Remaining Useful Life Prediction Based on Degradation Signals Using Monotonic B-splines with Infinite Support

Salman Jahani JAHANI@WISC.EDU
Department of Industrial and Systems Engineering
University of Wisconsin-Madison
Madison, WI 53706, USA

Raed Kontar ALKONTAR@UMICH.EDU
Department of Industrial and Operations Engineering
University of Michigan
Ann Arbor, MI 48109, USA

Shiyu Zhou SHIYUZHOU@WISC.EDU
Department of Industrial and Systems Engineering
University of Wisconsin-Madison
Madison, WI 53706, USA

Dharmaraj Veeramani RAJ.VEERAMANI@WISC.EDU
Department of Industrial and Systems Engineering
University of Wisconsin-Madison
Madison, WI 53706, USA

Abstract

Degradation modeling traditionally relies on monitoring degradation signals to model the underlying degradation process. In this context, failure is typically defined as the point where the degradation signal reaches a pre-specified threshold level. Many models assume that degradation signals are completely observed beyond the failure threshold, while the issue of truncated degradation signals still remains a challenge. Moreover, based on the physics of degradation process, the degradation signal should be inherently monotonic. However, it is almost inevitable that most of the sensor-based degradation signals are subject to noise which can lead to misleading prediction results. In this paper, a non-parametric approach to modeling and prognosis of degradation signals using B-splines in a mixed effects setting is proposed. In order to deal with the issue of truncated historical degradation signals, our approach is based on augmenting B-spline basis functions with functions of infinite support. Moreover, to model the degradation signal more accurately and robustly in a noisy setting, necessary and sufficient conditions to ensure monotonic evolution of the modeled signals are derived. Appropriate procedures for online updating of random coefficients of mixed effects model considering derived monotonicity constraints based on degradation data collected from an in-service unit are also presented. The performance of the proposed framework is investigated and benchmarked through analysis based on numerical studies and a case study using real-world data from automotive lead-acid batteries.

Keywords: Remaining useful life, Condition monitoring, B-splines, Monotone degradation path.

1. Introduction

Remaining useful life (RUL) estimation is essential in prognostics and health management of a system. RUL can be defined as the remaining life of a component until it is able to function in accordance with its intended purposes. With recent advances in sensor and information technology, growing attention is being paid to RUL prognosis. In general, RUL prognosis models can be categorized into two categories: Model-based prognostics and data-driven prognostics (Vanem, 2018). In model-based prognostics, the RUL prediction of the system is done using physical or mathematical models of the degradation phenomenon. However, in data-driven prognostics, the aim is to transform the data provided by the sensors into relevant models (parametric or non-parametric) of the degradation behavior (Medjaher et al., 2012). In model-based approaches, the prognosis accuracy is highly dependent on the accuracy of the physical model that is used. On the other hand, the independence regarding specific objects in data-driven approaches provides more flexibility which makes them widely used in RUL estimation and degradation modeling (Zhang et al., 2018). This paper deals with a data-driven prognostics approach for the estimation of RUL.

Contemporary data-driven prognostics approaches are mostly based on degradation signals. Degradation signals, also known as condition monitoring (CM) signals, are correlated with the physical degradation of the system and can be used to infer the unobservable underlying health status of the system (Kontar et al., 2017). In degradation data analysis, the component under study is considered failed once its degradation level reaches a certain pre-specified level for the first time (first hitting time). Specifically, most data driven prognostics approaches first start by establishing an initial population-level statistical model for describing the evolution of the degradation signals in the offline stage that provides the prior information. Then, the prediction of signal evolution for a new in-service unit is performed based on both prior information and the newly collected degradation data from the in-service unit to obtain the posterior information.

Most of existing data-driven prognostics methods in literature are based on parametric modeling of the degradation signal evolution. Such parametric models typically assume that the degradation signals behave according to a common functional form, referred to as common shape function (Liao and Köttig, 2014). The most commonly used approach in parametric modeling is based on using mixed effects models that contain a fixed effects part representing the average behavior of the population and a random effects part representing the individual units' characteristics. Lu and Meeker (1993) were among the first who proposed the use of a mixed effects model to characterize the degradation path of a population of units. Based on the general model of Lu and Meeker (1993) and general assumptions summarized by Wang (2000), several mixed effects models have been proposed for RUL estimation (Gebraeel et al., 2005, 2009; Rizopoulos et al., 2014).

Apart from the methods based on the mixed effects models, two other widely-used parametric approaches are Wiener processes and Gamma processes. Wiener processes are typically used for

degradation modelling where the degradation process may vary bi-directionally over time. Tseng and Peng (2004) proposed to model the cumulative degradation path of a product's quality characteristic using Wiener processes. Modeling degradation processes using the Wiener process has some limitations. Firstly, the Wiener processes have the Markov property which means that the evolution of the degradation process at each stage is independent of its past behavior. Also, the Wiener processes are based on the assumption that the mean degradation path is linear. On the other hand, the Gamma process is appropriate for modeling degradation processes that are monotonic and evolving in only one direction (Si et al., 2011; Sun et al., 2018). One limitation of using Gamma processes is that, similar to Wiener processes, they also assume that the degradation process has the independent increment property. Furthermore, the noise in the Gamma process should have a Gamma distribution with a special parameter structure (Son et al., 2016).

The parametric models are typically derived from the underlying physics of the degradation process or empirical evaluation of the degradation signals. However, in many applications it may be difficult to identify the underlying physics and the signal may not follow any functional form (Zhang et al., 2018). Moreover, imposing a parametric form always comes with one caveat that if the specified form is far from the truth the prognostics results are misleading (Kontar et al., 2018). One approach to overcome the challenges of parametric models is to consider non-parametric degradation models in which the functional form is learned from the degradation data. Zhou et al. (2011) used functional principal component analysis (FPCA) to non-parametrically model the degradation process using the historical data and a set of principal component scores. Moreover, an empirical Bayes approach is used in their paper to update the principal component scores of the degradation model in real-time for online monitoring of the components operating in the field. One major weakness of modeling the degradation signals using the FPCA approach is inaccurate estimation of the mean and covariance function under high noise level (Kontar et al., 2018; Zhou et al., 2012). Besides that, in a recent attempt towards non-parametric prognosis of degradation signals, Kontar et al. (2018) proposed using the multivariate Gaussian convolution processes (MGCP) to model the evolution of individual components taking into account the heterogeneity in data. Although their method can effectively model the evolution of degradation signals, it suffers from the issue of scalability, specifically when number of signals is large.

Despite the recent work on non-parametric methods, both the FPCA and MGCP require that data are densely observed over the whole domain of the experiment, i.e., degradation signals have the same support. As mentioned previously, failure is defined as the moment when the degradation signal reaches a certain pre-specified threshold level. In some engineering practices, it is possible to continue observing the degradation signal after it crosses the failure threshold as the failure does not necessarily imply component replacement (Shiau and Lin, 1999; Wang and Xu, 2010). However, in most cases, the system is shut down or replaced immediately once the degradation signal crosses the failure threshold. This means that no further observations can be made once

the unit has failed. Such signals are referred to as truncated degradation signals (Gebraeel et al., 2005; Virkler et al., 1979). Unfortunately, truncation results in wrong estimation of the remaining useful life for the MGCP and FPCA, since we are not able to observe the complete evolution of the historical signals over the whole domain of the experiment. This is intuitively understandable, as prediction for a Gaussian Process (GP) falls back to the mean when predicting far away from observation points (Rasmussen, 2004). Therefore, the prediction may never hit the threshold level in the case of truncation. One exception to this issue is a study by Zhou et al. (2012) where they applied FPCA on the transformed degradation signals. However, their approach cannot guarantee the monotonicity of the CM signal evolution, as the FPCA assumes that the signals indeed follow a GP, which is non-monotone. Also, the transformation approach requires a relatively large number of training units in the historical database.

In this paper, we propose a non-parametric approach to modeling and prognosis of degradation signals using B-spline basis functions in a mixed effects setting. Specifically, our approach is motivated by supplementing the basis with functions of infinite support as proposed in Schumaker (2007) and Corlay (2016). This framework offers numerous advantages. First, B-spline is inherently a non-parametric data-driven approach in which we are allowing the data to speak for themselves. Moreover, by augmenting the basis functions with infinite support we let the B-spline to cover the whole possible range of degradation signal evolution thus allowing to extrapolate the evolution of degradation signal beyond the threshold level. Therefore, it allows us to deal with the issue of truncated signals as mentioned earlier in this paper. Furthermore, it is well-known that the B-splines can be derived as polynomials with a certain degree depending on the degree of generating B-spline basis functions. This feature, further, allows us to control the evolution of the degradation signals by imposing appropriate constraints to ensure monotonicity in the resulting polynomial functions. This, in fact, helps with correct modeling and prognosis of the degradation signal when we are observing noisy data. Finally, it should be mentioned that as opposed to MGCP approach developed in Kontar et al. (2018), the number of parameters that needs to be estimated in the log-likelihood function is fixed and depends only on the degree and number of basis functions. This, indeed, offers a huge advantage over the multivariate Gaussian process based approaches that typically suffer from the scalability issues.

The remainder of this paper is structured as follows. Section 2 reviews degradation signal modeling and conventional Bayesian updating. Moreover, the concept of B-splines with infinite support is described in this section. In section 3, we derive necessary and sufficient conditions to guarantee monotonic evolution of degradation signals in a mixed effects model based on B-splines of infinite support. Moreover, appropriate procedures to conduct online updating based on the collected data from the new in-service unit considering the derived monotonicity constraints are discussed in this section. Section 4 presents numerical studies where the performance of the proposed approach is investigated under different scenarios. In section 5, the application of the

proposed method in a real-world case study is demonstrated. Finally, our concluding remarks are presented in section 6.

2. Degradation signal modeling and prognosis

2.1 Review of B-spline based mixed effects modeling for degradation signals and online updating

The mixed effects model has been widely used in degradation signal modeling and RUL estimation in literature due to its flexible model structure (Gebraeel et al., 2005). The main advantage of the mixed effects model is that it allows each unit to have its own parameters and degradation progression path. The degradation path of the ith unit at each time point t in the mixed effects model is typically defined as:

$$y_i(t) = x_i(t) + \varepsilon_i(t) = \boldsymbol{z}^T(t)\,\boldsymbol{\theta}_i + \varepsilon_i(t), \tag{1}$$

where $x_i(t)$ is the true but unobservable value of the degradation signal, $\varepsilon_i(t)$ is the measurement error which is assumed to be independent and normally distributed $N(0, \sigma^2)$, $\boldsymbol{z}^T(t)$ is a pre-specified time-dependent regression function and $\boldsymbol{\theta}_i = [\theta_{i,1}, \ldots, \theta_{i,d}]^T$ is the vector of random coefficients for unit i. The random coefficients in (1) allow the units to have distinct but similar degradation paths, and it is assumed that $\boldsymbol{\theta}_i$ follows a multivariate normal distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

In cases where the behavior of degradation signals is evident, we may consider the polynomials of appropriate order or other certain specific nonlinear functions of time (e.g., log(t)) as the regression function $\boldsymbol{z}^T(t)$. However, in most of the cases, it happens that identifying a model that best defines the evolution of degradation signals is difficult if not impossible. In this situation, one can use B-spline basis functions of an appropriate degree as a non-parametric regression approach to define $\boldsymbol{z}^T(t)$ in Eq. (1). The value of B-spline basis functions of degree n defined over a set of non-decreasing knot values at time t $\left(\mathbf{b}_n^T(t)\right)$ can be obtained recursively as a function of lower degree basis functions. It is worth mentioning that by properly increasing the number of knots or degree (n), B-splines can approximate arbitrary continuous functions to any given precision (De Boor et al., 1978).

The B-spline based mixed effects model for modeling the degradation path of the ith unit at time t can be defined as follows:

$$y_i(t) = \boldsymbol{b}_n^T(t)\,\boldsymbol{\theta}_i + \varepsilon_i(t), \tag{2}$$

where $\boldsymbol{b}_n(t) = [b_{1,n}(t), \ldots, b_{d,n}(t)]^T$ is the vector of B-spline basis functions evaluated at time t over a set of non-decreasing knot values. In order to use the mixed effects model in (2), first we need to estimate the parameters of the model such as $\boldsymbol{\mu}, \boldsymbol{\Sigma}$ and σ^2 in the offline stage based on the historical dataset of degradation signals. Then, in the online stage, the estimated model is used

as a prior and the Bayesian updating is performed to estimate the model parameters for the new in-service unit based on the degradation data collected from the new unit.

At the current time point t_h, when the prediction is to be made, assume there are s_i values of the degradation signal available from the in-service unit i, i.e. $\boldsymbol{y}_i(\boldsymbol{t}_{i,1:h}) = [\boldsymbol{y}_i(t_{i,1}), \boldsymbol{y}_i(t_{i,2}), ..., \boldsymbol{y}_i(t_{i,s_i})]^T$ and $\boldsymbol{t}_{i,1:h} = [t_{i,1}, t_{i,2}, ..., t_{i,s_i}]^T$ where $t_{i,s_i} \leq t_h$. Therefore, the posterior distribution of the random coefficients based on the newly collected data from the ith in-service unit up to time t_h, can be computed as $p(\boldsymbol{\theta}_i|\boldsymbol{y}_i(\boldsymbol{t}_{1:h})) \propto \ p(\boldsymbol{y}_i(\boldsymbol{t}_{i,1:h})|\boldsymbol{\theta}_i)\pi(\boldsymbol{\theta}_i)$ where $p(\boldsymbol{y}_i(\boldsymbol{t}_{i,1:h})|\boldsymbol{\theta}_i)$ represents the likelihood of the observation data and $\pi(\boldsymbol{\theta}_i)$ refers to the prior distribution estimated in the offline stage. Assuming normally distributed $\boldsymbol{\theta}_i$ and ε_i, the posterior is also normally distributed $p(\boldsymbol{\theta}_i|\boldsymbol{y}_i(\boldsymbol{t}_{i,1:h})) \sim N(\hat{\boldsymbol{\mu}}_{i,h}, \hat{\boldsymbol{\Sigma}}_{i,h})$ where $\hat{\boldsymbol{\mu}}_{i,h}$ and $\hat{\boldsymbol{\Sigma}}_{i,h}$ represent the mean and covariance matrix of the posterior. This, in fact, derives from the conjugate property of the normal distribution. Considering the prior parameter estimates as $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$, the closed form expression for the posterior mean and covariance matrix is as follows:

$$\begin{aligned} \hat{\boldsymbol{\mu}}_{i,h} &= \hat{\boldsymbol{\Sigma}}_{i,h}\left[\frac{\boldsymbol{b}_n^T(\boldsymbol{t}_{i,1:h})\,\boldsymbol{y}_i(\boldsymbol{t}_{i,1:h})}{\sigma^2} + \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0\right], \\ \hat{\boldsymbol{\Sigma}}_{i,h} &= \left[\boldsymbol{\Sigma}_0^{-1} + \left[\frac{\boldsymbol{b}_n^T(\boldsymbol{t}_{i,1:h})\,\boldsymbol{b}_n(\boldsymbol{t}_{i,1:h})}{\sigma^2}\right]^{-1}\right], \end{aligned} \tag{3}$$

where $\boldsymbol{b}_n(\boldsymbol{t}_{i,1:h}) = [\boldsymbol{b}_n(\boldsymbol{t}_{i,1}), \boldsymbol{b}_n(\boldsymbol{t}_{i,2}), \ldots, \boldsymbol{b}_n(\boldsymbol{t}_{i,h})]^T$ and $\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0$ and σ^2 can be replaced by their estimates $\hat{\boldsymbol{\mu}}_0, \hat{\boldsymbol{\Sigma}}_0$ and $\hat{\sigma}^2$ from the offline stage. With the updated mean and covariance matrix, the degradation values of the new in-service unit can be predicted for future time $t > t_h$. This procedure considers both the prior information coming from the average behavior of the population and specific behavior of the in-service unit up to time t_h.

It should be mentioned that the B-spline basis functions are always limited to the range of observed data points and vanish outside this range based on their definition. As a result, modeling the evolution of the degradation signals always depends on the specific range of the historical data over which the B-spline basis functions are defined. However, it often happens that the support range of degradation signals in our historical dataset is limited. In this case, when modeling the evolution of a specific unit, we reach to a state at which we are required to extrapolate outside the range of B-splines. The next subsection introduces infinite support B-splines as a solution to deal with this issue.

2.2 B-splines of infinite support

The classical B-spline basis functions have compact support and can only model data limited to this specific support range. This subsection reviews the extension of classical B-spline to include the basis functions with infinite support as proposed in Schumaker (2007).

Let k be a nonnegative integer and $\Gamma = \{\gamma_0 \leq \gamma_1 \leq ... \leq \gamma_{k-1}\}$ be a sorted collection of k knots, where if $k = 0$ then $\Gamma = \emptyset$. Then the B-spline basis function of degree 0 associated with the knots Γ is defined as follows:

$$\begin{cases} b_{0,0}(t) := I_{(-\infty, inf(\Gamma))}(t), & b_{k,0} := I_{[sup(\Gamma), +\infty)}(t), \\ b_{j,0}(t) := I_{[\gamma_{j-1}, \gamma_j)}(t), & 1 \leq j \leq k-1, \end{cases} \tag{4}$$

where $I_{(.)}$ denotes the indicator function. The B-spline basis functions with higher degree $1 \leq n \leq k$ can be defined using the induction formula given in appendix A. The collection of functions $(b_{j,n})_{0 \leq\ j<k+n+1}$ for $0 \leq n \leq k$ are called the B-spline basis functions of degree n. Appendix A also gives the induction formula for the B-spline basis functions of degree n where $n \geq k$. Figure 1 shows the B-spline basis functions with infinite supports of different degrees defined over $[-4, 4]$ with 6 equally spaced knot values placed at $\Gamma = \{-2.5, -1.5, -0.5, 0.5, 1.5, 2.5\}$. The classical B-spline basis functions in this case are limited to the range between $\gamma_0 = -2.5$ and $\gamma_5 = 2.5$. However, the basis functions with infinite support can extrapolate beyond this range as demonstrated in Figure 1.

One important property of B-spline basis functions is that their derivatives can be decomposed onto B-spline basis functions of lower degree. Therefore, with the same notation as above, the differentiation of B-splines basis functions (for $0 < n \leq\ k$) is as follows:

$$(b_{j,n})' = \begin{cases} -\frac{n}{C_0} b_{j,n-1} & 0 \leq j < min(n,k), \\ -\frac{n}{\gamma_n - \gamma_0} b_{n,n-1} & \text{for } j = n, \\ \frac{n}{\gamma_{j-1} - \gamma_{j-n-1}} b_{j-1,n-1} - \frac{n}{\gamma_j - \gamma_{j-n}} b_{j,n-1} & n+1 \leq j < k, \\ \frac{n}{\gamma_{k-1} - \gamma_{k-n-1}} b_{k-1,n-1} & \text{for } j = k, \\ \frac{n}{C_1} b_{j-1,n-1} & max(n,k)+1 \leq j < k+n+1, \end{cases} \tag{5}$$

where for the constants C_0 and C_1, we use the guideline proposed in Corlay (2016). In this study, we take $C_0 = C_1 = \frac{\gamma_{k-1} - \gamma_0}{k-1}$ if $\gamma_{k-1} > \gamma_0$ and $C_0 = C_1 = 1$ otherwise. By iterating over this decomposition, the mth derivative of B-spline basis function of degree n onto the basis of degree $n - m$ can be obtained. More information about the derivatives of B-splines and algorithms to evaluate and calculate them can be found in Schumaker (2007) and Corlay (2016).

A critical stage in evaluating B-splines is the placement of knot values which can be generally done using bisecting method with $O(log(k))$ complexity (Tjahjowidodo et al., 2017; Corlay, 2016; Aguilar et al., 2018). However the case of equally spaced knots, which is utilized in this paper, leads to further simplifications where all bounded spline basis functions have the same polynomial representation up to a parallel shift, and unbounded basis functions are symmetric. These properties can be exploited to save a significant amount of memory and computation. Regarding selecting the number of knots, we would like to also mention that it can be simply done using cross validation

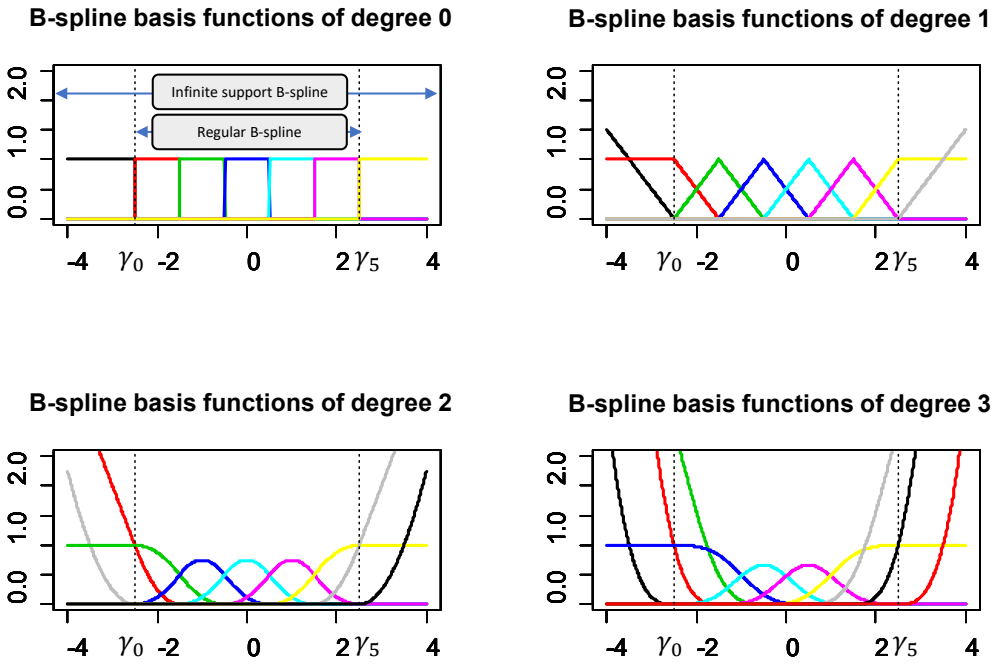

Figure 1: B-spline basis functions of infinite support with degrees 0, 1, 2 and 3.

or any of the well-established versions of statistical information criteria like Akaike Information Criteria (AIC) or Bayesian Information Criteria (BIC).

In this paper, we propose to use a non-parametric mixed effects model based on B-spline basis functions of infinite support as the regression function in (2). The advantage of using the infinite support B-spline is that, firstly, it is data-driven and does not assume the data to follow a specific functional form. Moreover, by properly constraining the random coefficients of B-spline basis functions in the mixed effects setting, one can ensure the monotonic degradation path of each unit as suggested by our domain knowledge. The next section discusses the development of such constraints and appropriate procedures for online updating of the infinite support B-spline based mixed effects model.

3. Remaining useful life prediction using monotonic B-splines with infinite support

In this section, we first derive the monotonicity conditions for infinite support B-Splines of degree 1 and degree 2 in a mixed effects setting. Then, the necessary and sufficient conditions for monotonic degradation of units in a mixed effects model with degree 3 B-spline basis function of infinite support will be derived. Finally, the appropriate procedures to conduct online updating considering these conditions will be discussed in section 3.2.

3.1 Monotonicity conditions for B-splines with infinite support

In degradation modeling, practitioners typically use B-splines to model degradation signals because of their flexibility in consistently approximating the evolution of signals based on collected data. However, in cases where we are collecting data from sensors, most often it happens that the sensor data are noisy and as a result, the collected data may contradict the underlying physics. In degradation modeling, for example, the degradation signal should be monotonically changing over time since the underlying health condition is always deteriorating unless some maintenance actions have been performed (Son et al., 2016; Byon et al., 2010; Zhou et al., 2012). In cases where the noise level of the collected data is high, non-parametric approaches like B-splines may lead to incorrect estimation of the degradation signal. To overcome this potential shortcoming, the ideal degradation modeling approach is the one that allows for flexibility while maintaining the monotonic evolution path of the degradation signal. This section discusses the required conditions for modeling of monotonic degradation signals using B-splines of infinite support in a mixed effects setting.

First, we derive the necessary and sufficient conditions for monotonicity of degree 1 and degree 2 B-splines with infinite support in a mixed effects model. As mentioned in section 2.2, the derivatives of B-spline basis functions are explicitly decomposed onto the basis of lower degree. Thus, non-negativity constraints on the first derivative of B-splines with infinite support translates into monotonicity constraints. The B-spline basis functions of degree n $(n \leq k)$ associated with the knots $\Gamma = \{\gamma_0 \leq \gamma_1 \leq \ldots \leq \gamma_{k-1}\}$ and given by the induction formula in appendix A are a collection of non-negative functions. Therefore, non-negativity of coefficients of basis functions in the mixed effects model is a sufficient condition for non-negativity of B-splines with infinite support. For B-spline of degree 0 and degree 1, this also happens to be a necessary condition (Yuan et al., 2017; Beliakov, 2002). Thus, for basis functions of degree 1 and degree 2 which decompose onto degree 0 and degree 1 basis functions, one can obtain the monotonicity condition using Eq. (5). The following lemma discusses the monotonicity constraints for degree 1 and degree 2 B-splines with infinite support.

Lemma 1 *Let* $\boldsymbol{\theta}_i = [\theta_{i,1}, \ldots, \theta_{i,n+k+1}]^T$ *be the coefficients of infinite support basis functions in the mixed effects model of Eq.* (2) *for unit* i *at time* t. *Let* $\Gamma = \{\gamma_0 \leq \gamma_1 \leq \ldots \leq \gamma_{k-1}\}$ *where* $k > n$ *be the set of non-decreasing knot values over which the infinite support B-spline is defined. Then, the necessary and sufficient condition for monotonicity of degree 1 and degree 2 B-spline with infinite support can be encoded in terms of the following constraints:*

$$\begin{aligned} \theta_{i,j} &\leq 0, \quad j = 1, \ldots, n, \\ \theta_{i,j} - \theta_{i,j-1} &\geq 0, \quad j = n+2, \ldots, k+1, \\ \theta_{i,j} &\geq 0, \quad j = k+2, \ldots, k+n+1. \end{aligned} \tag{6}$$

Proof. The B-spline of degree n with infinite support over the knot vector Γ is defined as follows:

$$Q_i^{(n)}(t) = \boldsymbol{b}_n^T(t)\,\boldsymbol{\theta}_i = \sum_{j=0}^{k+n} \theta_{i,j+1} b_{j,n}(t). \tag{7}$$

The first order derivative of B-spline in (7) can be calculated using the differentiation of B-spline basis functions given in (5) as follows:

$$\begin{aligned}
Q_i'^{(n)}(t) &= \sum_{j=0}^{k+n} \left(\theta_{i,j+1} b_{j,n}(t)\right)' \\
&= \sum_{j=0}^{n-1} \theta_{i,j+1} \left(\frac{-n}{C_0} b_{j,n-1}(t)\right) + \theta_{i,n+1} \left(\frac{-n}{\gamma_n - \gamma_0} b_{n,n-1}(t)\right) \\
&+ \sum_{j=n+1}^{k-1} \theta_{i,j+1} \left(\frac{n}{\gamma_{j-1} - \gamma_{j-n-1}} b_{j-1,n-1}(t) - \frac{n}{\gamma_j - \gamma_{j-n}} b_{j,n-1}(t)\right) \\
&+ \theta_{i,k+1} \left(\frac{n}{\gamma_{k-1} - \gamma_{k-n-1}} b_{k-1,n-1}(t)\right) + \sum_{j=k+1}^{k+n} \theta_{i,j+1} \left(\frac{n}{C_1} b_{j-1,n-1}(t)\right) \\
&= \sum_{j=0}^{n-1} \left(\frac{-n}{C_0} \theta_{i,j+1}\right) b_{j,n-1}(t) + \sum_{j=n+1}^{k} \left(\frac{n}{\gamma_{j-1} - \gamma_{j-n-1}} \left(\theta_{i,j+1} - \theta_{i,j}\right)\right) b_{j-1,n-1}(t) \\
&+ \sum_{j=k+1}^{k+n} \left(\frac{n}{C_1} \theta_{i,j+1}\right) b_{j-1,n-1}(t),
\end{aligned} \tag{8}$$

where, in the last equality, we rearrange the terms for basis functions. The monotonicity condition can be derived by imposing a non-negativity constraint on the first order derivative of B-splines with infinite support. The first order derivative of B-splines of degree 1 and degree 2 can be calculated in terms of basis functions of degree 0 and degree 1, respectively, as given in (8). For basis functions of degree 0 and degree 1, non-negativity of coefficients is both a necessary and sufficient condition for non-negativity of B-spline. Therefore, imposing non-negativity constraints on coefficients of basis functions in the last equality of Eq. (8) for degree 1 and degree 2 B-splines results in the conditions given in (6). As for the necessity of these conditions, it should be noted that the basis functions $b_{j-1,n-1}$ and $b_{j,n-1}$ are always non-negative. Moreover for a degree 2 (or 1) B-spline, each point in a knot interval has non-zero values in 3 (or 2) nearby basis functions (Please refer to Figure 1). Therefore if the conditions in (6) for the corresponding coefficients is not satisfied, the derivative of B-spline in that interval would be negative and as a result the B-spline function is not monotonically increasing. This happens for degree 2 and 1 B-splines simply because their derivatives can be written in forms of polynomials of degree 1 and 0. Thus the condition in (6) is both necessary and sufficient and the proof is complete. □

Using Lemma 1, monotonicity constraints for degree 1 and degree 2 B-splines with infinite support can be written in form of linear inequality constraints for each unit i as follows:

$$\boldsymbol{a} \leq \boldsymbol{A}\boldsymbol{\theta}_i \leq \boldsymbol{b}, \tag{9}$$

where

$$\begin{aligned}
\boldsymbol{A}_{d\times d} &= \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} & \boldsymbol{A}_{13} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} & \boldsymbol{A}_{23} \\ \boldsymbol{A}_{31} & \boldsymbol{A}_{32} & \boldsymbol{A}_{33} \end{bmatrix}, \\
\boldsymbol{A}_{11} &= \boldsymbol{A}_{33} = \boldsymbol{I}_{n\times n}, \\
\boldsymbol{A}_{13} &= \boldsymbol{A}_{31} = \boldsymbol{0}_{n\times n}, \\
\boldsymbol{A}_{21} &= \boldsymbol{A}_{23} = \boldsymbol{0}_{(k-n+1)\times n}, \\
\boldsymbol{A}_{12} &= \boldsymbol{A}_{32} = \boldsymbol{0}_{n\times(k-n+1)}, \\
\boldsymbol{A}_{22} &= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \dots & -1 & 1 \end{bmatrix},
\end{aligned} \tag{10}$$

$\boldsymbol{A}_{22}$ is a $(k-n+1)\times(k-n+1)$ matrix and $d = n+k+1$ is the number of basis functions. For $\boldsymbol{a}_{(k+n+1)\times 1}$ we have $\boldsymbol{a} = \left[\boldsymbol{a}_1^T, \boldsymbol{a}_2^T, \boldsymbol{a}_3^T\right]^T$ where $(\boldsymbol{a}_1)_{n\times 1} = [-\infty, \dots, -\infty]^T$, $(\boldsymbol{a}_2)_{(k-n+1)\times 1} = [-\infty, 0, \dots, 0]^T$ and $(\boldsymbol{a}_3)_{n\times 1} = [0, \dots, 0]^T$. Also, for $\boldsymbol{b}_{(k+n+1)\times 1}$ we have $\boldsymbol{b} = \left[\boldsymbol{b}_1^T, \boldsymbol{b}_2^T, \boldsymbol{b}_3^T\right]^T$ where $(\boldsymbol{b}_1)_{n\times 1} = [0, \dots, 0]^T$, $(\boldsymbol{b}_2)_{(k-n+1)\times 1} = [+\infty, \dots, +\infty]^T$ and $(\boldsymbol{b}_3)_{n\times 1} = [+\infty, \dots, +\infty]^T$.

The conditions in (9) should always be considered when updating the posterior distribution for the in-service unit in online stage in order to get a monotone degradation path. Unfortunately, sole application of the Bayesian updating procedure reviewed in section 2.1 cannot guarantee these conditions in updating the posterior distribution. In subsection 3.2.1, we introduce an approach for updating the posterior distribution of new in-service unit considering the linear inequality constraints in (9) to ensure monotonicity of the resulting modeled degradation signal in a mixed effects models based on degree 1 and degree 2 B-splines with infinite support.

B-splines of degree 2 are typically sufficient for modeling a wide variety of degradation signal evolution paths. Moreover, adding more knot values increases the accuracy of degree 2 B-splines. However, B-splines of degree 3 are also commonly used in literature due to their flexibility and ability to model curvature(Schumaker, 2007; Yuan et al., 2017). Thus, it is of prime importance to develop necessary and sufficient conditions to ensure monotonic evolution of the degradation signal using degree 3 B-splines of infinite support. In practice, B-splines of degree less than 4 are

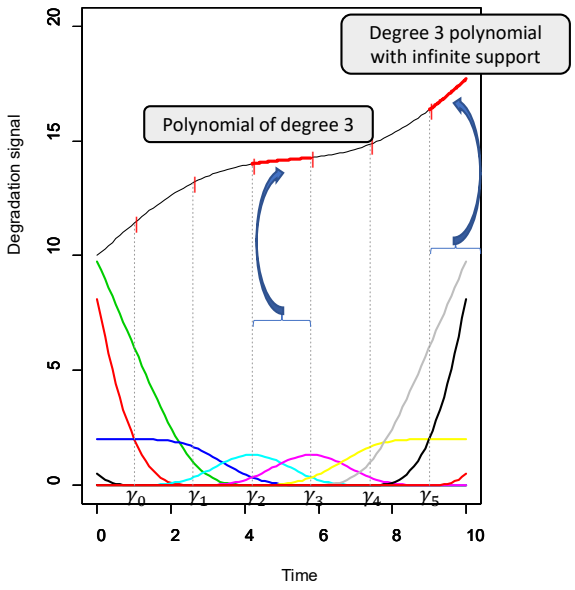

Figure 2: Piecewise polynomial construction of degradation signals using degree 3 B-splines with infinite support.

sufficient for modeling different continuous functions and higher degrees of the B-splines typically increases computational complexity and numerical instability (Schumaker, 2007).

For B-splines of degree 3 with infinite support, slightly more complex necessary and sufficient conditions are required. In this regard, we take advantage of the piecewise representation of B-splines. De Boor et al. (1978) and Schumaker (2007) state that for $t \in [\gamma_p, \gamma_{p+1}]$ where γ_p and γ_{p+1} are two consecutive knot values, the degree 3 B-spline basis function can be expressed as follows:

$$b_{j,3}(t) = \alpha_{0,p,j} + \alpha_{1,p,j} t + \alpha_{2,p,j} t^2 + \alpha_{3,p,j} t^3, \tag{11}$$

where $\alpha_{r,p,j}, r = 0, ..., 3$ are coefficients determined only by placement of the knots. Because of the local support property of B-splines, $b_{j,3}(t)$ are nonzero only when $j = p-3, \ldots, p$. Therefore, for B-spline of degree 3 with infinite support and $t \in [\gamma_p, \gamma_{p+1}]$ we have that:

$$\begin{aligned} Q_i^{(3)}(t) = \boldsymbol{b}_3^T(t)\,\boldsymbol{\theta}_i = \sum_{j=0}^{k+3} \theta_{i,j+1} b_{j,3}(t) = \sum_{j=p-3}^{p} \theta_{i,j+1} \sum_{r=0}^{3} \alpha_{r,p,j} t^r \\ = \sum_{r=0}^{3} \sum_{j=p-3}^{p} \left(\theta_{i,j+1} \alpha_{r,p,j}\right) t^r = \sum_{r=0}^{3} \eta_{r,p} t^r. \end{aligned} \tag{12}$$

Equation (12) shows that degree 3 B-splines model the degradation signal in each interval $[\gamma_p, \gamma_{p+1}]$ using polynomials of degree 3. Figure 2 illustrates the piecewise polynomial construction of a degradation signal using degree 3 B-spline basis functions with infinite support. Monotonicity condition for a degree 3 B-splines can be obtained by imposing monotonicity on $Q_i^{(3)}$ in each knot

interval. The following lemma gives the required conditions for monotonicity of a degree 3 B-spline with infinite support

Lemma 2 *Let* $\Gamma = \{\gamma_0 \leq \gamma_1 \leq \ldots \leq \gamma_{k-1}\}$ *where* $k > 3$ *be the set of non-decreasing knot values over which the infinite support B-spline is defined. Let* $\eta_{r,p}, r = 0, ..., 3$ *be the coefficients of derived degree 3 polynomial for* $t \in [\gamma_p, \gamma_{p+1}]$ *as defined in* (12)*. Then, the necessary and sufficient conditions for monotonicity of degree 3 B-spline with infinite support can be encoded in terms of the following constraints:*

$$\begin{gathered} I\left(3\eta_{3,p}\gamma_{p-1} \leq -\eta_{2,p} \leq 3\eta_{3,p}\gamma_p, \eta_{3,p} \geq 0\right)\left(\eta_{1,p} - \frac{\eta_{2,p}^2}{3\eta_{3,p}}\right) \geq 0, \quad p = 1, \ldots, k-1, \\ \eta_{1,p} + 2\eta_{2,p}\gamma_{p-1} + 3\eta_{3,p}\gamma_{p-1}^2 \geq 0, \quad p = 1, \ldots, k, \\ I\left(3\eta_{3,k}\gamma_{k-1} \leq -\eta_{2,k}, \eta_{3,k} \geq 0\right)\left(\eta_{1,k} - \frac{\eta_{2,k}^2}{3\eta_{3,k}}\right) \geq 0, \\ \eta_{1,k} + 2\eta_{2,k}\gamma_k + 3\eta_{3,k}\gamma_k^2 \geq 0. \end{gathered} \tag{13}$$

Proof. In order to obtain necessary and sufficient conditions for monotonicity of $Q_i^{(3)}(t)$, we need to ensure monotonicity in each interval $t \in [\gamma_p, \gamma_{p+1}]$. In this regard, one needs to just impose monotonicity on a degree 3 polynomial in each interval $t \in [\gamma_p, \gamma_{p+1}]$ according to Eq. (12). The polynomial $Q_i^{(3)}(t) = \sum_{r=0}^{3} \eta_{r,p} t^r$ is monotonically increasing on $t \in [\gamma_p, \gamma_{p+1}]$ if we force its derivative $Q_i'^{(3)}(t) = \eta_{1,p} + 2\eta_{2,p}t + 3\eta_{3,p}t^2$ to be positive for $t \in [\gamma_p, \gamma_{p+1}]$. It should be noted that $Q_i'^{(3)}(t)$ is minimized at $t = -\eta_{2,p}/(3\eta_{3,p})$ and the minimum value is $\eta_{1,p} - \eta_{2,p}^2/(3\eta_{3,p})$ if $\eta_{3,p} > 0$. Thus, the following three conditions ensure monotonicity of degree 3 B-splines with infinite support for $t \in [\gamma_p, \gamma_{p+1}]$.

$$\begin{gathered} I\left(3\eta_{3,p}\gamma_p \leq -\eta_{2,p} \leq 3\eta_{3,p}\gamma_{p+1}, \eta_{3,p} \geq 0\right)\left(\eta_{1,p} - \frac{\eta_{2,p}^2}{3\eta_{3,p}}\right) \geq 0, \\ Q_i'^{(3)}(\gamma_p) \geq 0, \quad Q_i'^{(3)}(\gamma_{p+1}) \geq 0. \end{gathered} \tag{14}$$

In order to obtain necessary and sufficient conditions for monotonicity of $Q_i^{(3)}(t)$, by local support property of B-splines, we only need to ensure monotonicity in each interval of the knot vector. We also need to add one dummy knot value at the end of the knot vector to determine the extent of extrapolation that we are interested in. This knot value is not required to be inside the range of the observed data point as with the regular B-splines. We denote this knot value by γ_k. This results in the conditioned given in (13). □

The posterior distribution of the in-service unit should always satisfy the conditions of Lemma 2 in the online stage to obtain monotone evolution of a degradation signal when using degree 3 B-splines of infinite support in a mixed effect model. *The constraints developed in* (13) *for online updating of the mixed effects model based on degree 3 B-splines with infinite support are non-linear,*

unlike the ones derived for the case of degree 1 and degree 2 B-splines with infinite support in (9). Subsection 3.2.2 introduces a procedure to update the posterior distribution of a new in-service unit considering the general form of non-linear inequality constraints derived in Lemma 2 to ensure monotonic evolution of the resulting modeled degradation signal in the mixed effects model based on degree 3 B-splines of infinite support.

3.2 Online updating of the mixed effects model based on the B-splines of infinite support with monotonicity constraints

This section discusses appropriate procedures to realize the necessary and sufficient conditions developed previously for online updating of the mixed effects model based on monotonic B-splines of infinite support. Specifically in subsection 3.2.1, we first introduce the constraint Kalman filter (CKF) approach tailored for updating the random coefficients of infinite support basis functions of degree 1 and degree 2 in mixed effects model considering the constraints given in Lemma 1. The CKF procedure is applicable for updating random coefficients considering linear inequality constraints. However, by Lemma 2, the monotonicity conditions for degree 3 B-splines with infinite support are in the form of nonlinear inequality constraints, and the CKF procedure is not applicable anymore. In order to deal with this issue, our approach to online updating of degree 3 B-splines is based on Monte Carlo (MC) Sampling. The detailed procedure for updating based on the MC sampling approach is discussed in subsection 3.2.2.

3.2.1 Analytical updating procedure based on the constrained Kalman filter for degree 1 and degree 2 B-splines with infinite support

The online Bayesian updating procedure based on collected data from new in-service unit i at time t_h results in a multivariate normal distribution denoted as $p\left(\boldsymbol{\theta}_i|\boldsymbol{y}_i\left(\boldsymbol{t}_{i,1:h}\right)\right) \sim N(\hat{\boldsymbol{\mu}}_{i,h}, \hat{\boldsymbol{\Sigma}}_{i,h})$ for random coefficients of a mixed effects model as discussed in section 2.1. In order to get a monotonic degradation signal for the new in-service unit in a mixed effects model based on degree 1 and degree 2 B-splines with infinite support, the posterior distribution should always satisfy the conditions provided by Lemma 1. This, indeed, results in updating the posterior multivariate normal distribution with linear inequality constraint given by Lemma 1. Enforcing the set of constraints given in (9) yields the truncated multivariate normal distribution denoetd as $LN\left(\hat{\boldsymbol{\mu}}_{i,h}, \hat{\boldsymbol{\Sigma}}_{i,h}; \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{A}\right)$ with the PDF as follows:

$$p\left(\boldsymbol{\theta}_i|\boldsymbol{y}_i\left(\boldsymbol{t}_{i,1:h}\right); \boldsymbol{\theta}_i \in \mathcal{C}_h\right) = \begin{cases} \frac{p\left(\boldsymbol{\theta}_i|\boldsymbol{y}_i\left(\boldsymbol{t}_{i,1:h}\right)\right)}{\int_{\mathcal{C}_h} p\left(\boldsymbol{\theta}_i|\boldsymbol{y}_i\left(\boldsymbol{t}_{i,1:h}\right)\right) d\boldsymbol{\theta}_i} = \xi_h^{-1} p\left(\boldsymbol{\theta}_i|\boldsymbol{y}_i\left(\boldsymbol{t}_{i,1:h}\right)\right) & \text{if } \boldsymbol{\theta}_i \in \mathcal{C}_h, \\ 0 & \text{otherwise,} \end{cases} \tag{15}$$

where ξ_h is a normalizing constant and $\mathcal{C}_h$ denotes the constraints' space.

First, in Lemma 3 we discuss transforming the multivariate normal distribution with linear inequality constraints ($\boldsymbol{\theta} \sim LN\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{A}\right)$) into a simpler form of a truncated multivariate nor-

mal distribution with box constraints denoted by $TN(\boldsymbol{\mu}^n, \boldsymbol{\Sigma}^n; \boldsymbol{a}, \boldsymbol{b})$ where $\boldsymbol{\mu}^n$ and $\boldsymbol{\Sigma}^n$ denote the new mean vector and covariance matrix after transformation.

Lemma 3 *Let* $\boldsymbol{\theta} \sim LN(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{A})$ *and* $\boldsymbol{Z} = \boldsymbol{A\theta}$. *Then* $\boldsymbol{Z} \sim TN(\boldsymbol{A\mu}, \boldsymbol{A\Sigma A}^T; \boldsymbol{a}, \boldsymbol{b})$ *and If* $\boldsymbol{A}^{-1}$ *exists, then* $\boldsymbol{\theta} = \boldsymbol{A}^{-1}\boldsymbol{Z}$.

Proof. We know $\boldsymbol{\theta} \sim LN(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{A})$ has PDF as follows:

$$C_L^{-1} exp\left\{-(\boldsymbol{\theta}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}-\boldsymbol{\mu})/2\right\} I_{(\boldsymbol{a}\leq \boldsymbol{A\theta}\leq \boldsymbol{b})}, \tag{16}$$

where C_L is the normalizing constant. Let us consider $\boldsymbol{Z} = \boldsymbol{A\theta}$. Then, according to the linear transformation property of multivariate normal distribution, the first part of Eq.(16) must change into the PDF of multivariate normal distribution $N\left(\boldsymbol{A\mu}, \boldsymbol{A\Sigma A}^T\right)$ for $\boldsymbol{Z}$. Thus, $\boldsymbol{Z}$ has the PDF as follows:

$$C_T^{-1} exp\left\{-(\boldsymbol{Z}-\boldsymbol{A\mu})^T \left(\boldsymbol{A\Sigma A}^T\right)^{-1}(\boldsymbol{Z}-\boldsymbol{A\mu})/2\right\} I_{(\boldsymbol{a}\leq \boldsymbol{Z}\leq \boldsymbol{b})}, \tag{17}$$

where C_T is the new normalizing constant after transformation. Equation (17) shows the PDF of $TN(\boldsymbol{A\mu}, \boldsymbol{A\Sigma A}^T; \boldsymbol{a}, \boldsymbol{b})$. Moreover, following the same steps as above, if $\boldsymbol{A}^{-1}$ exists, then $\boldsymbol{\theta} = \boldsymbol{A}^{-1}\boldsymbol{Z}$, and the proof is complete. □

The monotonicity conditions of degree 1 and degree 2 infinite support B-splines can be analytically imposed using Lemma 3. In this regard, online updating based on the linear inequality constraints of Lemma 1 reduces to a simpler form of box constraints on the coefficients of the mixed effects model. Thus, if we had an approach to update the random coefficients of B-spline basis functions in a box constrained multivariate normal distribution, then we could convert back the updated coefficients according to Lemma 3 into the corresponding linear inequality constrained multivariate normal distribution. Our approach for updating this simpler form of constraints is based on the constrained Kalman filter proposed in Wilhelm et al. (2012) and Son et al. (2016).

The online Bayesian updating procedure discussed in section 2.1 eventually gives $\hat{\boldsymbol{\mu}}_{i,h}$ and $\hat{\boldsymbol{\Sigma}}_{i,h}$ where $\hat{\boldsymbol{\mu}}_{i,h} \in \mathbb{R}^d$. The true individual parameter $\boldsymbol{\theta}_i$ for a specific unit is unobservable and should be estimated by the posterior mean $\hat{\boldsymbol{\mu}}_{i,h}$. However applying Lemma 1 and Lemma 3 respectively, suggest that the transformed vector of random coefficients should satisfy a specific set of box constraints. Thus, the posterior distribution of this new transformed vector of random coefficients should be truncated at certain boundaries. This can be achieved through a PDF truncation step in addition to the Bayesian updating step.

To be more specific, suppose that we transform the truncated posterior multivariate normal distribution with linear inequality constraints of Lemma 1 for unit i at time t_h $\left(LN\left(\hat{\boldsymbol{\mu}}_{i,h}, \hat{\boldsymbol{\Sigma}}_{i,h}; \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{A}\right)\right)$ into the new simpler truncated multivariate normal distribution with box constraints denoted by $TN\left(\hat{\boldsymbol{\mu}}^n_{i,h}, \hat{\boldsymbol{\Sigma}}^n_{i,h}; \boldsymbol{a}, \boldsymbol{b}\right)$ using Lemma 3 where $\hat{\boldsymbol{\mu}}^n_{i,h} = \boldsymbol{A}\hat{\boldsymbol{\mu}}_{i,h}$ and $\hat{\boldsymbol{\Sigma}}^n_{i,h} = \boldsymbol{A}\hat{\boldsymbol{\Sigma}}_{i,h}\boldsymbol{A}^T$ and $\boldsymbol{A}, \boldsymbol{a}$ and $\boldsymbol{b}$ are as given in (9). Moreover, we assume that $\boldsymbol{Z}_i = \boldsymbol{A\theta}_i$ is the new vector of coefficients for unit i after

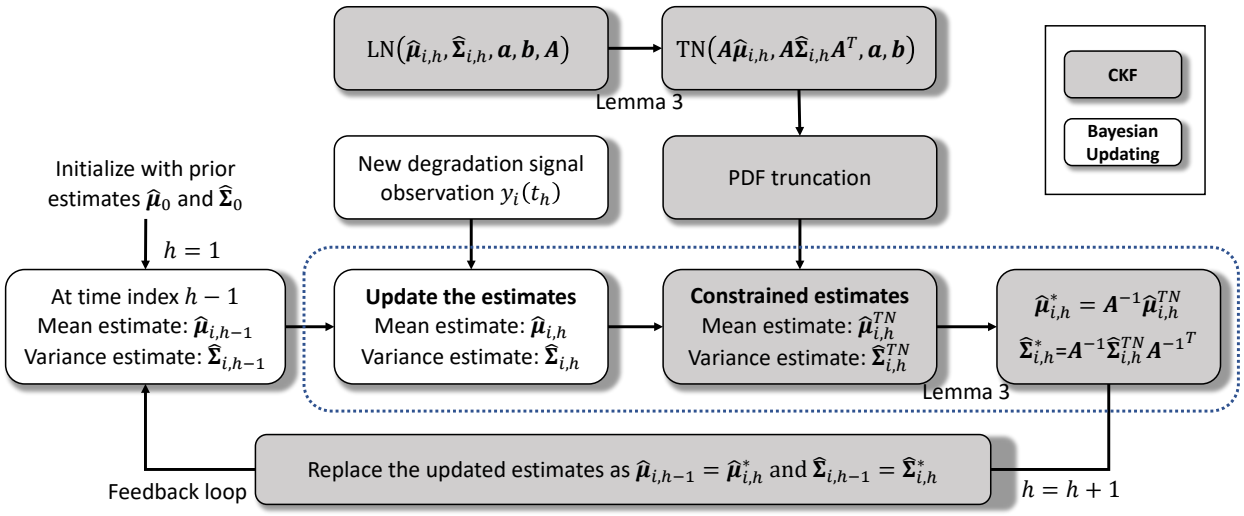

Figure 3: Summary of CKF procedure for online updating

transformation. The PDF for $TN\left(\hat{\boldsymbol{\mu}}_{i,h}^{n}, \hat{\boldsymbol{\Sigma}}_{i,h}^{n}; \boldsymbol{a}, \boldsymbol{b}\right)$ can be defined as follows:

$$f_T\left(\boldsymbol{Z}_i|\boldsymbol{y}_i\left(\boldsymbol{t}_{i,1:h}\right); \boldsymbol{a}, \boldsymbol{b}\right) = \frac{f\left(\boldsymbol{Z}_i|\boldsymbol{y}_i\left(\boldsymbol{t}_{i,1:h}\right)\right)}{F\left(\boldsymbol{b}|\boldsymbol{y}_i\left(\boldsymbol{t}_{i,1:h}\right)\right) - F\left(\boldsymbol{a}|\boldsymbol{y}_i\left(\boldsymbol{t}_{i,1:h}\right)\right)}, \quad \text{for} \quad a_j \leq z_{ij} \leq b_j, \tag{18}$$

where $f_T\left(\boldsymbol{Z}_i|\boldsymbol{y}_i\left(\boldsymbol{t}_{i,1:h}\right); \boldsymbol{a}, \boldsymbol{b}\right) = 0$ elsewhere, $f_T\left(\cdot\right)$ indicates the truncated density function of $\boldsymbol{Z}_i$ with box constraints and $f\left(\cdot\right)$ and $F\left(\cdot\right)$ denotes the probability density and cumulative density functions of multivariate normal distribution. After truncating the density, the mean $\hat{\boldsymbol{\mu}}_{i,h}^{TN} = \left[\hat{\mu}_{i,h,1}^{TN}, \ldots, \hat{\mu}_{i,h,d}^{TN}\right]^T$ and the covariance matrix $\widehat{Cov}_{i,h}^{TN}\left[z_{i,j}, z_{i,l}\right]$ for $j, l \in \{1, \ldots, d\}$ can be readily computed, as shown in appendix B where $TN\left(\hat{\boldsymbol{\mu}}_{i,h}^{n}, \hat{\boldsymbol{\Sigma}}_{i,h}^{n}; \boldsymbol{a}, \boldsymbol{b}\right) \approx N\left(\hat{\boldsymbol{\mu}}_{i,h}^{TN}, \hat{\boldsymbol{\Sigma}}_{i,h}^{TN}\right)$ due to preserving the reproducibility of conditional PDFs (Simon, 2006; Straka et al., 2012). The detailed procedure to obtain the truncated mean vector $\left(\hat{\boldsymbol{\mu}}_{i,h}^{TN}\right)$ and covariance matrix $\left(\hat{\boldsymbol{\Sigma}}_{i,h}^{TN}\right)$ is given in appendix B for the sake of completeness.

Applying the constrained Kalman filter approach, we can update the posterior distribution of transformed coefficients according to the set of box constraints given by Lemma 3. Finally, we need to transform back the updated mean vector and covariance matrix into the mean vector and covariance matrix of the originally truncated posterior distribution using Lemma 3. Figure 3 summarizes the overall procedure for online updating considering the monotonicity constraints for the mixed effects model based on degree 1 and degree 2 B-splines with infinite support using CKF.

3.2.2 General updating procedure based on Monte Carlo sampling for degree 3 B-spline with infinite support

The monotonicity conditions of degree 3 B-splines are in the form of non-linear inequality constraints as derived in Lemma 2. Therefore, the analytical procedure presented in the previous subsection is

no longer applicable. In order to conduct online updating in this situation, we propose to use a more general approach based on the MC sampling. To be more specific, suppose that the unconstrained updating of mixed effects parameters based on the Bayesian procedure of section 2.1 results in the posterior distribution $p\left(\boldsymbol{\theta}_i|\boldsymbol{y}_i\left(\boldsymbol{t}_{i,1:h}\right)\right) \sim N(\hat{\boldsymbol{\mu}}_{i,h}, \hat{\boldsymbol{\Sigma}}_{i,h})$ for unit i at time t_h. In order to model a monotonic degradation signal using degree 3 B-splines with infinite support, the random coefficients of mixed effects model should constantly satisfy the conditions derived by Lemma 2. The aim is to estimate the posterior distribution of random coefficients subject to constraints in (13), i.e., $p\left(\boldsymbol{\theta}_i|\boldsymbol{y}_i\left(\boldsymbol{t}_{i,1:h}\right); \boldsymbol{\theta}_i \in \mathcal{C}_h\right) \sim TN\left(\hat{\boldsymbol{\mu}}_{i,h}, \hat{\boldsymbol{\Sigma}}_{i,h}; \boldsymbol{\theta}_i \in \mathcal{C}_h\right)$ where $\mathcal{C}_h$ denotes the constraints' space. The PDF of $TN\left(\hat{\boldsymbol{\mu}}_{i,h}, \hat{\boldsymbol{\Sigma}}_{i,h}; \boldsymbol{\theta}_i \in \mathcal{C}_h\right)$ can be expressed similar to (15). Equation (15) presents a closed form description of the PDF of random variable $\boldsymbol{\theta}_i$ with respect to constraints in (13). Here, due to preserving the reproducibility of the conditional PDFs (Simon, 2006; Straka et al., 2012), the truncated PDF must be approximated by a Gaussian PDF as follows:

$$p\left(\boldsymbol{\theta}_i|\boldsymbol{y}_i\left(\boldsymbol{t}_{i,1:h}\right); \boldsymbol{\theta}_i \in \mathcal{C}_h\right) \approx \; N(\hat{\boldsymbol{\mu}}^c_{i,h}, \hat{\boldsymbol{\Sigma}}^c_{i,h}). \tag{19}$$

It should be noted that the Gaussian approximation of the truncated PDF in (19) has nonzero values outside the constrained region. However compared to the original unconstrained distribution, the volume of approximate Gaussian distribution above the constrained region is close to 1. Similar approximations can also be found in Straka et al. (2012) and Simon (2006). The moment of the PDF function in (19) can be written as follows:

$$\hat{\boldsymbol{\mu}}^c_{i,h} = \int \boldsymbol{\theta}_i p\left(\boldsymbol{\theta}_i|\boldsymbol{y}_i\left(\boldsymbol{t}_{i,1:h}\right); \boldsymbol{\theta}_i \in \mathcal{C}_h\right) d\boldsymbol{\theta}_i, \tag{20}$$

$$\hat{\boldsymbol{\Sigma}}^c_{i,h} = \int \left(\boldsymbol{\theta}_i - \hat{\boldsymbol{\mu}}^c_{i,h}\right)\left(\boldsymbol{\theta}_i - \hat{\boldsymbol{\mu}}^c_{i,h}\right)^T p\left(\boldsymbol{\theta}_i|\boldsymbol{y}_i\left(\boldsymbol{t}_{i,1:h}\right); \boldsymbol{\theta}_i \in \mathcal{C}_h\right) d\boldsymbol{\theta}_i. \tag{21}$$

The mean and covariance in (20) and (21) are difficult to solve analytically except for some special cases like linear inequality constraints (Straka et al., 2012). Here, we use MC sampling technique to approximate the integrals of Equations (20) and (21). The MC sampling technique is based on approximating integrals using dense sampling of the truncated PDF. In order to approximate the integrals in (20) and (21), suppose that N samples $\boldsymbol{\theta}_i^{(r)}, r = 1, 2, \ldots, N$, are drawn from the unconstrained posterior distribution $N(\hat{\boldsymbol{\mu}}_{i,h}, \hat{\boldsymbol{\Sigma}}_{i,h})$. We divide the samples into two groups: the samples satisfying the constraints $\mathcal{C}_h$ denoted as $\boldsymbol{\theta}_i^{c,(l)},\ l = 1, 2, \ldots, N^c,\ N^c \leq\ N$ and the ones not satisfying the constraints. Thus, the mean and covariance matrix of the truncated PDF in (19) can be approximated as follows:

$$\hat{\boldsymbol{\mu}}^c_{i,h} \approx \frac{1}{N^c}\sum_{l=1}^{N^c} \boldsymbol{\theta}_i^{c,(l)}, \tag{22}$$

$$\hat{\boldsymbol{\Sigma}}^c_{i,h} \approx \frac{1}{N^c - 1}\sum_{l=1}^{N^c} \left(\boldsymbol{\theta}_i^{c,(l)} - \hat{\boldsymbol{\mu}}^c_{i,h}\right)\left(\boldsymbol{\theta}_i^{c,(l)} - \hat{\boldsymbol{\mu}}^c_{i,h}\right)^T. \tag{23}$$

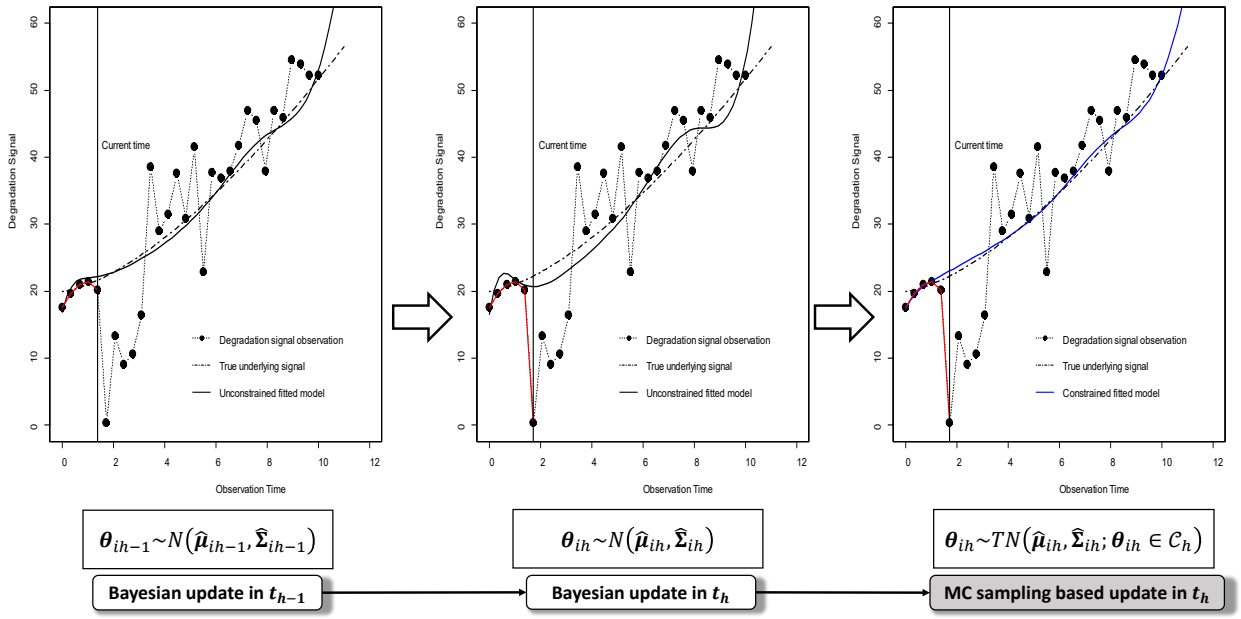

Figure 4: Online updating for the in-service unit using B-splines of degree 3 with infinite support.

Therefore, in order to find the approximate mean and covariance matrix of the truncated posterior PDF in (19), we need to simulate a large number of samples from $N(\hat{\boldsymbol{\mu}}_{i,h}, \hat{\boldsymbol{\Sigma}}_{i,h})$ and form the pool of samples satisfying the required constraints given in (13). Then the mean and covariance matrix of posterior distribution considering the constraints in (13) can be approximated using Equations (22) and (23). Figure 4 illustrates the online updating procedure considering the monotonicity constraints using MC sampling for a mixed effects model based on infinite support B-splines of degree 3 when we get a new observation from the in-service unit.

4. Numerical Study

In this section, the performance of the proposed mixed effects model framework will be investigated. Specifically, we first discuss the general procedure to simulate the degradation signals and evaluate the performance of different methods. Then, using the simulated signals, we demonstrate the advantages of our proposed method compared with some existing methods in the literature.

To motivate the application of the proposed approach, we consider an illustrative example. Let us assume we have signals generated from a true underlying function with noise. Specially we assume that the signals are generated from $R_y(t) = \omega_y t^{1.5} + \epsilon$, where $y = 1, 2, \ldots, 15$, $t \in [0, 10]$ and $\omega_y \sim N(1, 0.25^2)$. The observations are made at 30 evenly spaced points and the measurement noise is set to $\sigma_\varepsilon = 5$. Moreover, it is assumed that all the signals in the historical dataset are truncated after a specific observation time. Figure 5 demonstrates the signals generated in such a

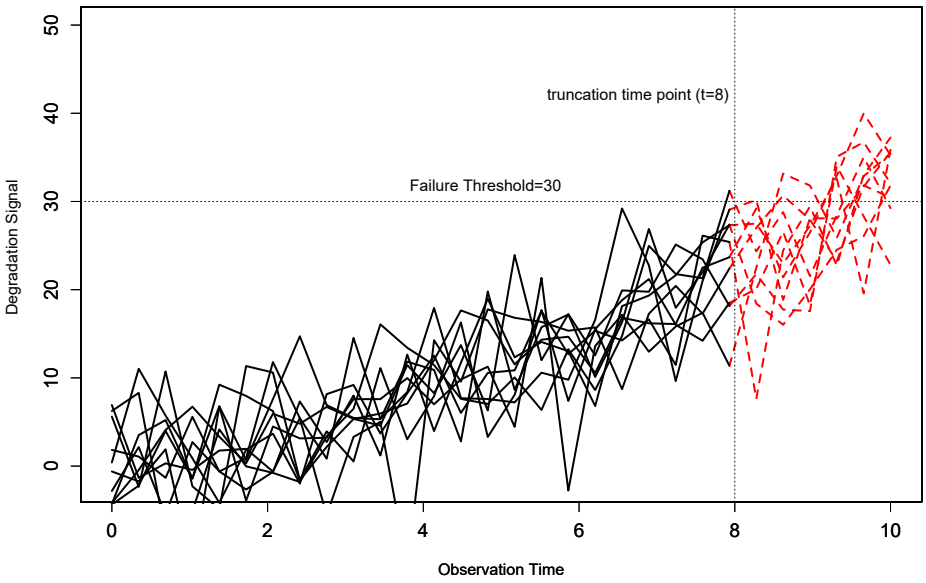

Figure 5: Truncated historical observations with noise.

setting where it is assumed that all the historical signals are observed until time point $t = 8$. The failure threshold for this specific example is set to TH=30.

As shown in Figure 5, the failure time of all units is not observed. Therefore, if the new in-service unit fails after $t = 8$, then we need to extrapolate to find its failure time. Considering these truncated historical signals, we can fit a mixed effects model to predict the remaining useful life for the new in-service unit. In this regard, we consider the case where the mixed effects model is used both with regular degree 3 basis functions and degree 3 basis functions with infinite support. It is assumed that we have partial observations for the new in-service unit and the aim is to predict its failure time. Figure 6 demonstrates the prognosis results for the new in-service unit.

From Figure 6, it can be clearly observed that the prediction results based on the regular B-spline are misleading. The main reason is that the regular B-spline, as with most of the non-parametric approaches, is limited to the range of historical observations which prevents it from extrapolating beyond this range. However, for the infinite support B-spline, there is no such limitation and the method can extrapolate and predict beyond the range. This extrapolation results in accurate estimation of the failure time in this method as can be seen from Figure 6. However, it should be mentioned that the extrapolation result of the proposed framework is indeed more accurate when the truncation point is closer to the failure time and the signal trend is similar after and before the truncation point.

Further, insight regarding the importance of monotonicity of the proposed approach can also be gained from Figure 6. As can be seen, the regular B-spline based approach with no constraints on

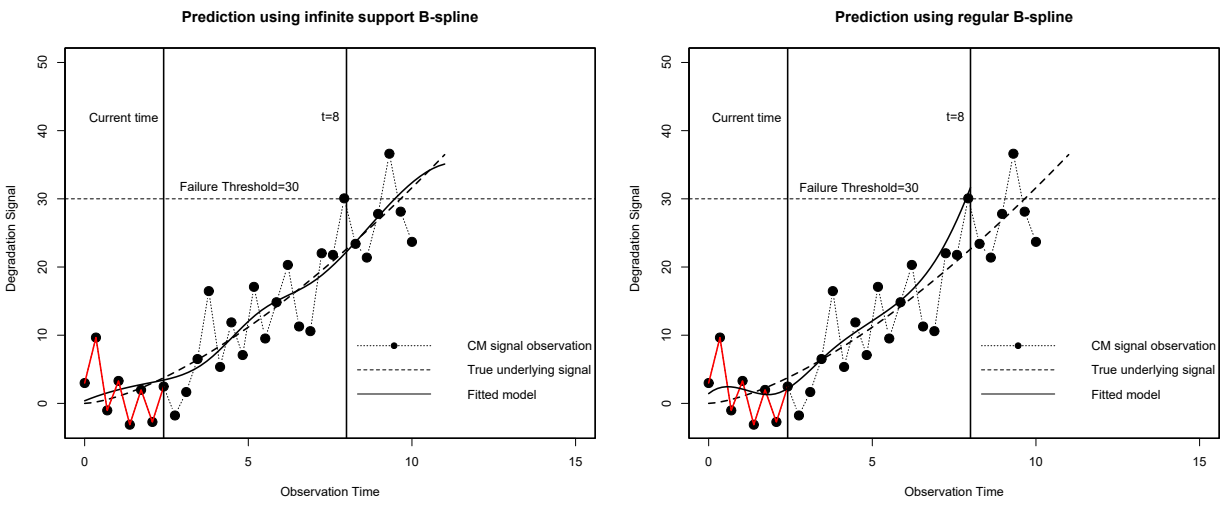

Figure 6: Prognostics using infinite support B-splines vs. regular B-splines.

the evolution of the degradation signals results in a signal which is not monotonically increasing. This contradicts the underlying domain knowledge that the degradation signal cannot decrease over time. This phenomenon, in practice, happens as a result of high level of noise in sensor data observations. Therefore, the most suitable approach is the one that allows for certain flexibility along with the ability to impose the domain knowledge. Imposing the appropriate monotonicity constraint in the infinite support B-spline approach can account for the high level of noise and adjust the prediction which, in turn, results in more accurate prediction of the failure time. This feature is well demonstrated in the evolution path of the in-service unit based on the two model settings in Figure 6.

Regarding the numerical study, we simulate the degradation signals from five different model settings. In each model setting, we report the prediction accuracy of different methods for partially observed in-service signal at varying time points t^*. Specifically, we report the prediction accuracy for different percentiles of observed in-service signal (i.e., 40%, 60% and 80%). Our procedure to compare different methods is based on simulating Y=15 degradation signals from a specific model setting and then randomly selecting one of them as the in-service unit r, and the rest as the historical dataset. This procedure is repeated for 1000 times and each time we report the absolute error (AE) as our criterion for comparing different methods. The absolute error between the true failure time T_r and the estimated failure time $\hat{T}_r(t^*)$ can be described as follows:

$$AE\left(t^*\right) = \left|\hat{T}_r\left(t^*\right) - T_r\right|, \tag{24}$$

where as mentioned before a unit fails when its degradation signal passes the failure threshold.

The performance of different approaches are compared using a series of boxplots highlighting the distribution of AE through 1000 simulation runs. Specifically, we compare the performance of

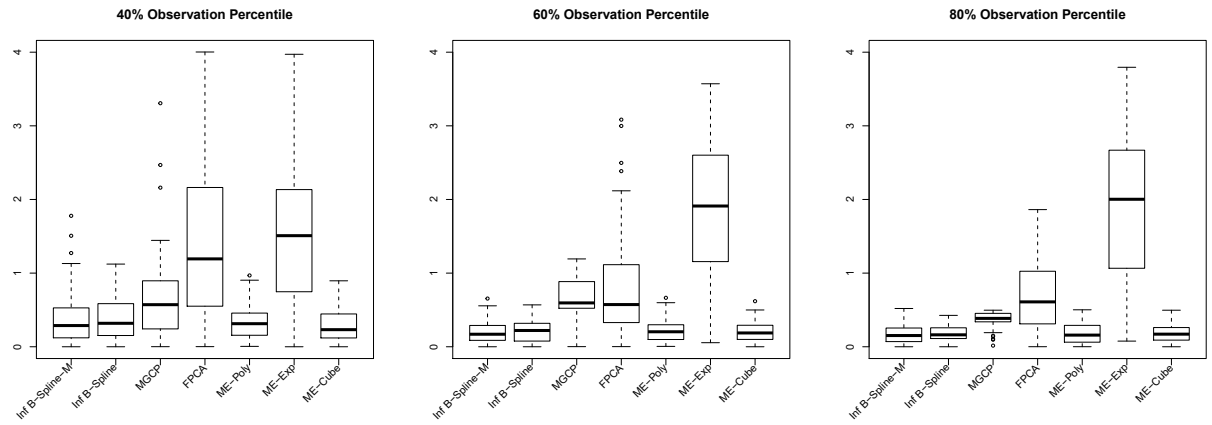

Figure 7: Prediction accuracy results of model setting I.

the proposed degree 3 infinite support B-spline with monotonicty constraint approach with four other methods proposed in the literature. The first one is the parametric mixed-effects model using an exponential function denoted by ME-Exp introduced by Gebraeel et al. (2005). The second method is the nonparametric FPCA approach proposed by Zhou et al. (2011). The third approach is the mixed-effects model using general polynomial function proposed by Son et al. (2013) and Rizopoulos (2011) denoted as ME-Poly. The appropriate degree of polynomial is determined using the Akaike information criteria (AIC). For this approach we also consider the case where the degree of polynomial is always fixed at 3 to have a measure of comparision with our proposed framework with degree 3 B-splines. We denote this approach as ME-Cube. Finally, we compare the performance of our approach with the MGCP approach of Kontar et al. (2018). In addition to the proposed framework with monotonicity constraints, we also consider the performance of infinite support B-spline without the monotonicity constraints to make the numerical study more complete. We denote the former as Inf B-spline-M, while the latter is denoted as Inf B-spline in our model settings.

In this study, we consider five model settings to compare the performance of different approaches. In model setting I, we highlight the importance of monotonicity in degradation modeling. Specifically in model setting I, we adopt the quadratic function defined in Kontar et al. (2018). In this regard, we consider the setting where the outputs for the Y curves are generated from the $R_y(t) = \omega_{1,y}^I t^2 + \epsilon$, where $y = 1, 2, \ldots, Y$, $t \in [0, 10]$ and $\omega_{1,y} \sim N(1, 0.25^2)$. Moreover it is assumed that the observations are made at 30 evenly spaced points, the measurement noise standard deviation is set to $\sigma_\varepsilon = 0.5$ and the failure threshold is set at TH=40. In order to benchmark the result with the correctly specified parametric form, we use the quadratic function in the polynomial mixed effects model (ME-Poly). The results are demonstrated in Figure 7.

From Figure 7, we can see that the performance of the proposed non-parametric approach based on the B-Splines of infinite support is comparable to that of the correctly specified parametric model

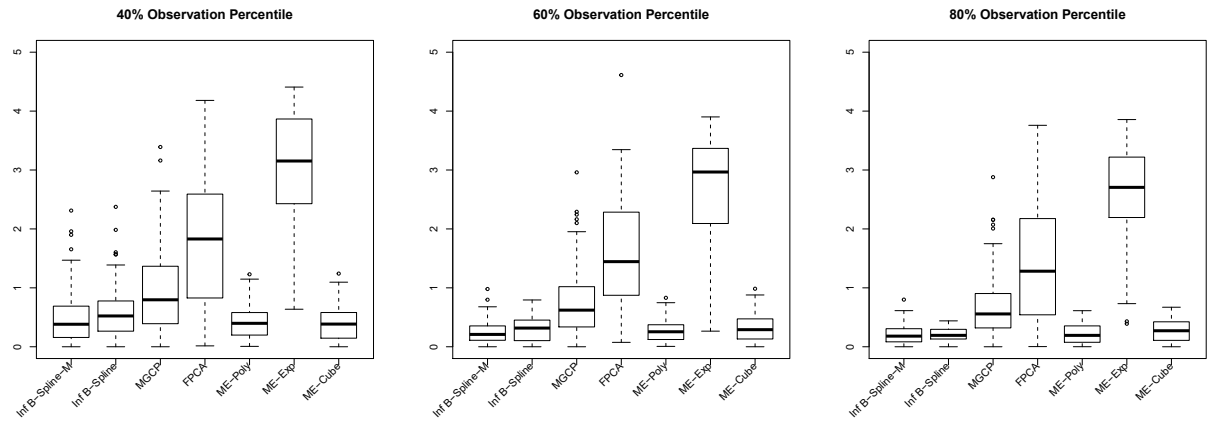

Figure 8: Prediction accuracy results of model setting II.

in ME-Poly approach. Specifically, in the earlier stage when only 40% of the in-service unit signal is available, the variance of prediction error for the proposed method is a little bit higher than the ME-Poly approach which is based on the true parametric function. This happens because the ME-Poly approach which is based on the true parametric form, can inherently capture the trend even with a small number of observations. However, our proposed method requires more observations to capture the true underlying trend and its prediction accuracy improves as it gets more data. Moreover, we can see that by increasing the order of polynomial of mixed effect model in ME-Cube approach we get similar performance as in ME-Poly.

From the results in Figure 7, we can gain insight regarding the importance of the monotonicity constraint considered in our approach. The degradation signals generated in model setting I are inherently monotone and imposing monotonicity through our proposed framework increases the accuracy of its predictions. However, other non-parametric approaches considered in this study cannot guarantee monotonicity. This is, in fact, the main reason that their prediction errors are higher than that of our proposed method. It should also be mentioned that an important issue with MGCP approach is scalability as mentioned in Kontar et al. (2018). This issue becomes more prominent as we increase the number of signals in the historical dataset which leads the optimization of log-likelihood to converge to local optima. This, in turn, results in inferior result for the MGCP approach.

In order to further highlight the importance of extrapolation in our proposed approach, we conduct a numerical study with doubly truncated signals. Specifically, in model settings II, we adopt the signals introduced in model setting I with one modification. In model setting II, we assume that in addition to the failure threshold, all the signals in the offline stage are truncated after time $t = 7$. The prediction accuracy results of this model setting are given in Figure 8.

As can be seen from Figure 8, the performance of the infinite support B-spline approach proposed in this paper is comparable to the ME-Poly approach which is based on the correctly specified

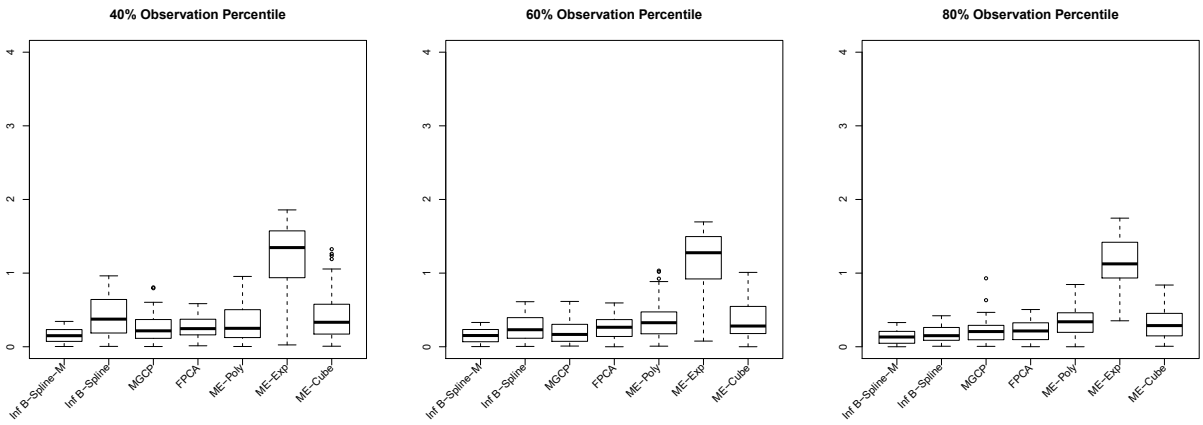

Figure 9: Prediction accuracy results of model setting III.

parametric approach. As mentioned earlier, in this model setting, no historical observations are available after time $t = 7$. However, this does not significantly affect the performance of the ME-Poly approach as it is inherently based on the correctly specified parametric form. This, in fact, happens because even if the signals in the historical dataset are truncated after time $t = 7$, this approach can learn the true behavior of degradation signals based on the initial observations. Regarding the infinite support based B-spline approach, however, the evolution of the in-service degradation signal is extrapolated based on the behavior of the historical degradation signals until time $t = 7$. As can be seen from Figure 8, the performance of this extrapolation is comparable to that of the true parametric form in different observation percentiles. However, as for the other non-parametric approaches, the predictions outside the range of historical data converge to an overall mean function which result in wrong estimation of the failure time. This can be seen from the result of numerical study in Figure 8.

In order to further investigate the performance of our approach, we conduct three more numerical studies. Specifically, in model setting III, we focus on a nonlinear signal function adapted from Kontar et al. (2018). The output for Y curves are generated from $R_y(t) = 1.5t + \omega_{1,y}^{III} \sin(t) + \omega_{2,y}^{III}$, where $y = 1, 2, \ldots, Y$ and $t \in [0, 10]$. For the yth output, $\omega_{1,y}^{III} \sim U(0.8, 1.2)$ and $\omega_{2,y}^{III} \sim U(0, 7)$. It is assumed that there are 35 evenly spaced observations per signal. The measurement noise standard deviation is set to $\sigma_\varepsilon = 0.01$ and the failure threshold is set at TH=14. In model setting IV, we randomly sample 10 observations from each signal to further evaluate the robustness of our approach when sparse data are available. Moreover, model setting V investigates the performance of the proposed method when the noise level is high. Specifically, we set the measurement noise standard deviation to $\sigma_\varepsilon = 0.05$. Figure 9 demonstrates the results of model setting III.

From Figure 9, we can see that in different observation percentiles, the median and variance of the prediction errors of the proposed method are less than those of the other parametric and non-parametric approaches proposed in the literature. Moreover, it can be seen that the variance

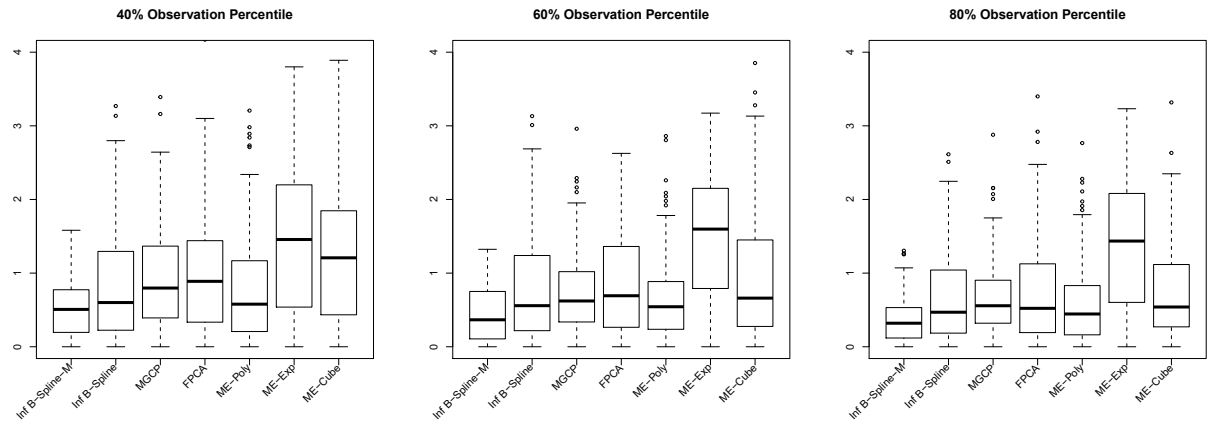

Figure 10: Prediction accuracy results of model setting IV.

and median of prediction errors decreases with increase in the observation percentiles. In fact, as more data from the in-service unit becomes available in the online stage, the accuracy of different methods gets better. We can also get an interesting insight by comparing the results of the parametric and non-parametric approaches. As can be seen from Figure 9, the prediction based on the parametric approaches of ME-Poly, ME-Exp and ME-Cube leads to worse results than that of the non-parametric approaches in different observation percentiles. This signals the prominent hazard of defining a wrong parametric model in predicting the RUL. On the other hand, the non-parametric approaches enjoy enough flexibility to adapt with the specific trend of the signal. Moreover comparing the performance of Inf B-Spline-M and Inf B-Spline models, we can see that imposing monotonicity constraints increases the prediction accuracy of the proposed framework for nonlinear signal function forms.

Results from Figure 10 further demonstrate the robustness of different approaches to sparsity in observations from the in-service unit. It can be seen that sparsity in observations leads to increase in the prediction errors of different approaches. However, the mean and variance of absolute error of predictions based on our proposed method remains less than those of other approaches. Moreover, Figure 11 showcases the impact of increase in the noise of observations. It can be seen that increase in the level of noise increases the prediction error of the different approaches; however, the performance of our proposed approach remains superior to that of other methods.

From the result of numerical study we can see that the importance of monotonicity constraint is, in fact, more notable in the early stage of prediction when the data is rare. This can be observed more vividly comparing the performance of the infinite support B-spline with and without monotonicity constraints in all our model settings. As can be seen from Figures 7-11, the infinite support B-spline with monotonicity constraint typically performs better than other approaches in the early stage of RUL prediction when we cannot learn the monotonic trend of signal based on available data. This better performance gets clearer, especially, when we have scarcity or higher

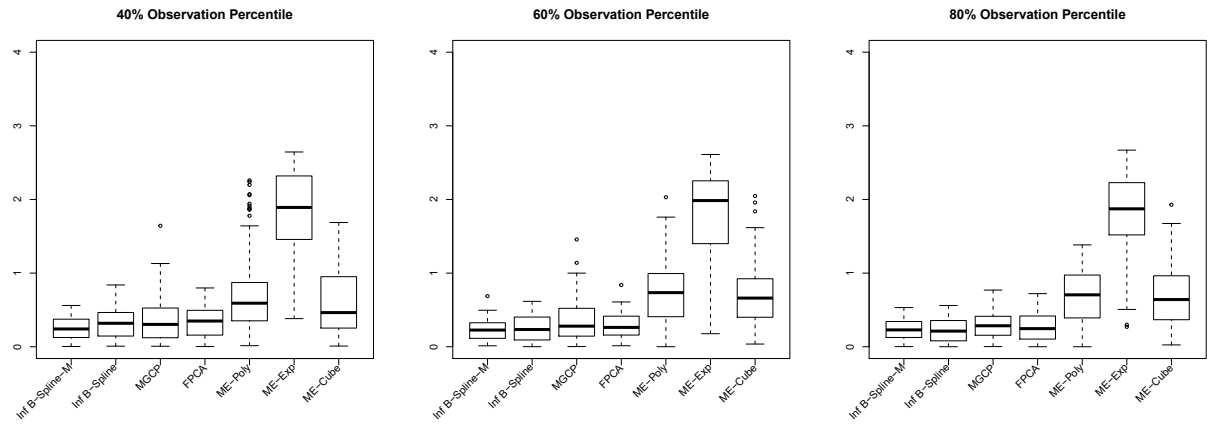

Figure 11: Prediction accuracy results of model setting V.

noise level in our observations as discussed in model settings IV and V. The scarcity, insufficient sampling data and high measurement noise are, indeed, inevitable in most manufacturing systems which makes it crucial for decision making and control policies to have a robust measure of RUL.

5. Real-world case study

In this section, application of the proposed prognosis procedure on the real-world data collected from an automotive lead-acid battery aging test is demonstrated. The dataset contains 14 batteries and each battery fails when its degradation signal reaches a prespecified threshold level. Figure 12 demonstrates the degradation signal evolution of batteries from this dataset. The failure threshold is defined as 5.4 and all the units are considered failed once their degradation signal hits the failure threshold which results in truncated signals. Due to confidentiality, the data presented in Figure 12 has been slightly modified from the original data, but this modification does not affect our study here.

In order to compare the performance of different methods, we use the leave-one-out cross validation approach. Specifically, we consider that one of the 14 batteries is the in-service unit and the model fitting in offline stage is performed for the remaining 13 ones. Then, the predictions for the in-service battery is performed for different percentiles of its lifespan (i.e., 40%, 60% and 80%). Both degree 3 and degree 2 infinite support B-spline with monotonicity constraint are used in this case study. We denote the former as Inf B-Spline-D2 while the latter is denoted as Inf B-Spline-D3. The whole procedure is repeated 14 times and the AE is calculated each time. Figure 13 summarizes the results for the case study.

To the best of authors' knowledge, there is no general physical model for the degradation of lead acid batteries. Therefore, a quadratic degradation path is used for the ME-Poly approach. This further highlights the importance of non-parametric approaches in degradation modeling when

there is no general physical model. As can be seen from Figure 13, in different percentiles of the observed signal of the in-service unit, our approach based on degree 3 B-splines of infinite support performs relatively better than the other approaches proposed in literature. The main reason for the superior performance of our method is that it can effectively deal with truncated signals using the basis functions of infinite support. This feature combined with the monotonicity constraint on the evolution of the degradation signal guarantee that predictions are not limited to the range of historical data as with other non-parametric approaches. On the contrary, in non-parametric

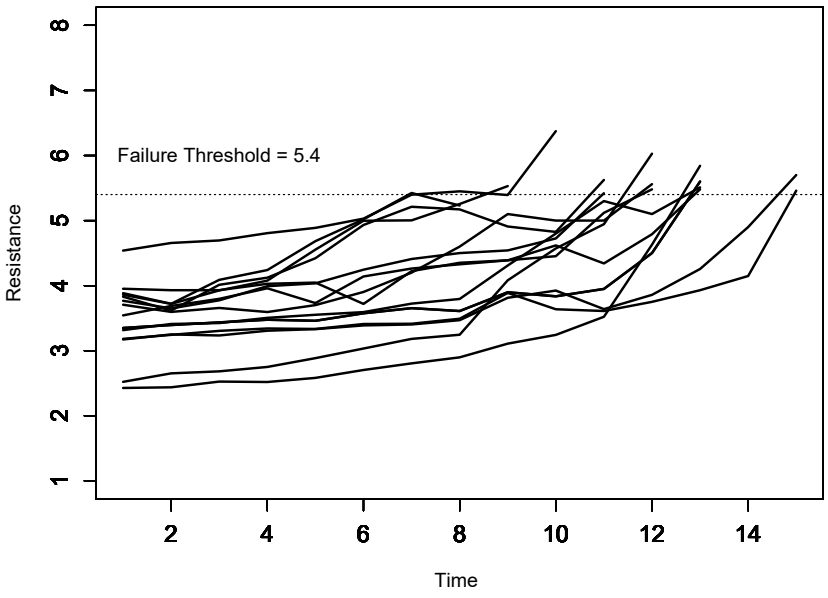

Figure 12: Battery resistance data from an accelerated aging test.

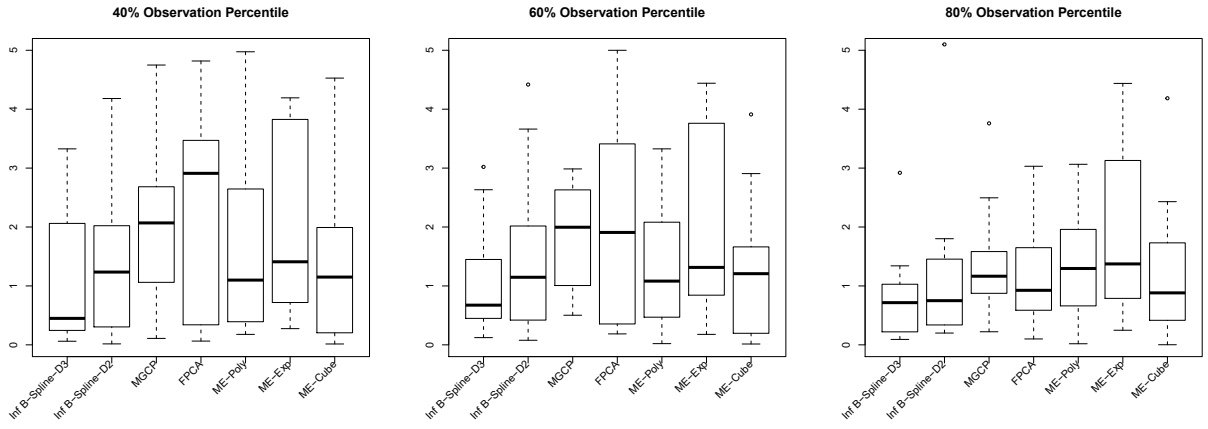

Figure 13: Prediction accuracy results for case study.

approaches like MGCP, predictions outside the range of historical data converge to an overall mean function which is typically defined as zero. This, indeed, results in wrong estimation of the RUL as the predications may never hit the failure threshold. However, this is not an issue in our approach which is based on B-splines of infinite support. Moreover comparing B-splines models, we can see that the degree 3 B-splines which provide higher flexibility have a better performance in terms of the absolute error compared with the degree 2 B-spline. Thus, the case study further confirms the results of the numerical study that, in the case of truncated degradation signals, the performance of our proposed approach is superior to the methods proposed in the literature.

6. Conclusion

In this study, a non-parametric approach to modeling and prognosis of degradation signals using B-splines in a mixed effects framework is proposed. Unlike most of non-parametric approaches that rely on a historical sample of complete degradation signals, our approach is suitable for modeling truncated signals. In this regard, our approach is based on the B-splines augmented with basis functions of infinite support. One advantage of using this framework is that it allows us to extrapolate the evolution of the in-service degradation signal beyond the range of truncated degradation signals from our historical dataset.

Moreover, the B-spline setting allows us to encode the inherent monotonic evolution path of degradation signals. In this regard, the necessary and sufficient conditions to guarantee monotonic evolution of degradation signals in a mixed effects model based on B-splines of infinite support are derived. Our approach for updating random coefficients of the proposed mixed effects model with B-splines of degree 1 and degree 2 considering the derived monotonicity constraints is based on transforming the resulted truncated multivariate normal distribution of coefficients into a simpler form of box constrained multivariate normal distribution and updating using the constraint Kalman filter. Regarding the B-splines of degree 3, slightly more complicated necessary and sufficient monotonicity conditions based on polynomial representation of the B-spline basis functions are developed. The derived conditions for updating the mixed effects model in the online stage in this case leads to a truncated multivariate normal distribution with non-linear inequality constraints. In order to estimate the mean and covariance matrix of the resulting truncated PDF, we resort to the Monte Carlo sampling technique. The results of our numerical study along with the real-world case study confirm the advantage of the proposed framework in dealing with truncated signals.

In this paper, we have only considered modeling of one degradation signal. However, most of the contemporary real-time monitoring systems tend to collect more than one degradation signal. In such cases, the need for developing more sophisticated degradation modeling approaches arises that can deal with both the issues of monotonicity and truncated degradation signal paths. This can be considered as a potential future research direction that we intend to pursue.

Appendices

A. Induction Formula for Infinite Support B-splines

The induction formula for the infinite support B-spline of degree $1 \leq n \leq k$ is as follows:

$$\begin{cases} b_{j,n}(t) := \frac{\gamma_j - t}{C_0} b_{j,n-1}(t), & j = 0, \\ b_{j,n}(t) := b_{j-1,n-1}(t) + \frac{\gamma_j - t}{C_0} b_{j,n-1}(t), & 1 \leq j < min(n,k), \\ b_{j,n}(t) := \begin{cases} b_{j-1,n-1}(t) + \frac{\gamma_j - t}{\gamma_j - \gamma_{j-n}} b_{j,n-1}(t), & if k > n \\ b_{j-1,n-1}(t) + b_{j,n-1}(t), & if k = n \end{cases} & j = min(n,k) \\ b_{j,n}(t) := \frac{t - \gamma_{j-n-1}}{\gamma_{j-1} - \gamma_{j-n-1}} b_{j-1,n-1}(t) + \frac{\gamma_j - t}{\gamma_j - \gamma_{j-n}} b_{j,n-1}(t), & min(n,k) + 1 \leq j < max(n,k), \\ b_{j,n}(t) := \begin{cases} b_{j,n-1}(t) + \frac{t - \gamma_{j-n-1}}{\gamma_{j-1} - \gamma_{j-n-1}} b_{j-1,n-1}(t), & if k > n \\ b_{j-1,n-1}(t) + b_{j,n-1}(t), & if k = n \end{cases} & j = max(n,k) \\ b_{j,n}(t) := \frac{t - \gamma_{j-n-1}}{C_1} b_{j-1,n-1}(t) + b_{j,n-1}(t), & max(n,k) + 1 \leq j < k + n, \\ b_{j,n}(t) := \frac{t - \gamma_{j-n-1}}{C_1} b_{j-1,n-1}(t), & j = k + n \end{cases} \tag{25}$$

where the term $\frac{t-\gamma_{j-n-1}}{\gamma_{j-1}-\gamma_{j-n-1}}$ and $\frac{\gamma_j - t}{\gamma_j - \gamma_{j-n}}$ are replaced by 1 and 0 respectively, when their denominators are equal to zero. Regarding the choice of constants C_0 and C_1, we use the guideline proposed in Corlay (2016). In this paper, we take $C_0 = C_1 = \frac{\gamma_{k-1} - \gamma_0}{k-1}$ if $\gamma_{k-1} > \gamma_0$ and $C_0 = C_1 = 1$ otherwise.

The B-spline basis functions of degree $n > k$ associated with knots Γ are defined as follows:

$$\begin{cases} b_{j,n}(t) := \frac{\gamma_j - t}{C_0} b_{j,n-1}(t), & j = 0, \\ b_{j,n}(t) := b_{j-1,n-1}(t) + \frac{\gamma_j - t}{C_0} b_{j,n-1}(t), & 1 \leq j < min(n,k), \\ (b_{j,n})_{min(n,k) \leq\ j < max(n,k)+1} & \text{any basis of } \rho_{max(n,k)-min(n,k)}(\mathbb{R}) = \rho_{n-k}(\mathbb{R}) \\ b_{j,n}(t) := \frac{t - \gamma_{j-n-1}}{C_1} b_{j-1,n-1}(t) + b_{j,n-1}(t), & max(n,k) + 1 \leq j < k + n, \\ b_{j,n}(t) := \frac{t - \gamma_{j-n-1}}{C_1} b_{j-1,n-1}(t), & j = k + n \end{cases} \tag{26}$$

B. Details of Constrained Kalman Filter

Using the moment generating function (MGF), the mean $\hat{\boldsymbol{\mu}}_{ih}^{TN} = \left[\hat{\mu}_{ih,1}^{TN}, \ldots, \hat{\mu}_{ih,d}^{TN}\right]^T$ and the covariance matrix $\widehat{Cov}^{TN}[\theta_{i,j}, \theta_{i,l}]$ for $j, l \in \{1, \ldots, d\}$ of the $TN\left(\hat{\boldsymbol{\mu}}_{i,h}^{n}, \hat{\boldsymbol{\Sigma}}_{i,h}^{n}; \boldsymbol{a}, \boldsymbol{b}\right)$ with density function given in (18) can be computed as follows:

$$\hat{\mu}_{ih,j}^{TN} = \hat{\mu}_{ih,j}^{n} + \psi_j\left(\boldsymbol{a}, \boldsymbol{b}; \hat{\boldsymbol{\Sigma}}_{ih}^{n}\right) = \hat{\mu}_{ih,j}^{n} + \sum_{l=1}^{d} \sigma_{ih,j,l}\left\{\varphi_l(a_l) - \varphi_l(b_l)\right\}, \tag{27}$$

and

$$\widehat{Cov}^{TN}\left[z_{i,j}, z_{i,l}\right] = \sigma_{ih,j,l} + \xi_{j,l}\left(\boldsymbol{a}, \boldsymbol{b}; \hat{\boldsymbol{\Sigma}}_{ih}^{n}\right) + \eta_{j,l}\left(\boldsymbol{a}, \boldsymbol{b}; \hat{\boldsymbol{\Sigma}}_{ih}^{n}\right) - \psi_j\left(\boldsymbol{a}, \boldsymbol{b}; \hat{\boldsymbol{\Sigma}}_{ih}^{n}\right)\psi_l\left(\boldsymbol{a}, \boldsymbol{b}; \hat{\boldsymbol{\Sigma}}_{ih}^{n}\right), \tag{28}$$

where $\sigma_{ih,j,l}$ denotes covariance between the jth and the lth parameters. The $\xi_{j,l}\left(\boldsymbol{a}, \boldsymbol{b}; \hat{\boldsymbol{\Sigma}}_{ih}^{n}\right)$ and $\eta_{j,l}\left(\boldsymbol{a}, \boldsymbol{b}; \hat{\boldsymbol{\Sigma}}_{ih}^{n}\right)$ functions used in (27) and (28) are respectively defined as follows

$$\xi_{j,l}\left(\boldsymbol{a}, \boldsymbol{b}; \hat{\boldsymbol{\Sigma}}_{ih}^{n}\right) = \sum_{q=1}^{d} \frac{\sigma_{ih,j,q}\sigma_{ih,l,q}\left\{a_q\varphi_q\left(a_q\right) - b_q\varphi_q\left(b_q\right)\right\}}{\sigma_{ih,q,q}}, \tag{29}$$

and

$$\eta_{j,l}\left(\boldsymbol{a}, \boldsymbol{b}; \hat{\boldsymbol{\Sigma}}_{ih}^{n}\right) = \sum_{q=1}^{d}\sigma_{ih,j,q}\sum_{q\neq\ o}\left[\left(\sigma_{ih,l,o} - \frac{\sigma_{ih,q,o}\sigma_{ih,l,q}}{\sigma_{ih,q,q}}\right) \times\left\{\varphi_{q,o}\left(a_q, a_o\right) + \varphi_{q,o}\left(b_q, b_o\right) - \varphi_{q,o}\left(a_q, b_o\right) - \varphi_{q,o}\left(b_q, a_o\right)\right\}\right] \tag{30}$$

where φ_l and $\varphi_{q,o}$ are univariate and bivariate marginal density functions. The univariate marginal density is $\varphi_l\left(x\right) = \int_{\boldsymbol{\Omega}_{-l}} f_T\left(x, \boldsymbol{Z}_{-l}\right)\ d\boldsymbol{Z}_{-l}$ where $\boldsymbol{Z}_{-l}$ is a $(d-1)$-dimensional vector that excludes $z_{i,l}$ and $\boldsymbol{\Omega}_{-l} = \left\{a_j \leq z_{i,j} \leq b_j : \forall j \neq l, j = 1, \ldots, d\right\}$. The bivariate marginal density is $\varphi_l\left(x, y\right) = \int_{\boldsymbol{\Omega}_{-q-o}} f_T\left(x, y, \boldsymbol{Z}_{-q-o}\right)\ d\boldsymbol{Z}_{-q-o}$ where $\boldsymbol{Z}_{-q-o}$ is a $(d-2)$-dimensional vector $\boldsymbol{Z}_{-q-o} = \left[z_{i,1}, \ldots, z_{i,q-1}, z_{i,q+1}, \ldots, z_{i,o-1}, z_{i,o+1}, \ldots, z_{i,d}\right]^T$ that excludes both $z_{i,q}$ and $z_{i,o}$ for $q \neq\ o$, and $\boldsymbol{\Omega}_{-q-o} = \left\{a_j \leq z_{i,j} \leq b_j : \forall j \neq q \text{and} \forall j \neq o, j = 1, \ldots, d\right\}$ for $q \neq o$.

References

Aguilar, E., Elizalde, H., Cárdenas, D., Probst, O., Marzocca, P. and Ramirez-Mendoza, R. A. (2018), 'An adaptive curvature-guided approach for the knot-placement problem in fitted splines', *Journal of Computing and Information Science in Engineering* **18**(4), 041013.

Beliakov, G. (2002), 'Monotone approximation of aggregation operators using least squares splines', *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* **10**(06), 659–676.

Byon, E., Ntaimo, L. and Ding, Y. (2010), 'Optimal maintenance strategies for wind turbine systems under stochastic weather conditions', *IEEE Transactions on Reliability* **59**(2), 393–404.

Corlay, S. (2016), 'B-spline techniques for volatility modeling', *Journal of Computational Finance* **19**.

De Boor, C., De Boor, C., Mathématicien, E.-U., De Boor, C. and De Boor, C. (1978), *A practical guide to splines*, Vol. 27, Springer-Verlag New York.

Gebraeel, N., Elwany, A. and Pan, J. (2009), 'Residual life predictions in the absence of prior degradation knowledge', *IEEE Transactions on Reliability* **58**(1), 106–117.

Gebraeel, N. Z., Lawley, M. A., Li, R. and Ryan, J. K. (2005), 'Residual-life distributions from component degradation signals: A bayesian approach', *IiE Transactions* **37**(6), 543–557.

Kontar, R., Son, J., Zhou, S., Sankavaram, C., Zhang, Y. and Du, X. (2017), 'Remaining useful life prediction based on the mixed effects model with mixture prior distribution', *IISE Transactions* **49**(7), 682–697.

Kontar, R., Zhou, S., Sankavaram, C., Du, X. and Zhang, Y. (2018), 'Nonparametric modeling and prognosis of condition monitoring signals using multivariate gaussian convolution processes', *Technometrics* **60**(4), 484–496.

Liao, L. and Köttig, F. (2014), 'Review of hybrid prognostics approaches for remaining useful life prediction of engineered systems, and an application to battery life prediction', *IEEE Transactions on Reliability* **63**(1), 191–207.

Lu, C. J. and Meeker, W. O. (1993), 'Using degradation measures to estimate a time-to-failure distribution', *Technometrics* **35**(2), 161–174.

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.

Rasmussen, C. E. (2004), Gaussian processes in machine learning, *in* 'Advanced lectures on machine learning', Springer, pp. 63–71.

Rizopoulos, D. (2011), 'Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data', *Biometrics* **67**(3), 819–829.

Rizopoulos, D., Hatfield, L. A., Carlin, B. P. and Takkenberg, J. J. (2014), 'Combining dynamic predictions from joint models for longitudinal and time-to-event data using bayesian model averaging', *Journal of the American Statistical Association* **109**(508), 1385–1397.

Schumaker, L. (2007), *Spline functions: basic theory*, Cambridge University Press.

Shiau, J.-J. H. and Lin, H.-H. (1999), 'Analyzing accelerated degradation data by nonparametric regression', *IEEE Transactions on reliability* **48**(2), 149–158.

Si, X.-S., Wang, W., Hu, C.-H. and Zhou, D.-H. (2011), 'Remaining useful life estimation–a review on the statistical data driven approaches', *European journal of operational research* **213**(1), 1–14.

Simon, D. (2006), *Optimal state estimation: Kalman, H infinity, and nonlinear approaches*, John Wiley & Sons.

Son, J., Zhou, Q., Zhou, S., Mao, X. and Salman, M. (2013), 'Evaluation and comparison of mixed effects model based prognosis for hard failure', *IEEE Transactions on Reliability* **62**(2), 379–394.

Son, J., Zhou, S., Sankavaram, C., Du, X. and Zhang, Y. (2016), 'Remaining useful life prediction based on noisy condition monitoring signals using constrained kalman filter', *Reliability Engineering & System Safety* **152**, 38–50.

Straka, O., Duník, J. and Šimandl, M. (2012), 'Truncation nonlinear filters for state estimation with nonlinear inequality constraints', *Automatica* **48**(2), 273–286.

Sun, B., Yan, M., Feng, Q., Li, Y., Ren, Y., Zhou, K. and Zhang, W. (2018), 'Gamma degradation process and accelerated model combined reliability analysis method for rubber o-rings', *IEEE Access* .

Tjahjowidodo, T. et al. (2017), 'A direct method to solve optimal knots of b-spline curves: An application for non-uniform b-spline curves fitting', *PloS one* **12**(3), e0173857.

Tseng, S.-T. and Peng, C.-Y. (2004), 'Optimal burn-in policy by using an integrated wiener process', *IIE Transactions* **36**(12), 1161–1170.

Vanem, E. (2018), 'Statistical methods for condition monitoring systems', *International Journal of Condition Monitoring* **8**(1), 9–23.

Virkler, D. A., Hillberry, B. and Goel, P. (1979), 'The statistical nature of fatigue crack propagation', *Journal of Engineering Materials and Technology* **101**(2), 148–153.

Wang, W.-b. (2000), 'A model to determine the optimal critical level and the monitoring intervals in condition-based maintenance', *International Journal of Production Research* **38**(6), 1425–1436.

Wang, X. and Xu, D. (2010), 'An inverse gaussian process model for degradation data', *Technometrics* **52**(2), 188–197.

Wilhelm, S. et al. (2012), 'Moments calculation for the doubly truncated multivariate normal density', *arXiv preprint arXiv:1206.5387* .

Yuan, Y., Chen, N. and Zhou, S. (2017), 'Modeling regression quantile process using monotone b-splines', *Technometrics* **59**(3), 338–350.

Zhang, Z., Si, X., Hu, C. and Lei, Y. (2018), 'Degradation data analysis and remaining useful life estimation: A review on wiener-process-based methods', *European Journal of Operational Research* .

Zhou, R., Gebraeel, N. and Serban, N. (2012), 'Degradation modeling and monitoring of truncated degradation signals', *IIE Transactions* **44**(9), 793–803.

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* pp. 1586–1610.