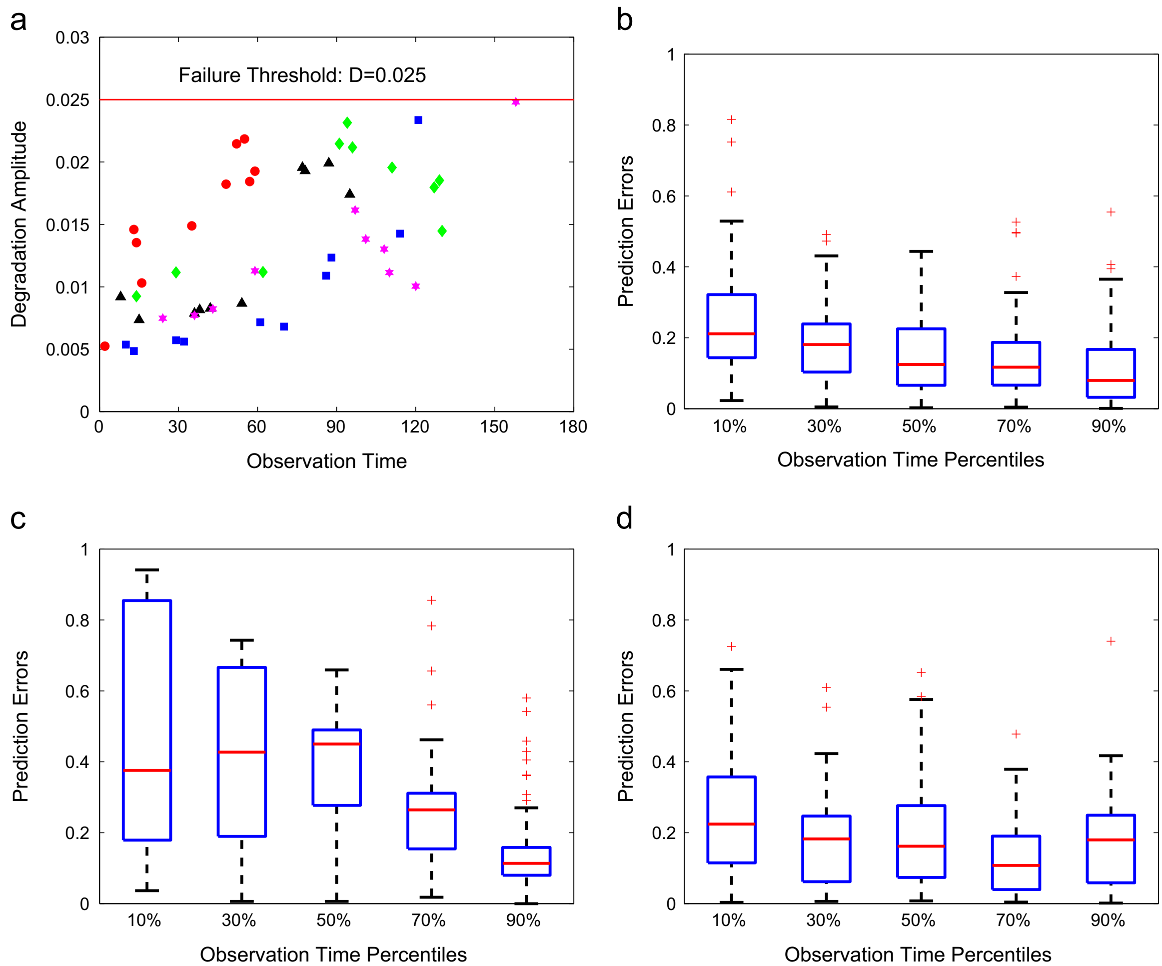

Fig. 6. Plots of prediction errors under the sparse signals scenario. (a) A sample of sparse degradation signals. (b) Prediction errors using the functional regression model. (c) Prediction error using a classical FPCA model. (d) Prediction errors using the axis-transformation FPCA model.

in our model compared to the 'classical FPCA' model. Even though the variance of the prediction errors at the 90th percentile in our approach is higher than 'classic-FPCA', the mean error remains relatively smaller. The increased variance phenomenon can be attributed to the fact that fewer components tend to have long lifetimes. This fact coupled with the sparsity of the data creates a significant level of variability at the 90th life percentile when using our approach.

By comparing Figs. 6 and 7, we can see that our approach edges a little over the 'axis-transformation FPCA' model proposed by [31]. One observation that is clear is that the variance of the prediction errors for the benchmark model change from one life percentile to another. In addition they are consistently greater than those of our model. This observation is even more pronounced in the case of the fragmented signals (compare plots (b) and (d) of Fig. 7). We believe this observation may be attributed to the fact that benchmark model is restricted to monotonic degradation signals. Consequently, for the relatively noisy signals used in this case study, the monotonicity condition is often violated resulting in poor performance of the model. Relative to the sparse signals, the effect of the noise is more visible in the fragmented case.

7.2. *Results and analysis for complete signals*

For the complete signal scenario, we compared our functional regression model to the exponential Brownian model (designated as 'exp-Brown') presented in [13]. The model consists of an random coefficients model with an exponential trend function and a Brownian motion error term. As mentioned earlier, this model was used as a benchmark because there was enough observations in the complete signals to identify a suitable degradation trend. In the same spirit, no parametric benchmark models are used for the sparse and fragmented scenarios since the significant levels of missing data are not enough to identify suitable trend functions. Box plots of the prediction errors are summarized in Fig. 8. The plots show that the functional regression model performed well compared to the 'exp-Brown' model. Variance was significantly smaller at the 10th, 30th, and 90th life percentiles when using the functional regression model. The benchmark model edges slightly over our model at the 50th life percentile in terms of the variance of the prediction error, but there seems to be no significant difference in terms of the mean of the prediction error. Based on the plots, it seems that our functional regression model still retains its effectiveness in the case of complete signals.

8. Summary

Many industrial applications provide the capability of real-time monitoring of performance and degradation. However, a common challenge in many industries is how to utilize this information given that often times there are missing data. In this paper, we presented a functional regression model capable of predicting and updating, in real-time, the remaining lifetime of engineering systems. Our approach is best suited for applications in which degradation signals have different forms of missing data, i.e., sparse or fragmented data.

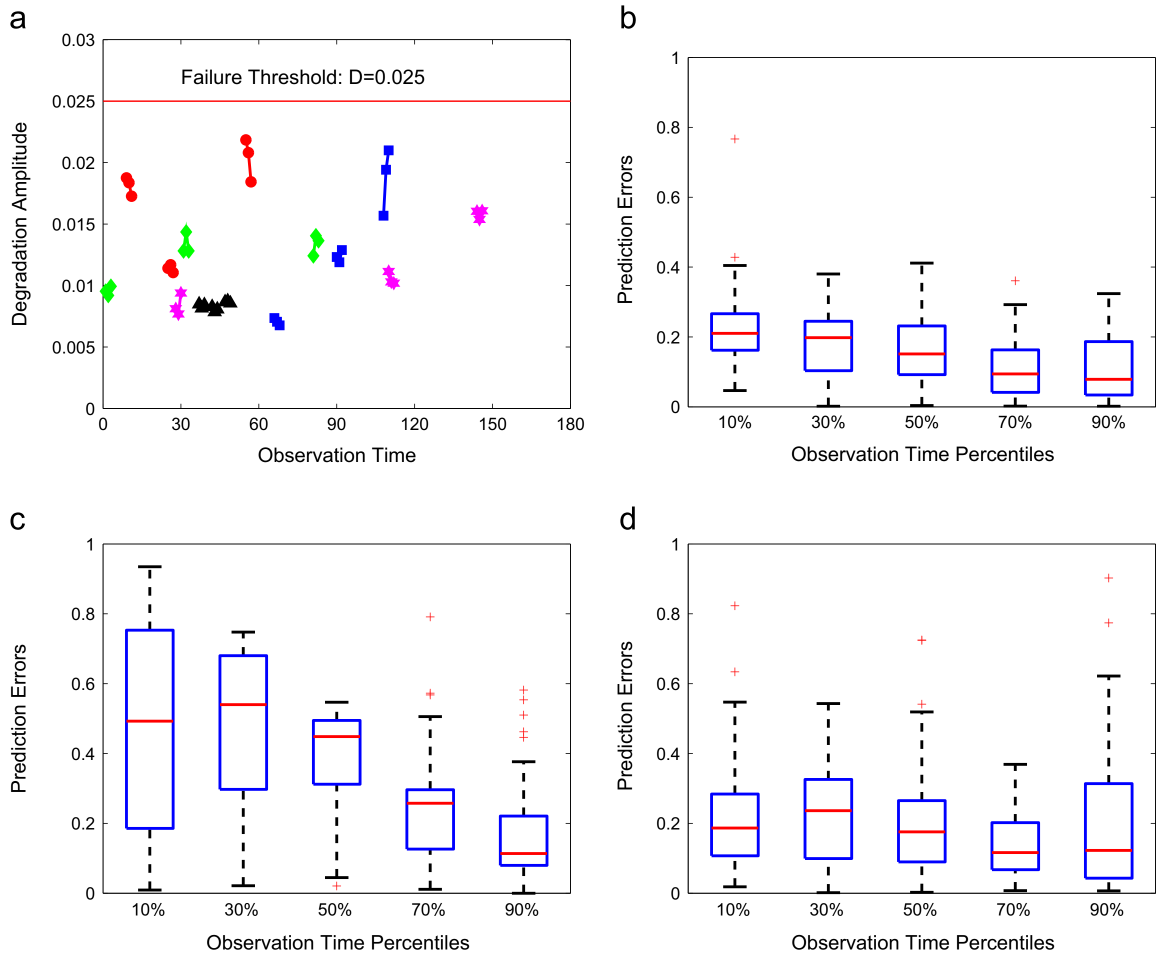

Fig. 7. Plots of prediction errors under the fragmented signals scenario. (a) example of fragmented degradation signals. (b) prediction error for functional regression. (c) prediction error for classical FPCA. (d) prediction error for the axis-transformation FPCA model.

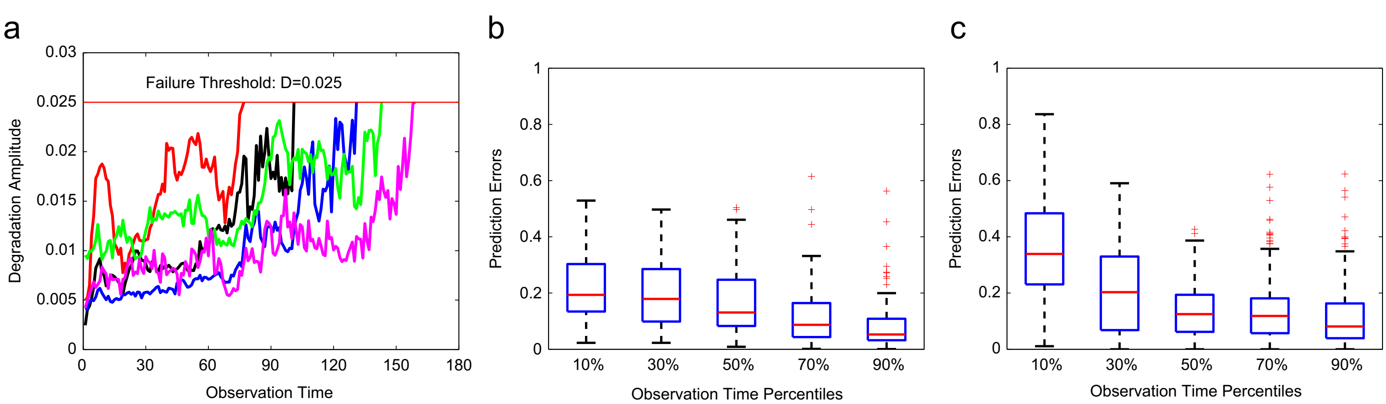

Fig. 8. Plots of prediction errors under complete signals scenario. (a) example of complete degradation signals. (b) prediction error for functional regression. (c) prediction error for exp-Brown.

This is especially beneficial since many parametric models no longer become a viable option due to the data sparsity.

Our methodology was based on using FPCA to identify a general nonparametric trend for degradation signals pertaining to a population of similar components. An adaptive functional regression model was then used to model the relationship between the FPC-scores and the time-to-failure of the components. Real-time signals observed from validation components (assumed to be operating in the field) were incorporated into the model and used to update the predicted time-to-failure of each fielded component based on their unique degradation characteristics. The model was validated using two sets of degradation data, crack growth and bearing vibration data. The performance of the model was benchmarked against other nonparametric and parametric models. The investigation was performed for complete, sparse, and fragmented signal scenarios. Results indicated that the performance of our proposed model was more robust compared and provided relatively failure predictability in comparison to the other benchmarks used in the study. This was particularly true to sparse and fragmented degradation signals. In the case of complete signals that had no missing data, our model performance at least as good as the benchmark parametric model.

The model proposed in this paper is limited to applications with a single failure mode. However, we believe it is possible to extend this modeling framework to encompass multiple failure modes. From the sensing perspective, the model used a single type of sensor observation. This is also a limitation from a practical standpoint as often time multiple sensors are being used to monitor a single component. Consequently, future research is still needed to account for this aspect

Acknowledgments

We would like to thank the associate editor and the referees for their constructive comments. This research was supported by USA National Science Foundation (NSF) Grants CMMI-1200639.

Appendix A. Degradation model estimation

Due to the missing signal observations, the model expressed in Eq. (5) is estimated using the "pooled" historical degradation data. This allows us to borrow information across different components, which improves the estimation process. First, we denote the degradation signal of component i, as $S_i(t_{ij})$, where $j=1,\dots,m_i$, and m_i is the number of observation time points of signal i. (Recall, $i=1,\dots,n$, and n is the number of signals). $\{t_{ij}\}_{j=1,\dots,m_i}$ are sparsely observed times points on the time domain $[0,M]$ for signal i.Using these notation, we arrive at the following form:

$$\begin{aligned} S_i(t_{ij}) &= \mu(t_{ij})+X_i(t_{ij})+\epsilon_i(t_{ij}) \\ &= \mu(t_{ij})+\sum_{k=1}^{K}\xi_{ik}\phi_k(t_{ij})+\epsilon_i(t_{ij}). \end{aligned} \tag{A.1}$$

Next, we illustrate how to estimate the mean and covariance functions, and the FPC-scores.

A.1. Estimating the mean function

The mean function $\mu(t)$ is estimated using a smoothing technique, specifically local regression [37]. In local regression, a smoothing window with a given bandwidth is defined around a point in the domain of the mean function. A smooth function is then approximated in that neighborhood using the available signal observations. We use a local linear approximation of the mean function, i.e, $\mu(t_{ij})\approx c_0+c_1(t_{ij}-t)$. The estimates of the linear coefficients, $\hat{c}_0$ and $\hat{c}_1$ are chosen to minimize the following function:

$$\min_{c_0,c_1}\sum_{i=1}^{n}\sum_{j=1}^{m_i}W\left(\frac{t_{ij}-t}{d_\mu}\right)\{S_i(t_{ij})-c_0-c_1(t-t_{ij})\}^2, \tag{A.2}$$

where t is the objective fitting time point, t_{ij} are the observed time points within the smoothing window, d_μ is the bandwidth of the smoothing window, which is selected by using the one-curve-leave-out cross-validation method (see [21] for additional details), and $W(\cdot)$ is a Gaussian kernel function that assigns more weight to those observations close to point t. $\mu(t)$ is estimated as follows: $\hat{\mu}(t)=\hat{c}_0(t)$.

A.2. Estimating the Covariance Function

The covariance function $C(t,t')$ is estimated using the demeaned signal data $S_i(t_{ij})-\hat{\mu}(t_{ij})$, where $\hat{\mu}(t)$ is the estimated mean function obtained from the previous step. The raw covariance surface is denoted by $G_i(t_{ij},t_{ik})=Cov(S_i(t_{ij})-\hat{\mu}(t_{ij}),S_i(t_{ik})-\hat{\mu}(t_{ik}))$. Recall that $C_i(t_{ij},t_{ik})=Cov(X_i(t_{ij}),X_i(t_{ik}))$, thus we have $G_i(t_{ij},t_{ik})=C_i(t_{ij},t_{ik})+\sigma^2\delta_{t_{ij}t_{ik}}$, where $\delta_{t_{ij}t_{ik}}=1$, if $t_{ij}=t_{jk}$, and 0 otherwise. In order to estimate $C_i(t_{ij},t_{ik})$, we only consider the off-diagonal elements of the raw covariance surface $G_i(t_{ij},t_{ik})$ (since the diagonal elements contain an additional element, the error variance). In other words, only $G_i(t_{ij},t_{ik}), t_{ij}\neq t_{ik}$ are used to estimate $C_i(t_{ij},t_{ik})$. We also use local regression to estimate the covariance function as follows [37]:

$$\begin{aligned} \min_{c_0,c_1,c_2}\sum_{i=1}^{n}\sum_{1\le j\neq k\le m_i}W\left(\frac{t_{ij}-t}{d_G},\frac{t_{ik}-t'}{d_G}\right)\{G_i(t_{ij},t_{ik})-c_0-c_1(t-t_{ij}) \\ -c_2(t'-t_{ik})\}^2, \end{aligned} \tag{A.3}$$

where d_G is the smoothing window bandwidth, which is also selected by using one-curve-leave-out cross-validation, and $W(\cdot)$ is a bivariate Gaussian kernel function.

$C(t,t')$ is estimated as $\hat{C}(t,t')=\hat{c}_0(t,t')$. Recall that $C(t,t')=\sum_{k=1}^{\infty}\lambda_k\phi_k(t)\phi_k(t')$ where the eigenfunctions $\phi_k(t)$ and the eigenvalues λ_k can now be estimated by solving the following eigen equations:

$$\int_{[0,M]}\hat{C}(t,t')\hat{\phi}_k(t)\,dt=\hat{\lambda}_k\hat{\phi}_k(t'), \tag{A.4}$$

where $\int_{[0,M]}\hat{\phi}_k(t)\times\hat{\phi}_m(t)\,dt=1$, if $m=k$, and 0 otherwise. Eq. (A.4) is solved by discretizing the estimated covariance surface $\hat{C}(t,t')$ (details can be found in [21]).

A.3. Estimating the error term

To estimate the error terms, we let $\hat{D}(t)$ represent a smoothed function estimated using the diagonal elements of the raw covariance surface i.e. $C(t,t')+\sigma^2\boldsymbol{I}$. Next, we define $\tilde{C}(t)$ as a smoothed function based on the covariance surface, $C(t,t')$. In other words, $\hat{D}(t)$ is calculated based on the "raw" diagonal elements, and hence it contains the variance of the error term, whereas $\tilde{C}(t)$ is estimated based on covariance matrix with the "smoothed" diagonal elements, and hence so it does not contain the variance of the error term. Using these two quantities, we can estimate σ^2 using the following expression:

$$\hat{\sigma}^2=\frac{1}{|M|}\int_{[0,M]}\left\{\hat{D}(t)-\tilde{C}(t)\right\}dt, \tag{A.5}$$

where $|M|$ denotes the length of $[0,M]$.

$\hat{D}(t)$ is estimated using local linear smoothers similar to those presented in Eq. (A.2) with $G_i(t_{ij},t_{ij})$ as input. Local quadratic smoothers are used to estimate $\tilde{C}(t)$ since they tend to capture the shape of the surface better. Consequently, local quadratic smoothers are used to estimate $\tilde{C}(t)$ by rotating the coordinates by 45° as proposed in [23] (since the covariance surface $C(t,t')$ is maximal along the diagonal as noted by [38]). By rotating the axes by 45° clockwise, we have the following;

$$\begin{pmatrix} t_{ij}^r \\ t_{ik}^r \end{pmatrix}=\begin{pmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ -\sqrt{2}/2 & \sqrt{2}/2 \end{pmatrix}\begin{pmatrix} t_{ij} \\ t_{ik} \end{pmatrix}$$

where (t_{ij},t_{ik}) denote the points in the original axes, and (t_{ij}^r,t_{ik}^r) represent the corresponding points on the rotated axes.

The rotated $C(t,t')$ is estimated using the quadratic local smoothers as follows:

$$\begin{aligned} \min_{c_0,c_1,c_2}\sum_{i=1}^{n}\sum_{1\le j\neq k\le m_i}W\left(\frac{t_{ij}^r-t}{d_G},\frac{t_{ik}^r-t'}{d_G}\right)\{G_i(t_{ij}^r,t_{ik}^r)-c_0-c_1(t-t_{ij}^r) \\ -c_2(t'-t_{ik}^r)^2\}^2, \end{aligned} \tag{A.6}$$

where d_G is the smoothing bandwidth; and $W(\cdot)$ is the bivariate Gaussian kernel function. If we let $\hat{C}^r(t)=\hat{c}_0(t,t')$ be the rotated estimate of $C(t,t')$, then $\tilde{C}(t)=\hat{C}^r(0,t/\sqrt{2})$, and hence, σ^2 can be estimated using Eq. (A.5).

A.4. Estimating the FPC-scores

Typically, with complete degradation signals the FPC-scores are given as $\xi_{ik}=\int_{[0,M]}(X_i(t)-\mu(t))\phi_k(t)\,dt$ and can be estimated by numerical integration using $\hat{\xi}_{ik}=\sum_{j=1}^{m_i}(S_i(t_{ij})-\hat{\mu}(t_{ij}))\hat{\phi}_k(t_{ij})$

$(t_{ij}-t_{i(j-1)})$, where $t_{i0}=0$. However, when dealing with sparse and fragmented signals, numerical integration may not be suitable, especially when the degradation signals are too sparse. Consequently, we use a method proposed by [23] and is known as Principal Analysis by Conditional Expectation (PACE). Asymptotic results reported by the authors demonstrate that the PACE method is well-suited for estimating FPC-scores when repeated measurements are irregularly spaced and the number of observations per subject is limited. To illustrate this method, we begin by defining the following vectors; $\tilde{\mathbf{X}}_i=(X_i(t_{i1}),\ldots,X_i(t_{im_i}))^T$, $\tilde{\mathbf{S}}_i=(S_i(t_{i1}),\ldots,S_i(t_{im_i}))^T$, $\hat{\boldsymbol{\mu}}_i=(\hat{\mu}(t_{i1}),\ldots,\hat{\mu}(t_{im_i}))^T$, and $\hat{\boldsymbol{\phi}}_{ik}=(\hat{\phi}_k(t_{i1}),\ldots,\hat{\phi}_k(t_{im_i}))^T$. Based on the assumption that ξ_{ik} and ϵ_i are jointly Gaussian, the FPC-scores of the ith signal can be estimated using the following conditional expectation:

$$\hat{\xi}_{ik}=\hat{E}[\xi_{ik}|\tilde{\mathbf{S}}_i]=\hat{\lambda}_k\hat{\boldsymbol{\phi}}_{ik}^T\hat{\boldsymbol{\Sigma}}_{\mathbf{S}_i}^{-1}(\tilde{\mathbf{S}}_i-\hat{\boldsymbol{\mu}}_i), \tag{A.7}$$

where $\hat{\boldsymbol{\Sigma}}_{\mathbf{S}_i}=cov(\tilde{\mathbf{S}}_i,\tilde{\mathbf{S}}_i)=cov(\tilde{\mathbf{X}}_i,\tilde{\mathbf{X}}_i)+\hat{\sigma}^2\mathbf{I}_{m_i}$. $\hat{\boldsymbol{\Sigma}}_{\mathbf{S}_i}$ is a $m_i\times m_i$ matrix and its $(j,k)th$ element is $\hat{\boldsymbol{\Sigma}}_{\mathbf{S}_i}(j,k)=\hat{C}(t_{ij},t_{ik})+\hat{\sigma}^2\delta_{t_{ij}t_{ik}}$, where $\delta_{t_{ij}t_{ik}}=1$, if $t_{ij}=t_{jk}$ and 0 otherwise.

References

[1] Lu C, Tao L, Fan H. An intelligent approach to machine component health prognostics by utilizing only truncated histories. Mech Syst Signal Process 2014;42(1):300–13.

[2] Yu M, Wang D. Model-based health monitoring for a vehicle steering system with multiple faults of unknown types. IEEE Trans Ind Electron 2014;61 (7):3574–86.

[3] Yu M, Wang D, Luo M. Model-based prognosis for hybrid systems with mode-dependent degradation behaviors. IEEE Trans Ind Electron 2014;61(1):546–54.

[4] Ramasso E, Gouriveau R. Remaining useful life estimation by classification of predictions based on a neuro-fuzzy system and theory of belief functions. IEEE Trans Reliab 2014;63(2):555–66.

[5] Wu B, Tian Z, Chen M. Condition-based maintenance optimization using neural network-based health condition prediction. Qual Reliab Eng Int 2013;29(8):1151–63.

[6] Herzog MA, Marwala T, Heyns PS. Machine and component residual life estimation through the application of neural networks. Reliab Eng Syst Saf 2009;94(2):479–89.

[7] Gebraeel N, Lawley M, Liu R, Parmeshwaran V. Residual life predictions from vibration-based degradation signals: a neural network approach. IEEE Trans Ind Electron 2004;51(3):694–700.

[8] Cadini F, Zio E, Avram D. Model-based Monte Carlo state estimation for condition-based component replacement. Reliab Eng Syst Saf 2009;94 (3):752–8.

[9] Zio E, Peloni G. Particle filtering prognostic estimation of the remaining useful life of nonlinear components. Reliab Eng Syst Saf 2011;96(3):403–9.

[10] Myötyri E, Pulkkinen U, Simola K. Application of stochastic filtering for lifetime prediction. Reliab Eng Syst Saf 2006;91(2):200–8.

[11] Li M, Meeker WQ. Application of Bayesian methods in reliability data analyses. J Qual Technol 2014;46:1–23.

[12] Peng W, Huang HZ, Xie M, Yang Y, Liu Y. A Bayesian approach for system reliability analysis with multilevel pass-fail, lifetime and degradation data sets. IEEE Trans Reliab 2013;62(3):689–99.

[13] Gebraeel N, Lawley M, Li R, Ryan J. Residual-life distributions from component degradation signals: a Bayesian approach. IIE Trans 2005;37:543–57.

[14] Meeker WQ, Escobar LA. Statistical methods for reliability data. New York: Wiley; 1998.

[15] Si XS, Wang W, Hu CH, Zhou DH. Estimating remaining useful life with three-source variability in degradation modeling. IEEE Trans Reliab 2014;63 (1):167–90.

[16] Wang ZQ, Hu CH, Wang W, Si XS. An additive Wiener process-based prognostic model for hybrid deteriorating systems. IEEE Trans Reliab 2014;63(1):208–22.

[17] Doksum KA, Hoyland A. Models for variable-stress accelerated life testing experiments based on Wiener processes and the inverse Gaussian distribution. Theory Probab Its Appl 1993;37(1):137–9.

[18] Lawless J, Crowder M. Covariates and random effects in a gamma process model with application to degradation and failure. Lifetime Data Anal 2004;10 (3):213–27.

[19] Ramsay JO, Silverman BW. Functional data analysis. New York: Springer; 1997.

[20] Ramsay JO, Silverman BW. Applied functional data analysis: methods and case studies. New York: Springer; 2002.

[21] Rice J, Silverman B. Estimating the mean and covariance structure nonparametrically when the data are curves. J R Stat Soc (Ser B) 1991;53:233–43.

[22] Yao F, Müller HG. Functional quadratic regression. Biometrika 2010;97 (1):49–64.

[23] Yao F, Müller HG, Wang LJ. Functional data analysis for sparse longitudinal data. J Am Stat Assoc 2005;100:577–90.

[24] Müller HG, Zhang Y. Time-varying functional regression for predicting remaining lifetime distributions from longitudinal trajectories. Biometrics 2005;61:1064–75.

[25] Hall P, Müller HG, Yao F. Modelling sparse generalized longitudinal observations with latent Gaussian processes. J R Stat Soc: Ser B (Stat Methodol) 2008;70(4):703–23.

[26] Sentürk D, Nguyen DV. Varying coefficient models for sparse noise-contaminated longitudinal data. Stat Sin 2011;21:1831–56.

[27] Sentürk D, Dalrymple LS, Mohammed SM, Kaysen GA, Nguyen DV. Modeling time-varying effects with generalized and unsynchronized longitudinal data. Stat Med 2013;32(17):2971–87.

[28] James GM, Hastie TJ, Sugar CA. Principal component models for sparse functional data. Biometrika 2000;87(3):587–602.

[29] Wu Y, Liu Y. Functional robust support vector machines for sparse and irregular longitudinal data. J Comput Graph Stat 2013;22(2):379–95.

[30] Zhou RR, Serban N, Gebraeel N. Degradation modeling applied to residual lifetime prediction using functional data analysis. Ann Appl Stat 2011;5 (2B):1586–610.

[31] Zhou RR, Gebraeel N, Serban N. Degradation modeling and monitoring of truncated degradation signals. IIE Trans 2011;44(9):793–803.

[32] Zhou RR, Serban N, Gebraeel N, Müller HG. A functional time warping approach to modeling and monitoring truncated degradation signals. Technometrics 2014;56(1):67–77.

[33] Gervini D. Detecting and handling outlying trajectories in irregularly sampled functional datasets. Ann Appl Stat 2009;3(4):1758–75.

[34] Mercer J. Functions of positive and negative type, and their connection with the theory of integral equations. Philos Trans R Soc Lond Ser A, Contain Pap Math Phys Character 1909;23:415–46.

[35] Virkler DA, Hillberry BM, Goel PK. The statistical nature of fatigue crack propagation. J Eng Mater Technol 1979;101:148–53.

[36] Gebraeel N, Elwany A, Pan J. Residual life predictions in the absence of prior degradation knowledge. IEEE Trans Reliab 2009;58(1):106–17.

[37] Fan J, Gijbels I. Local polynomial modelling and its applications. London: Chapman and Hall; 1996.

[38] Yao F, Müller HG, Clifford AJ, Dueker SR, Follett J, Lin Y, Buchholz BA, Vogel JS. Shrinkage estimation for functional principal component scores with application to the population kinetics of plasma folate. Biometrics 2013;59 (3):676–85.