Functional Principal Component Analysis for Extrapolating Multistream Longitudinal Data

Seokhyun Chung and Raed Kontar

***Abstract*—In this article, we present a nonparametric approach to predict the evolution of multistream longitudinal data. Our approach first decomposes each stream into a linear combination of eigenfunctions and their corresponding functional principal component (FPC) scores. A Gaussian process prior for the FPC scores is then induced based on a functional semimetric that introduces a similarity measure across streams. Finally, an empirical Bayesian updating strategy is derived to update the established prior using real-time stream data. Empirical evidence shows that the proposed framework outperforms state-of-the-art approaches and can effectively account for heterogeneity as well as achieve high predictive accuracy.**

***Index Terms*—Bayesian inference, functional principal component (FPC) analysis, Gaussian processes (GPs).**

I. Introduction

AMONG various environments, where longitudinal data are gathered, the environment covered in this study is a *multistream* and *real-time* environment. Recent progress in sensor and data storage technologies has facilitated data collection from multiple sensors in real-time as well as the accumulation of historical signals from multiple similar units during their operational lifetime. This data structure where multiple signals across different units are collected is referred to as multistream longitudinal data. Examples include: vital health signals from patients collected through wearable devices [1], [2], battery degradation signals from cars on the road [3], [4], and energy usage patterns from different smart home appliances [5].

In this article, we propose an efficient approach to extrapolate multistream data for an in-service unit through borrowing strength from other historical units. An illustrative example is provided in Fig. 1. In this figure, there are N historical units and an in-service unit, whose index is denoted by r. Each unit has M identical sensors from which each respective signal forms a stream. Multistream data from the in-service unit are partially observed up to the current time instance t^*. Our goal is to extrapolate stream data from the in-service unit r over a future period $t \geq t^* \in \mathcal{T}$, where $\mathcal{T}$ is the time domain of interest.

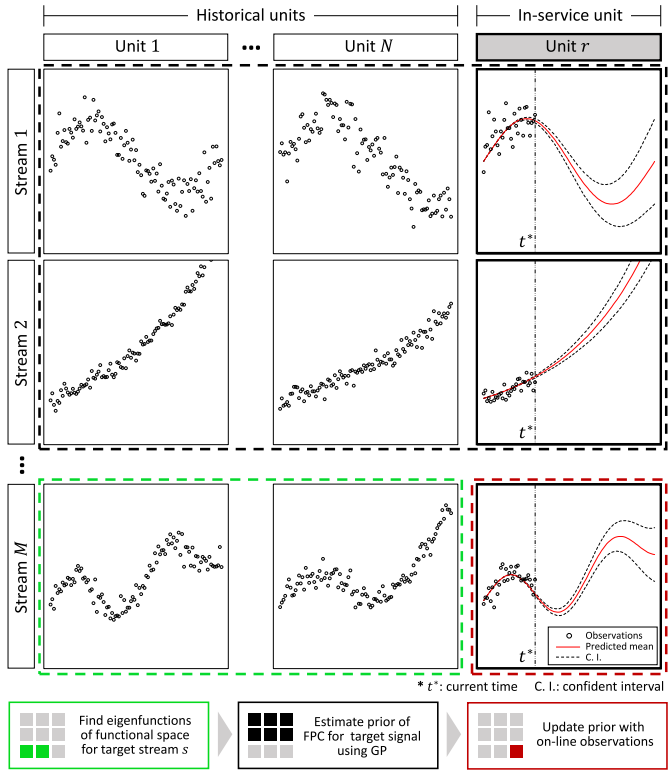

Fig. 1. Extrapolation of multistream longitudinal data for an in-service unit.

In mathematical notation, let $I = \{1, \ldots, N, r\}$ and $I^{\mathcal{H}} = \{1, \ldots, N\}$ be the respective index sets for all available units including the in-service unit r and the units in our historical dataset, respectively. For each unit $i \in I$, we have M streams of data, where $l \in L = \{1, \ldots, M\}$. For the ith unit, the history of observed data for a specific stream l is denoted as $X_i^{(l)}(t_{iu})$, where $\{t_{iu} : u = 1, \ldots, p_i^{(l)}\} \subset \mathcal{T}$ represents the observation time points and $p_i^{(l)}$ represents the number of observations for signal l of unit i. The underlying principle of our model is borrowing strength from a sample of curves $\{X_i^{(l)}(t) : i \in I^{\mathcal{H}}, l \in L\}$ to predict individual trajectories $X_r^{(l)}(t)$ over a future time period $t \geq t^* \in \mathcal{T}$. Without loss of generality, throughout the article, we focus on predicting stream $s \in L$, which we refer to as the target stream. Note that the target stream in Fig. 1 is stream M. To achieve this goal, we exploit functional principal component analysis (FPCA, [6]), which is a nonparametric tool for functional data analysis. Indeed, FPCA has recently drawn

Manuscript received February 4, 2020; revised June 13, 2020 and September 9, 2020; accepted October 11, 2020. This work was supported by NSF-Cyperphysical Systems program-1931950. Associate Editor: D. W. Coit. *(Corresponding author: Raed Kontar.)*

The authors are with the Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI 48109 USA (e-mail: seokhc@umich.edu; alkontar@umich.edu).

Color versions of one or more of the figures in this article are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TR.2020.3035084

increased attention due to its flexibility, uncertainty quantification capabilities, and the ability to handle sparse and irregular data [7]–[10]. However, advances in FPCA fall short of handling *multistream* data and *real-time* predictions.

Our overall framework is summarized at the bottom of Fig. 1. Specifically, historical signals $X_i^{(s)}(t)$, $i \in I^{\mathcal{H}}$, from the target stream s are decomposed into a linear combination of orthonormal eigenfunctions that form their functional space. The coefficients of the linear combination are called functional principal component (FPC) *scores*. With the assumption that the target signal of the in-service unit $X_r^{(s)}(t)$ lies in the same functional space, a proper estimation of FPC scores associated with $X_r^{(s)}(t)$ is required. Here, we propose to establish a prior on these FPC scores using information from streams $l \in L^{-s} = L \backslash \{s\}$. *However, such a prior needs to define a similarity measure across random functions rather than random variables*. To this end, we propose a Gaussian process (GP) prior for FPC scores of $X_r^{(s)}(t)$ using a functional semimetric that measures similarities of streams L^{-s} between historical units and the in-service unit r. The underlying principle is that unit $X_r^{(s)}(t)$ will exhibit more commonalities with historical units that exhibit similar trends in streams L^{-s}. For example, consider that stream s denotes degradation trajectories and L^{-s} denotes external factors such as temperature. Then, $X_r^{(s)}(t)$ will share more commonalities with a subset of historical signals $X_i^{(s)}(t)$, $i \in I^{\mathcal{H}}$, degrading under similar external factors (similar temperatures). This approach allows us to address heterogeneity in the data. Finally, an empirical Bayesian updating strategy is derived to update the established prior using real-time stream data obtained from the in-service unit.

In this article, we propose an FPCA-based model that provides *individualized predictions* for each stream in a multistream environment. Our model is able to automatically *account for heterogeneity* in the data and screen the sharing of information across different streams. We achieve this by introducing a GP prior based on a *functional semimetric* that measures similarities between streams of historical units and the in-service unit. We then derive a *computationally efficient Bayesian updating strategy* to update predictions when additional data are collected in real-time. We achieve the computational efficiency through building our model upon FPCA that features linearity and orthogonality. We demonstrate the advantageous features of our approach compared to state-of-the-art methods using both synthetic and simulation-based data. Here, we note that the derived functional semimetric can be of independent interest as it introduces a valid measure of similarities across random functions rather than random variables, where such measures are well-established.

The rest of this article is structured as follows. In Section II, we review a range of related work. In Section III, we briefly revisit the FPCA. In Section IV, we discuss our proposed model. In Section V, we discuss computational efficiency and time complexity of our model. Numerical experiments using synthetic data and simulation-based data are provided in Section VI. Finally, Section VII concludes this article. Technical proofs and a detailed code can be found in the appendix while a detailed code can be found online.[1]

II. Related Work

There has been extensive literature on the extrapolation of longitudinal signals under a single stream setting. However, literature has mainly focused on parametric models due to their computational efficiency and ease of implementation [11]–[14]. Unfortunately, parametric modeling is vulnerable to model misspecifications, and if the specified form is far from the truth, predictive results will be misleading. To address this issue, recent attempts at non-parametric approaches have been based on FPCA [15]–[17] or multivariate (multioutput) GPs (MGP) [18]–[22]. These studies show that such nonparametric approaches outperform parametric models in case where functional forms are complex and exhibit heterogeneity. Nevertheless, the foregoing works have dealt with only single stream cases. In this article, we exploit FPCA as MGP, despite its many recent success stories, faces intrinsic challenges in our setting: 1) Extrapolation: the empirical best linear unbiased predictor in the GP is an interpolant [23]. 2) Real-time inference: real-time updating in MGP implies refitting the model, where MGP (even with recent variational and low rank approximations) suffers from a nonconvex parameter space and high computational complexity, which limit its real-time applicability. 3) Multistream setting: To the best of our knowledge, there is no literature tackling MGPs in a multistream setting.

Multistream longitudinal data, also known as multivariate[2] longitudinal data analysis in the statistics literature, recently has drawn wide attention. The main objective of the analysis is to jointly model the covariance across multiple outcomes and units. Various methodologies such as pairwise fitting of mixed modeling [24], latent variable hidden Markov model accounting for heterogeneity [25], and nonparametric modeling [26], [27] are proposed. In practice, this sort of data often arises in medical-related applications, e.g., clinical drug trials [28], skin cancer chemoprevention [29], dietary intervention trial on breast cancer survivors [30], etc. The readers interested in a comprehensive review are referred to [31]. However, the above studies mainly focus on the analysis of multivariate longitudinal data rather than building a prediction model that extrapolates incomplete signals, which is important in many applications.

On the other hand, in engineering fields, the few literature that addressed the extrapolation in multistream settings have focused on data fusion approaches. Data fusion in this case refers to aggregating all streams into a single stream using fusion mechanisms. In health-related applications, this fused stream is coined as a health-index, which is often derived through a weighted combination of the M data streams [32], [33]. Such methods require regularly sampled observations and enforce

[1] [Online]. Available: https://github.com/UMDataScienceLab/FPCAmultistream.

[2] Remark that the term "multivariate" in longitudinal data analysis stands for a different notion in the GP-related literature as a multivariate GP corresponds to a multioutput GP.

strong parametric assumptions. An alternative data fusion approach includes multivariate FPCA [34]–[37]. However, since data fusion methods are operated by aggregating multistreams into a single or a smaller group of streams, they are not capable of predicting individual stream trajectories and thus have limited applications.

III. Brief Review of FPCA

From an FPCA perspective, longitudinal signals observed in a given time domain can be decomposed into a linear combination of orthonormal basis functions with corresponding FPC scores as coefficients. Therefore, FPCA can be regarded as a dimensionality reduction method in which a signal corresponds to a vector in a functional space defined by the basis functions. The basis functions are referred to as eigenfunctions. Assume that the longitudinal signals, over a given time domain $\mathcal{T}$, are generated from a square-integrable stochastic process $X(t)$ with its mean $\mathbb{E}[X(t)] = \mu(t)$ and covariance defined by a positive semidefinite kernel $G(t_1, t_2) = \text{Cov}(X(t_1), X(t_2))$ for $t_1, t_2 \in \mathcal{T}$. Using Mercer's theorem on $G(t_1, t_2)$, we have

$$G(t_1, t_2) = \sum_{k=1}^{\infty} \lambda_k \phi_k(t_1) \phi_k(t_2)$$

where $\phi_k(t)$ presents the kth eigenfunction of the linear Hilbert–Schmidt operator $G : L^2(\mathcal{T}) \to L^2(\mathcal{T}), G(f) = \int_{\mathcal{T}} G(t_1, t_2) f(t_1) dt_1$ ordered by the corresponding eigenvalues λ_k, $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$. The eigenfunctions $\phi_k(t)$ form a set of orthonormal basis in the Hilbert space $L^2(\mathcal{T})$. Following the Karhunen–Loéve decomposition, the centered stochastic process $X(t) - \mu(t)$ can then be expressed as

$$X(t) - \mu(t) = \sum_{k=1}^{\infty} \xi_k \phi_k(t) + \epsilon(t)$$

where $\xi_k = \int_{\mathcal{T}} (X(t) - \mu(t)) \phi_k(t)$ presents FPC scores associated with $\phi_k(t)$. The scores are uncorrelated normal random variables with zero-mean and variance λ_k; that is, $\mathbb{E}[\xi_k] = 0, \forall k \in \mathbb{N}$ and $\mathbb{E}[\xi_{k_1} \xi_{k_2}] = \delta_{k_1 k_2} \lambda_{k_1}, \forall k_1, k_2 \in \mathbb{N}$, where $\delta_{k_1 k_2}$ denotes the Kronecker delta. Also, $\epsilon(t)$ is additive Gaussian noise.

This idea of projecting signals onto a functional space spanned by eigenfunctions was first introduced by [38] for growth curves in particular. Basic principles [39], [40] and theoretical characteristics [41]–[43] were then developed. These ideas were expanded to longitudinal data settings in the seminal work of [44]. After that, the FPCA was applied and extended to a wide variety of applications, where multiple works tackled fast and efficient estimation of the underlying covariance surface [7], [8], [45]–[47].

IV. Extrapolation of Multistream Longitudinal Data

A. FPCA for Signal Approximation

Now, we discuss our proposed nonparametric approach for extrapolation of multistream longitudinal data. Hereon, unless there is ambiguity, we suppress subscripting the target stream with (s). Using historical signals $X_i(t)$, $i \in I^{\mathcal{H}}$, we decompose the target stream s as

$$X_i(t) = \mu(t) + f_i(t) + \epsilon(t), \quad i \in I^{\mathcal{H}} \tag{1}$$

where $f_i(t)$ represents random effects characterizing stochastic deviations across different historical signals in stream s and $\epsilon(t)$ denotes additive noise. We assume $f_i(t)$ and $\epsilon(t)$ are independent. Through an FPCA decomposition, we have that $f_i(t) = \sum_{k=1}^{\infty} \xi_{ik} \phi_k(t)$. This decomposition is an infinite sum, however, only a small number of eigenvalues are commonly significantly nonzero. For these values, the corresponding scores ξ_{ik} will also be approximately zero. Therefore, we approximate this decomposition as $f_i(t) = \sum_{k=1}^{K} \xi_{ik} \phi_k(t)$, where K is the number of significantly nonzero eigenvalues.

$$X_i(t) = \mu(t) + \sum_{k=1}^{K} \xi_{ik} \phi_k(t) + \epsilon(t), \quad i \in I^{\mathcal{H}}. \tag{2}$$

Here, we follow the standard estimation procedures in [8] and [48] to estimate the model parameters, where $\mu(t)$ is obtained by local linear smoothers [49], while K is selected to minimize the modified Akaike criterion. Now, given that the in-service unit r lies in the same functional space spanned by $\phi_k(t)$, our task is to find the individual distribution of ξ_{rk} using the partially observed multistream data from unit r. Specifically, we aim to find $X_r(t) = \mu(t) + \sum_{k=1}^{K} \xi_{rk} \phi_k(t) + \epsilon(t)$.

B. Estimation for Prior Distribution of FPC Scores Via GP

Next, we estimate the prior distribution of ξ_{rk} based on the key premise that $X_r(t)$ will behave more similarly to $X_i(t)$ for some units $i \in I^{\mathcal{H}}$ whose signals $X_i^{(l)}(t)$ for $l \in L^{-s}$ are similar to the corresponding signals of the in-service unit $X_r^{(l)}(t)$. To this end, for $k \in \{1, \ldots, K\}$, we model a functional relationship between $[\xi_{1k}, \ldots, \xi_{Nk}, \xi_{rk}]'$ and $X_i^{(l)}(t), \ldots, X_N^{(l)}(t), X_r^{(l)}(t)$ for $l \in L^{-s}$ as

$$[\boldsymbol{\xi}'_k, \xi_{rk}]' = \mathcal{G}_k \left(\mathbf{X}_1^{(-s)}(t), \ldots, \mathbf{X}_N^{(-s)}(t), \mathbf{X}_r^{(-s)}(t) \right) + \boldsymbol{\epsilon}_{\xi_k} \tag{3}$$

where $\mathbf{X}_i^{(-s)}(t) = [X_i^{(l)}(t)]_{l \in L^{-s}}$ for $i \in I$, $\boldsymbol{\xi}_k = [\xi_{ik}]_{i \in I^{\mathcal{H}}}$, and $\boldsymbol{\epsilon}_{\xi_k} \sim \mathcal{N}(\mathbf{0}, \sigma_k^2)$.

The idea here, is to model $\mathcal{G}_k$ as a GP with a covariance function defined by a similarity measure between the observed signals, i.e., a functional similarity measure. Specifically, for any $k \in \{1, \ldots, K\}$, the vector of FPC scores $[\boldsymbol{\xi}'_k, \xi_{rk}]'$ will follow a multivariate Gaussian distribution

$$\begin{bmatrix} \boldsymbol{\xi}_k \\ \xi_{rk} \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}_{N+1}, \begin{bmatrix} \mathbf{C}_k + \sigma_k^2 \mathbf{I}_N & \mathbf{c}_k \\ \mathbf{c}'_k & c_k^{(r)} \end{bmatrix} \right) \tag{4}$$

where $\mathbf{C}_k \in \mathbb{R}^{N \times N}$ is constructed such that its (i, j)th element is $h(i, j; \boldsymbol{\theta}_k)$ for $i, j \in I^{\mathcal{H}}$, $\mathbf{c}_k = [h(i, r, \boldsymbol{\theta}_k)]_{i \in I^{\mathcal{H}}}$, $c_k^{(r)} = h(r, r, \boldsymbol{\theta}_k)$, and h denotes a covariance function defined as

$$h(i, j; \boldsymbol{\theta}_k) = \alpha_k \exp\left(-\frac{1}{2} \sum_{l \in L^{-s}} \frac{\left\| X_i^{(l)}(t) - X_j^{(l)}(t) \right\|_l^2}{\left(\beta_k^{(l)} \right)^2} \right)$$

in which $\|\cdot\|_l$ is a semimetric providing a similarity measure across functions, and α_k and $\beta_k^{(l)}$ are nonnegative hyperparameters for streams $l \in L^{-s}$. For notational simplicity, we introduce $\boldsymbol{\theta}_k = [\alpha_k, \beta_k^{(1)}, \ldots, \beta_k^{(s-1)}, \beta_k^{(s+1)}, \ldots, \beta_k^{(M)}]'$.

To show the validity of the GP (4), we provide the following proposition.

Proposition 1: The matrix $\left[\begin{smallmatrix} \mathbf{C}_k + \sigma_k^2 \mathbf{I}_N & \mathbf{c}_k \\ \mathbf{c}_k' & c_k^{(r)} \end{smallmatrix}\right]$ corresponding to the covariance function $h(i, j; \boldsymbol{\theta}_k)$ a valid covariance matrix.

Proof: See Section A in Appendix. ■

One possible semimetric that represents the similarity between two signals can be derived based on FPCA. Let $\mathcal{T}_{t^*} \subset \mathcal{T}$ denote the time domain for observations up to t^*. Note that we define $\mathcal{T}_{t^*}$ since the signals of the in-service unit r are available only for $t \in \mathcal{T}_{t^*}$. For $i, j \in I, l \in L^{-s}$, and $t \in \mathcal{T}_{t^*}$, the semimetric based on FPCA for two signals $X_i^{(l)}(t)$ and $X_j^{(l)}(t)$ can be represented as

$$\left\|X_i^{(l)}(t) - X_j^{(l)}(t)\right\|_l = \sqrt{\sum_{k=1}^{K(l)} \left(\int_{\mathcal{T}_{t^*}} \left[X_i^{(l)}(t) - X_j^{(l)}(t)\right] \psi_k^{(l)}(t) dt \right)^2} \quad (5)$$

where $\psi_k^{(l)}(t)$ is kth eigenfunction derived by the FPCA on $X_i^{(l)}(t)$ for $i \in I$ and $t \in \mathcal{T}_{t^*}$, and $K(l)$ is the number of eigenfunctions. We would like to point out that $\int_{\mathcal{T}_{t^*}} [X_i^{(l)}(t) - X_j^{(l)}(t)]\psi_k^{(l)}(t)dt$ is the difference between the FPC scores of $X_i^{(l)}(t)$ and $X_j^{(l)}(t)$ associated with $\psi_k^{(l)}(t)$, which implies that this metric measures the Euclidean distance between two vectors composed of the corresponding FPC scores.

In order to optimize the hyperparameter $\boldsymbol{\Theta}_k = [\boldsymbol{\theta}_k', \sigma_k]'$ for the multivariate Gaussian distribution (4), we maximize the marginal log-likelihood function of $\boldsymbol{\xi}_k = [\xi_{1k}, \ldots, \xi_{Nk}]'$ given $\mathbf{X}_1^{(-s)}(t), \ldots, \mathbf{X}_N^{(-s)}(t)$. Let $\mathcal{X}$ denote the observations of signals $X_i^{(l)}(t)$ for $l \in L^{-s}$ units $i \in I^{\mathcal{H}}$, that is $\mathcal{X} = \{X_i^{(l)}(t) | t \in T_i^{(l)}, l \in L^{-s}, i \in I^{\mathcal{H}}\}$ where $T_i^{(l)} = \{t_{iu} | i \in I^{\mathcal{H}}, l \in L^{(-s)}, u = 1, \ldots, p_i^{(l)}\}$. Also, let z_{ik} denote the true underlying latent values corresponding to the FPC scores ξ_{ik} and let $\boldsymbol{z}_k = [z_{1k}, \ldots, z_{Nk}]'$. Then the marginal likelihood is given as

$$\begin{aligned} P(\boldsymbol{\xi}_k | \mathcal{X}, \boldsymbol{\Theta}_k) &= \int P(\boldsymbol{\xi}_k | \boldsymbol{z}_k, \mathcal{X}, \boldsymbol{\Theta}_k) P(\boldsymbol{z}_k | \mathcal{X}, \boldsymbol{\Theta}_k) d\boldsymbol{z}_k \\ &= \int \prod_{i \in I^{\mathcal{H}}} P(\xi_{ik} | z_{ik}, \mathcal{X}, \boldsymbol{\Theta}_k) P(\boldsymbol{z}_k | \mathcal{X}, \boldsymbol{\Theta}_k) d\boldsymbol{z}_k \end{aligned}$$

where $P(\xi_{ik} | z_{ik}, \mathcal{X}, \boldsymbol{\Theta}_k) = \mathcal{N}(0, \sigma_k^2)$ and $P(\boldsymbol{z}_k | \mathcal{X}, \boldsymbol{\Theta}_k) = \mathcal{N}(\mathbf{0}_N, \mathbf{C}_k)$. The second equality follows from the fact that the error is an additive Gaussian noise. Thus, $\prod_{i \in I^{\mathcal{H}}} P(\xi_{ik} | z_{ik}, \mathcal{X}, \boldsymbol{\Theta}_k) = \mathcal{N}(\mathbf{0}_N, \sigma_k^2 \mathbf{I}_N)$ and the log-likelihood of $P(\boldsymbol{\xi}_k | \mathcal{X}, \boldsymbol{\Theta}_k)$ is

$$\log P(\boldsymbol{\xi}_k | \mathcal{X}, \boldsymbol{\Theta}_k) = -\frac{1}{2} \left\langle \boldsymbol{\Xi}_k, (\mathbf{C}_k + \sigma_k^2 \mathbf{I}_N)^{-1} \right\rangle_{tr} - \log \left| \mathbf{C}_k + \sigma_k^2 \mathbf{I}_N \right| - \frac{n}{2} \log 2\pi$$

where $\langle \mathbf{A}, \mathbf{B} \rangle_{\text{tr}} = \text{trace}(\mathbf{AB})$ and $\boldsymbol{\Xi}_k = \boldsymbol{\xi}_k \boldsymbol{\xi}_k'$. As a consequence, the optimized hyperparameters denoted by $\boldsymbol{\Theta}_k^* = [\boldsymbol{\theta}_k^{*'}, \sigma_k^*]'$ are found by maximizing the marginal log-likelihood. More formally, we have

$$\boldsymbol{\Theta}_k^* = [\boldsymbol{\theta}_k^{*'}, \sigma_k^*]' = \text{argmax}_{\boldsymbol{\Theta}_k} \log P(\boldsymbol{\xi}_k | \mathcal{X}, \boldsymbol{\Theta}_k).$$

Following multivariate normal theory, the posterior predictive distribution of ξ_{rk}, given (4) and $\boldsymbol{\Theta}_k^*$, is derived as

$$P(\xi_{rk} | \boldsymbol{\xi}_k, \mathcal{X}, \boldsymbol{\Theta}_k^*) = \mathcal{N}\left(\hat{\xi}_{rk}, \hat{\sigma}_{rk}^2\right) \quad (6)$$

with

$$\begin{aligned} \hat{\xi}_{rk} &= \hat{\mathbf{c}}_k' (\hat{\mathbf{C}}_k + (\sigma_k^*)^2 \mathbf{I}_N)^{-1} \boldsymbol{\xi}_k \\ \hat{\sigma}_{rk}^2 &= \hat{c}_k^{(r)} - \hat{\mathbf{c}}_k' (\hat{\mathbf{C}}_k + (\sigma_k^*)^2 \mathbf{I}_N)^{-1} \hat{\mathbf{c}}_k. \end{aligned}$$

where $\hat{\mathbf{c}}_k, \hat{\mathbf{C}}_k$, and $\hat{c}_k^{(r)}$ indicate the corresponding covariances $\mathbf{c}_k, \mathbf{C}_k$, and $c_k^{(r)}$ established with the hyperparameter $\boldsymbol{\theta}_k^{*'}$.

Here, we note that for each $k \in \{1, \ldots, K\}$, we can derive $\hat{\xi}_{rk}$ and $\hat{\sigma}_{rk}^2$ using an independent GP as the FPC scores from different orthonormal basis functions are uncorrelated. This facilitates scalability of computation as for different k, we can derive $\{\hat{\xi}_{rk}\}$ and $\{\hat{\sigma}_{rk}^2\}$ in parallel. As shown in the computational complexity derivations in Section V, this aspect is important specifically in a real-time environment, where predictions need to be continuously updated.

Now combining (2) and (6), we obtain the predictive mean $\hat{X}_r(t)$ and variance $\hat{\sigma}_r^2(t)$ of $X_r(t)$ as follows:

$$\begin{aligned} \hat{X}_r(t) &= \mu(t) + \sum_{k=1}^{K} \hat{\xi}_{rk} \phi_k(t) \\ \hat{\sigma}_r^2(t) &= \sigma_\mu^2(t) + \sum_{k=1}^{K} \hat{\sigma}_{rk}^2 \phi_k^2(t) + \sigma_\epsilon^2. \end{aligned} \quad (7)$$

Here, note that $\mu(t)$, $\sigma_\mu^2(t)$, and σ_ϵ^2 are model parameters corresponding to the estimated FPCA model in (2), where σ_ϵ^2 denotes the estimated variance of $\epsilon(t)$. Here, we recall that $\hat{X}_r(t) = \hat{X}_r^{(s)}(t)$ as the index (s) is dropped for the target stream.

C. Empirical Bayesian Updating With Online Data

In the previous section, we derive a prior for $\hat{X}_r(t)$ and $\hat{\sigma}_r^2(t)$ using data observed from streams $l \in L^{-s}$. Here, we develop an empirical Bayesian approach to update $\hat{X}_r(t)$ and $\hat{\sigma}_r^2(t)$ given the target stream (s) observations from the in-service unit r. Specifically, given the prior distributions $P(\xi_{rk}) = \mathcal{N}(\hat{\xi}_{rk}, \hat{\sigma}_{rk}^2)$ for each k and given the observations $\mathbf{X}_r(\mathbf{t})$ at $\mathbf{t} = [t_{r1}, \ldots, t_{rp_r}]'$, the posterior $P(\xi_{rk} \mid \mathbf{X}_r(\mathbf{t}))$ is given in Proposition 2.

Proposition 2: Given that $X_r(t) = \mu(t) + \sum_{k=1}^{K} \xi_{rk} \phi_k(t) + \epsilon(t)$, where the prior distribution of ξ_{rk} is $\mathcal{N}(\hat{\xi}_{rk}, \hat{\sigma}_{rk}^2)$, and the FPC scores are pairwise independent. Then, the posterior distribution of the FPC scores, such that $\xi_{rk}^* = P(\xi_{rk} \mid \mathbf{X}_r(\mathbf{t}))$, is given as

$$[\xi_{r1}^*, \ldots, \xi_{rK}^*]' \sim \mathcal{N}(\boldsymbol{\xi}^*, \boldsymbol{\Sigma}^*)$$

where

$$\boldsymbol{\xi}^* = \boldsymbol{\Sigma}^* \left(\boldsymbol{\Sigma}_0^{-1} \boldsymbol{\mu}_0 + \frac{1}{\sigma_\epsilon^2} \boldsymbol{\Phi}(\mathbf{t})' (\mathbf{X}_r(\mathbf{t}) - \boldsymbol{\mu}(\mathbf{t})) \right)$$

$$\boldsymbol{\Sigma}^* = \left(\frac{1}{\sigma_\epsilon^2} \boldsymbol{\Phi}(\mathbf{t})' \boldsymbol{\Phi}(\mathbf{t}) + \boldsymbol{\Sigma}_0^{-1} \right)^{-1}$$

with

$$\boldsymbol{\mu}_0 = \left[\hat{\xi}_{r1}, \ldots, \hat{\xi}_{rK} \right]'$$

$$\boldsymbol{\Sigma}_0 = diag\left(\hat{\sigma}_{r1}^2, \ldots, \hat{\sigma}_{rK}^2 \right)$$

$$\mathbf{X}_r(\mathbf{t}) = [X_r(t_{r1}), \ldots, X_r(t_{rp_r})]'$$

$$\boldsymbol{\mu}(\mathbf{t}) = [\mu(t_{r1}), \ldots, \mu(t_{rp_r})]'$$

$$\boldsymbol{\Phi}(\mathbf{t}) = \begin{bmatrix} \phi_1(t_{r1}) & \ldots & \phi_K(t_{r1}) \\ \vdots & \ddots & \vdots \\ \phi_1(t_{rp_r}) & \ldots & \phi_K(t_{rp_r}) \end{bmatrix}.$$

Proof: See Section B in Appendix. ■

Based on the updated FPC scores for in-service unir r, the posterior predicted mean $\hat{X}_r^*(\tilde{t})$, of $X_r(t)$ for any future time point $\tilde{t} \geq t^*$, where $\tilde{t} \in \mathcal{T}$ is given as

$$\hat{X}_r^*(\tilde{t}) = \mu(\tilde{t}) + \sum_{k=1}^{K} \xi_{rk}^* \phi_k(\tilde{t}).$$

Similarly, the posterior variance $(\hat{\sigma}_r^*(\tilde{t}))^2$ can be computed as

$$\hat{\sigma}_r^{*2}(\tilde{t}) = \sigma_\mu^2(t) + \sum_{k_1=1}^{K} \sum_{k_2=1}^{K} \left(\boldsymbol{\Sigma}^* \right)_{k_1 k_2} \phi_{k_1}(\tilde{t}) \phi_{k_2}(\tilde{t}) + \sigma_\epsilon^2$$

where $(\boldsymbol{\Sigma}^*)_{k_1 k_2}$ indicates the (k_1, k_2)th element of the covariance matrix $\boldsymbol{\Sigma}^*$.

Despite our focus on the target stream s, we note that our framework can predict every individual stream for the in-service unit r. This ability to provide individualized predictions is a key feature of the proposed methodology compared to the data fusion literature that predicts a single aggregated signal. Further, one differentiating factor is that we allow irregularly sampled data from each stream, where time points of each signal $\{t_{iu}^{(l)} : u = 1, \ldots, p_i^{(l)}\} \subset \mathcal{T}$ do not need to be identical or regularly spaced across streams. Indeed, such situations are quite common in practice because most multistream data are gathered from different types of sensors. Therefore, the proposed approach is applicable to a wide array of practical situations.

V. Computational Complexity

Since we work in the regime of streaming data, the frequency with which we receive data is very high. Moreover, we collect data from multisensors, which usually leads to a larger scale data than of a single sensor system. Due to this, our model needs to be efficient in terms of the time taken to make each update.

Here, we note that the proposed Bayesian updating can be done efficiently due to two important features of FPCA: *linearity* and *orthogonality*. Despite being nonparametric, the FPCA features a linear combination of orthogonal eigenfunctions and their FPC scores. Thus, online updates can be done in closed form as Proposition 2. Furthermore, since the eigenfunctions are orthogonal, the estimation procedure of the GP priors placed on FPC scores in Section IV-B is parallelizable. This implies that reestimating and updating the model in a real-time setting can be done instantaneously.

More rigorous complexity analysis is provided as follows. With the assumption that all signals from M streams have Q observations, the complexity of multivariate FPCA for multistream data is $\mathcal{O}(M^2NQ^2 + M^3Q^3)$ [35]. In our model, the computationally expensive steps are the FPCA for the target stream (see Section IV-A) and the implementation of GP for estimating the FPC scores (see Section IV-B). Following [47], the complexity of the former is $\mathcal{O}(QN^2 + N^3)$. While complexity of a GP with an $N \times N$ covariance matrix is $\mathcal{O}(N^3)$ [50]. Given that we implement K independent GP models the complexity of estimating the FPC scores is $\mathcal{O}(KN^3)$. Combining the above observations, we conclude that the complexity of our procedure is $\mathcal{O}(QN^2 + N^3 + KN^3)$. Typically, we have that $M, N, K \ll Q$, also, in real-time Q is increasing. Thus, our model is clearly more efficient than multivariate FPCA and applicable in a real-time streaming environment.

VI. Numerical Case Study

A. General Settings

In this section, we discuss the general settings used to assess the proposed model, denoted as FPCA-GP. We evaluate the model by performing experiments with both synthetic and simulation-based data. We report the prediction accuracy at varying time points t^* for the partially observed unit r. Specifically, for the time domain $\mathcal{T} = [a, b]$, we assume that the online signals from the in-service unit r are partially observed in the range of $[a, t^* = a + \gamma(b - a)]$, referred as γ-observation. We set $\gamma = 25\%, 50\%$, and 75% for every case study. In the extrapolation interval $[t^*, b]$, we use the mean absolute error (MAE) between the true signal value $X_r^T(t)$ and its predicted value $\hat{X}_r^{(s)}(t)$ at U evenly spaced test points (denoted as t_u for $u = 1, \ldots, U$) as the criterion to evaluate our prediction accuracy.

$$\text{MAE} = \frac{1}{U} \sum_{u=1}^{U} \left| \hat{X}_r^{(s)}(t_u) - X_r^T(t_u) \right|, \; t_u \in [t^*, b]. \quad (8)$$

We report the distribution of the errors across G repetitions using a group of boxplots representing the MAE for the testing unit r at different γ-observation percentiles. Further, we benchmark our method with three other reference methods for comparison: 1) The FPCA approach for single stream settings denoted as FPCA-B. In this method, we only consider the target stream s [15], [20]. Note that we incorporate our Bayesian updating procedure to update predictions as new data is observed. 2) The empirical Bayes random effects model with a general polynomial function, whose degree is determined through an Akaike information criteria [11], [51], [52]. We denote this method as EBRE. The EBRE model intrinsically applies a Bayesian updating scheme as more data are obtained from the in-service unit. 3) The sparse convolved MPG denoted as SMGP [18]. We

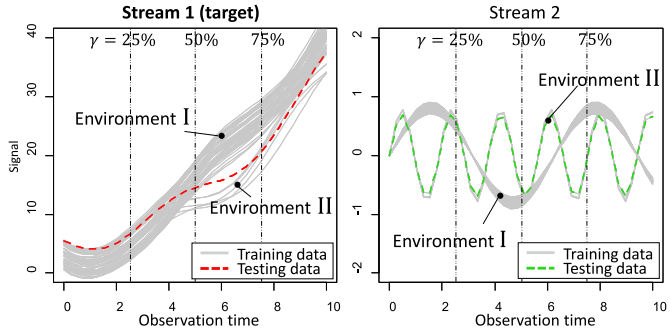

Fig. 2. Illustration of generated true curves (90% heterogeneity case).

choose the SMGP among various MGP-based models (e.g., [20], [22]) because SMGP can scale to our dataset. We follow model settings for the SMGP given by [18]. Note that the SMGP is also performed only on a single target stream. Detailed codes for the reference methods are included in online supplement.

B. Numerical Study With Synthetic Data

First, we show the numerical results of the proposed model performed on synthetic data. For this experiment, we assume that two streams $l \in \{1, 2\}$ of data are observed from two different sensors embedded in each unit. The target stream of interest is $l = 1$. To generate signals possessing heterogeneity, we suppose there are two separating environments, denoted by environment I and II. We generate signals for each unit using different underlying functions depending on which environment the unit is in. This is illustrated in Fig. 2. As shown in this figure, the underlying trend of the target stream ($l = 1$) will vary under different profiles of stream $l = 2$. To relate this setting with real-world application, consider $l = 1$ as the degradation level and $l = 2$ as the temperature profile. Then, from Fig. 2, we have that units operating under different temperature profiles will exhibit different trends.

We generate a training set of $N = 50$ units and one testing unit r whose signals are partially observed. Also, we repeat the experiment $G = 100$ times. Historical units are operated in either environment I or II, whereas the in-service unit is operated in environment II. The population of historical units is created according to three levels of heterogeneity: 1) *0% heterogeneity* where all units in the historical database are operated under environment II (similar to that of the testing unit). 2) *50% heterogeneity* where 25 units are distributed to each environment. 3) *90% heterogeneity* where only five units are assigned to environment II. Conducting the experiments across a homogeneous setting and a heterogeneous setting, where the in-service unit belongs to the minority group with only 10% ratio, will allow us to investigate the robustness of our approach.

For units in environment I, the signals from respective streams $l = 1, 2$ are generated according to $X_i^{(1)}(t) = 0.3t^2 - 2\sin(w_{1,i}\pi t) + w_{2,i}^{\mathrm{I}}$ and $X_i^{(2)}(t) = 2w_{1,i}\sin t$, where $w_{1,i} \sim \mathcal{N}(0.4, 0.03^2)$ and $w_{2,i}^{\mathrm{I}} \sim \mathcal{U}(0, 5)$, where $\mathcal{U}$ denotes the uniform distribution. For units in environment II, we generate the signals as $X_i^{(1)}(t) = 0.3t^2 - 2\sin(w_{1,i}\pi t^{0.85}) + 3(\arctan(t - 5) + \frac{\pi}{2}) + w_{2,i}^{\mathrm{II}}$ and $X_i^{(2)}(t) = 2w_{1,i}\sin 0.3t$, where $w_{2,i}^{\mathrm{II}} \sim \mathcal{U}(1.5, 6.5)$. Measurement error $\epsilon_1 \sim \mathcal{N}(0, 0.3^2)$ for stream 1 and $\epsilon_2 \sim \mathcal{N}(0, 0.05^2)$ for stream 2 are assumed.

Fig. 2 illustrates training signals $X_i^{(1)}(t)$ in the case of 90% heterogeneity. It is crucial to note that at early stages (ex: $\gamma = 25\%$), it is hard to distinguish between the two different trends in stream $l = 1$. We model that on purpose to check if our model can leverage information from stream $l = 2$ to uncover the underlying heterogeneity at early stages. *This in fact is a common feature in many health-related applications, as many diseases remain dormant at early stages and it is only through measuring other factors we can predict their evolution early on.*

The results are illustrated in Figs. 3, 4, and Table I. Based on the results, we can obtain some important insights. First, the FPCA-GP clearly outperforms the FPCA-B. This is specifically obvious at early stages ($\gamma = 25\%$) and when the data exhibits heterogeneity (90% and 50% heterogeneity). This confirms the ability of our model to borrow strength from information across different streams to discern the heterogeneity and enhance predictive accuracy at early stages. This result is very motivating specifically since at $\gamma = 25\%$ data from in-service unit r is sparse and all signals in stream s have similar behavior, which makes it hard to uncover future heterogeneity. It further implies that the FPC scores of the testing unit are appropriately estimated by the proposed approach, as shown in the first column of Fig. 3. From the figure, we observe that the estimated prior mean from the FPCA-GP appropriately follows the signals in environment II, whereas the prior mean from the FPCA-B follows the signals in environment I, which is the majority. Second, as expected, prediction errors significantly decrease as the percentiles increase. Thus, our Bayesian updating framework is able to efficiently utilize new collected data and provide more accurate predictions as t^* increases. Third, the results show that EBRE behaved the worst and its predictions accuracy merely decreases at later stages. This result illustrates the vulnerabilities of parametric modeling and demonstrates the ability of our nonparametric modeling to avoid model misspecifications. Fourth, the results confirm that even in the case, where other streams have no effect on the target stream (*0% heterogeneity*) the FPCA-GP is competitive compared to FPCA-B. This highlights the robustness of the FPCA-GP. Finally, the FPCA-GP significantly outperforms the SMGP in both predictive performance and computation time. This shows that an MGP-based model is vulnerable even in the case of the moderate number of outputs ($N = 50$) as they are required to optimize the negative likelihood in a high-dimensional nonconvex space [22]. Furthermore, the results in Table I showcase that the FPCA-GP performs considerably better than SMGP in a computational perspective even though the SMGP involves the target stream only. This supports the discussion on complexity in Section V.

C. Numerical Study With Simulation-Based Data

In this section, we discuss the numerical study using simulation-based data provided by the National Aeronautics

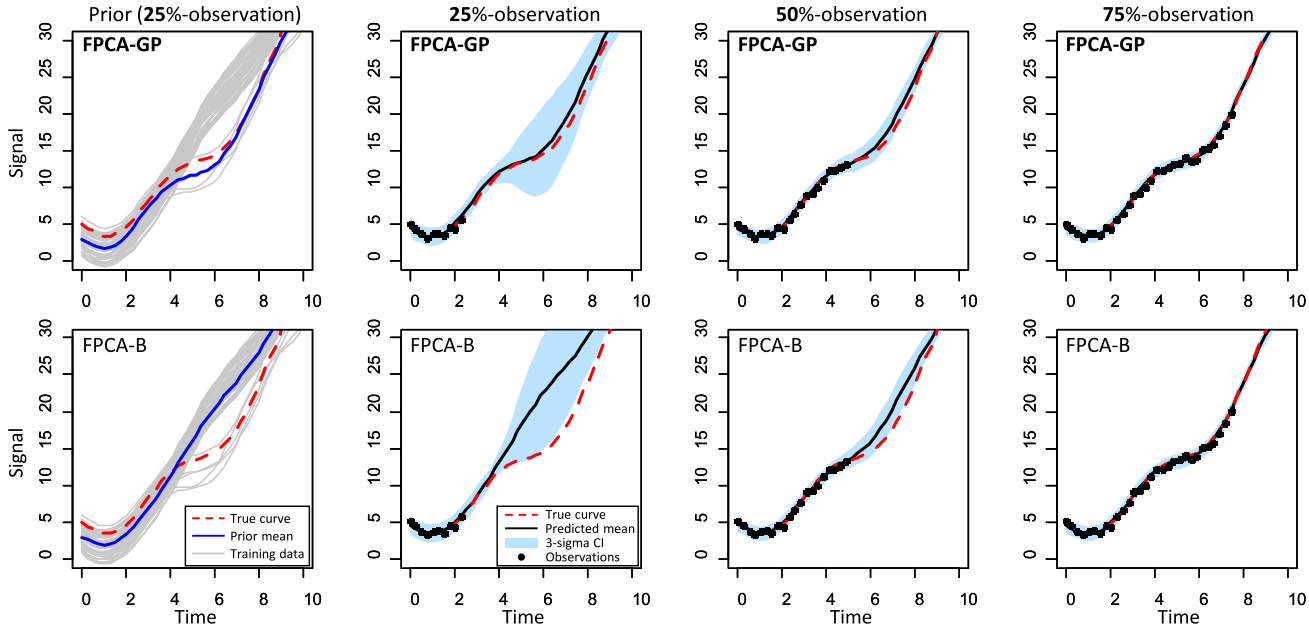

Fig. 3. Illustration of the FPCA-GP and FPCA-B prediction (90% heterogeneity case). The first column illustrates the respective prior mean of FPCA-GP and FPCA-B before updating in the case of 25%-observation.

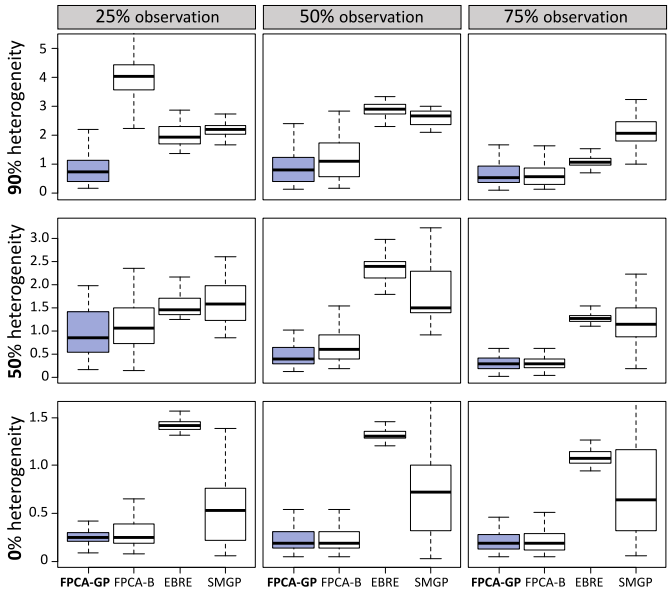

Fig. 4. Box plots of MAEs for comparative models for synthetic data.

TABLE I
COMPUTATION TIME (SECONDS) WITH DIFFERENT HETEROGENEITY RATIO (H-RATIO)

H-ratio	**FPCA-GP**	FPCA-B	EBRE	SMGP
25%	0.69	0.15	0.2	110.24
(std.)	(0.08)	(0.03)	(0.04)	(91.73)
50%	0.68	0.15	0.21	117.34
(std.)	(0.08)	(0.03)	(0.03)	(96.81)
75%	0.69	0.15	0.2	115.56
(std.)	(0.11)	(0.03)	(0.03)	(92.75)
Avg.	0.69	0.15	0.20	114.38

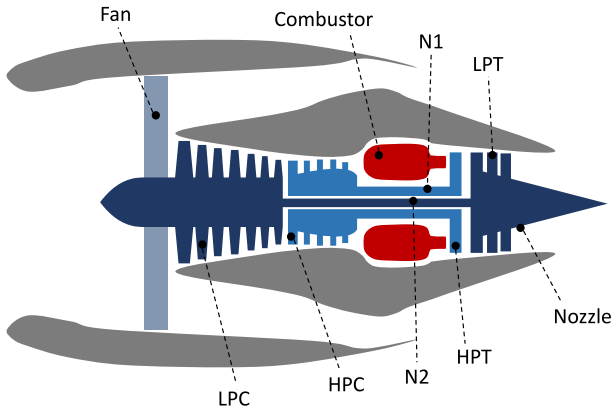

Fig. 5. Schematic diagram of turbofan engine (reproduced figure of the original imagein [55]).

and Space Administration (NASA). The dataset contains degradation signals collected from multiple sensors on an aircraft turbofan engine. This dataset was generated from a simulation model, developed in MATLAB Simulink, called commercial modular aero-propulsion system simulation (C-MAPSS). This system simulates degradation signals from multiple-sensors, installed in several components of an aero turbofan engine, under a variety of environmental conditions. The list of the components includes Fan, LPC, HPC, and LPT, and is illustrated in Fig. 5. The dataset is composed of 21 sensor streams from 100 training and 100 testing units. The sensors are listed in Table II. Refer to [53] for more details about turbofan engine data. The dataset is available at [54]. Following the analysis of [32], we preprocess the dataset as follows. We select the 11 most crucial streams: T24, T50, P30, Nf, Ps30, phi, NRf, BPR, htBleed, W31, and

TABLE II
LIST OF SENSORS ON TURBOFAN ENGINE

Index	Symbol	Description	Units
1	T2	Total temperature at fan inlet	°R
2	T24	Total temperature at LPC outlet	°R
3	T30	Total temperature at HPC outlet	°R
4	T50	Total temperature at LPT outlet	°R
5	P2	Pressure at fan inlet	psia
6	P15	Total pressure in bypass-duct	psia
7	P30	Total pressure at HPC outlet	psia
8	Nf	Physical fan speed	rpm
9	Nc	Physical core speed	rpm
10	epr	Engine pressure ratio (P50/P2)	-
11	Ps30	Static pressure at HPC outlet	psia
12	phi	Ratio of fuel flow to Ps30	pps/psi
13	NRf	Corrected fan speed	rpm
14	NRc	Corrected core speed	rpm
15	BPR	Bypass ratio	-
16	farB	Burner fuel-air ratio	-
17	htBleed	Bleed enthalpy	-
18	Nf_dmd	Demanded fan speed	rpm
19	PCNfR_dmd	Demanded corrected fan speed	rpm
20	W31	HPT coolant bleed	lbm/s
21	W32	LPT coolant bleed	lbm/s

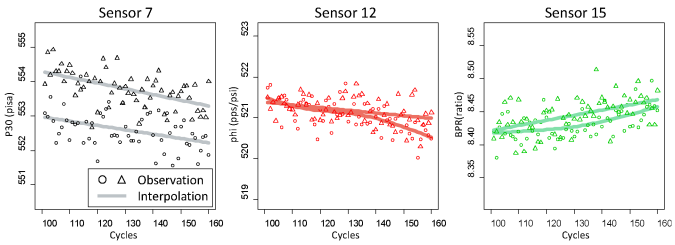

Fig. 6. Selective examples for degradation signals from turbofan engine data. Signals from two example units are illustrated.

W32. Some signals from these streams are shown in Fig. 6. To remove the nondefective phase of the signals, we truncate the time range (0,100] and predict the testing signal over the time range $(100, 160]$.

Table III demonstrates that the MAE results of stream 2, 4, 7, 8, 12, and 15. Note that we choose these streams since they have shown to have the largest impact on failure [35]. Note that we include the standard deviation of MAE across the testing units. Table IV presents the MAE results of the worst cases for the chosen sensor. We provide the results because it is important in real-applications to achieve robust performance for every possible unit.

The results clearly show that our approach is far more superior than benchmarks for the simulation-based data. For all provided cases, the mean of MAE for the FPCA-GP is less than that of the FPCA-B and the SMGP. Once again this highlights the importance of leveraging information from all streams of the data. Another important insight from this study is that our model was able to outperform the EBRE even though the curves from Fig. 6 seem to exhibit a clear parametric trend. This further highlights the robustness of our method and its ability to safeguard against parametric misspecifications. Finally, we find that FPCA-GP outperforms the other benchmark models even in the worst predictions, except for a few cases.

VII. CONCLUSION

In this article, we developed a nonparametric statistical model that can extrapolate individual signals in a multistream data setting. Using both synthetic and simulation-based data, we demonstrated our model's ability to borrow strength across all streams of data, predict individual streams, account for heterogeneity, and provide accurate real-time predictions, where an empirical Bayesian approach updates our predictor as new data is observed in real-time. Complexity analysis was discussed and this showed that the method enjoys computational efficiency comparing to existing multivariate FPCA methods. The proposed method outperformed benchmarks in both synthetic and simulation-based data. In particular, the numerical results showed that our model well revealed the future behavior of signals even at the early stages, which implies a huge applicability in various fields.

We conclude by discussing possible directions for future work. The current approach is built upon FPCA, which may suffer difficulty to deal with multivariate inputs. For example, one may have spatial-temporal data as inputs of the target stream in a multistream system. Tackling such a case will be an interesting problem and we leave it as future work. Another challenge inherited from FPCA is that the time domain of observations across units within a stream is assumed to be identical. A promising direction regarding the issue for future work can be to propose an extrapolation method for the data in multistream environments, especially for the case, where time domains vary across the units.

APPENDIX

A. Proof of Proposition 1

Proof: Hereon, unless there is ambiguity, we suppress superscripts (l) on $X_i^{(l)}(t)$ and $\beta_k^{(l)}$ and let $K = K(l)$ in the proof where this does not lead to confusion. Also, let $\langle \cdot, \cdot \rangle$ and $\langle \cdot, \cdot \rangle_{\mathcal{F}}$ denote the inner product in the Hilbert space and an inner product space $\mathcal{F}$, respectively.

Note that the matrix $\begin{bmatrix} \mathbf{C}_k + \sigma_k^2 \mathbf{I}_N & \mathbf{c}_k \\ \mathbf{c}_k' & c_k^{(r)} \end{bmatrix}$ is a valid covariance matrix if and only if $h(i, j, \boldsymbol{\theta}_k)$ is a valid kernel. Let us define the kernels

$$\bar{h}_l(i, j, \boldsymbol{\theta}_k) = \exp\left(- \frac{\left\| X_i(t) - X_j(t) \right\|_l^2}{2\beta_k^2} \right),\ l \in \{1, \ldots, L^{-s}\}. \tag{9}$$

Note that $h(i, j, \boldsymbol{\theta}_k)$ is the sum of the product of nonnegative hyperparameter α_k and $\bar{h}_l(i, j, \boldsymbol{\theta}_k)$ for $l \in \{1, \ldots, L^{-s}\}$. Following kernel theory [56], $h(i, j, \boldsymbol{\theta}_k)$ is a valid kernel if $\bar{h}_l(i, j, \boldsymbol{\theta}_k)$ for $l \in \{1, \ldots, L^{-s}\}$ are valid. Let us now prove that $\bar{h}_l(i, j, \boldsymbol{\theta}_k)$ is valid.

Definition 3: A kernel is a two-argument real-valued function over $\mathcal{A} \times \mathcal{A}$ ($\kappa : \mathcal{A} \times \mathcal{A} \to \mathbb{R}$) such that for any $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{A}$

$$\kappa(\mathbf{x}_1, \mathbf{x}_2) = \left\langle \omega(\mathbf{x}_1), \omega(\mathbf{x}_2) \right\rangle_{\mathcal{F}} \tag{10}$$

for some inner-product space $\mathcal{F}$ such that $\forall \mathbf{x}_1 \in \mathcal{A}, \omega(\mathbf{x}_1) \in \mathcal{F}$.

To prove that κ is a valid kernel, showing that a mapping ω exists that gives (10) is sufficient. We thus aim to show that our proposed kernel $\bar{h}$ can induce the mapping.

TABLE III
MEAN AND STANDARD DEVIATION (STD.) OF COMPARATIVE MODELS PERFORMED ON THE NASA DATASET

	Sensor 2 (T24)			Sensor 4 (T50)			Sensor 7 (P30)			Sensor 8 (Nf)			Sensor 12 (phi)			Sensor 15 (BPR)		
Model	25%	50%	75%	25%	50%	75%	25%	50%	75%	25%	50%	75%	25%	50%	75%	25%	50%	75%
FPCA-GP	**24.65**	**24.79**	**25.46**	**3.26**	**3.21**	**3.19**	**31.72**	**31.94**	**32.13**	**2.56**	**2.52**	**2.47**	**24.39**	**25.07**	**25.02**	**1.62**	**1.62**	**1.57**
(std.)	(2.12)	(2.76)	(4.26)	(0.36)	(0.42)	(0.48)	(3.41)	(4.85)	(4.69)	(0.32)	(0.41)	(0.44)	(2.18)	(3.62)	(4.17)	(0.18)	(0.19)	(0.33)
FPCA-B	26.62	26.49	26.97	3.49	3.37	3.31	35.91	34.56	33.69	2.83	2.70	2.59	27.93	26.76	26.04	1.76	1.75	1.63
(std.)	(2.97)	(3.18)	(4.06)	(0.45)	(0.56)	(0.54)	(4.76)	(5.89)	(5.31)	(0.35)	(0.56)	(0.53)	(3.80)	(4.19)	(4.66)	(0.28)	(0.31)	(0.36)
EBRE	26.62	26.62	26.47	3.51	3.38	3.34	36.06	34.99	34.42	2.81	2.72	2.61	27.54	26.85	26.03	1.79	1.77	1.65
(std.)	(2.82)	(3.22)	(3.82)	(0.45)	(0.59)	(0.55)	(4.86)	(6.05)	(5.60)	(0.36)	(0.59)	(0.55)	(3.49)	(4.33)	(4.79)	(0.28)	(0.32)	(0.37)
SMGP	29.58	29.80	29.28	4.66	5.09	5.84	43.85	49.09	51.47	3.28	3.30	3.22	34.48	35.86	34.03	2.06	2.01	2.05
(std.)	(4.75)	(5.38)	(8.23)	(1.24)	(1.66)	(2.19)	(11.05)	(16.57)	(21.43)	(0.74)	(0.88)	(1.23)	(9.38)	(10.20)	(9.63)	(0.42)	(0.47)	(0.65)

All values except for the sensor 4 are scaled by $\times 10^{-2}$. The best result in each case is boldfaced. Note that the increasing of MAE along with the case of 25%, 50%, and 75%-observation does not mean the predictive accuracy decreases, since the number of time point Q is different.

TABLE IV
MAEs OF THE WORST CASES OF COMPARATIVE MODELS PERFORMED ON THE NASA DATASET

	Sensor2 (T24)			Sensor 4 (T50)			Sensor 7 (P30)			Sensor 8 (Nf)			Sensor 12 (phi)			Sensor 15 (BPR)		
	25%	50%	75%	25%	50%	75%	25%	50%	75%	25%	50%	75%	25%	50%	75%	25%	50%	75%
FPCA-GP	**29.28**	**30.45**	**34.49**	**4.46**	**4.76**	4.41	**40.11**	**43.46**	**41.42**	**3.13**	**3.18**	**3.39**	**30.01**	**33.85**	**39.49**	**1.97**	**2.01**	2.43
FPCA-B	33.09	33.69	34.63	4.63	5.12	**4.32**	44.15	46.84	48.09	3.71	3.81	3.66	34.14	34.70	39.50	2.55	2.40	2.39
EBRE	33.41	33.92	35.70	4.56	5.08	4.43	44.04	47.36	50.56	3.76	3.91	3.93	35.59	36.70	40.61	2.62	2.40	**2.35**
SMGP	40.52	40.74	52.19	7.09	8.68	11.35	70.40	90.19	113.61	5.32	6.62	7.69	60.92	65.00	57.48	3.19	3.74	4.15

All values except for the sensor 4 are scaled by $\times 10^{-2}$. The best result in each case is boldfaced.

By the equation (5), the semimetric $\big\|X_i(t) - X_j(t)\big\|_l$ can be presented as

$$\begin{aligned}&\big\|X_i(t) - X_j(t)\big\|_l \\ &= \sqrt{\sum_{k=1}^{K}\left(\int_{\mathcal{T}_{t^*}} [X_i(t) - X_j(t)]\psi_k(t)dt\right)^2} \\ &= \sqrt{\sum_{k=1}^{K}\left(\int_{\mathcal{T}_{t^*}} X_i(t)\psi_k(t)dt - \int_{\mathcal{T}_{t^*}} X_j(t)\psi_k(t)dt\right)^2}.\end{aligned} \tag{11}$$

Note that $\int_{\mathcal{T}_{t^*}} X_i(t)\psi_k(t)dt \in \mathbb{R}$ is an inner product. Consider a real-valued vector $\mathbf{x}(i) = [x_1(i), \ldots, x_K(i)]' \in \mathbb{R}^K$, where $x_k(i) = \int_{\mathcal{T}_{t^*}} X_i(t)\psi_k(t)dt$. Then, the semimetric (11) can be represented as the Euclidean norm $\big\|\mathbf{x}(i) - \mathbf{x}(j)\big\|_2$, which allows us to represent the function $\bar{h}(i, j, \boldsymbol{\theta}_k)$ as

$$\bar{h}_l(i, j, \boldsymbol{\theta}_k) = \exp\left(-\frac{\big\|\mathbf{x}(i) - \mathbf{x}(j)\big\|_2^2}{2\beta_k^2}\right). \tag{12}$$

Without loss of generality, let $\beta_k^2 = 1$. Then

$$\begin{aligned}&\bar{h}_l(i, j; \boldsymbol{\theta}_k) \\ &= \exp\left(-\frac{\big\|\mathbf{x}(i) - \mathbf{x}(j)\big\|_2^2}{2}\right) \\ &= \exp\left(-\frac{\big\|\mathbf{x}(i)\big\|_2^2}{2}\right) \cdot \exp\left(-\frac{\big\|\mathbf{x}(j)\big\|_2^2}{2}\right) \\ &\quad \cdot \exp\Big(\big\langle \mathbf{x}(i), \mathbf{x}(j)\big\rangle\Big) \\ &= \exp\left(-\frac{\big\|\mathbf{x}(i)\big\|_2^2}{2}\right) \cdot \exp\left(-\frac{\big\|\mathbf{x}(j)\big\|_2^2}{2}\right) \\ &\quad \cdot \sum_{n=0}^{\infty} \frac{\big\langle \mathbf{x}(i), \mathbf{x}(j)\big\rangle^n}{n!} \\ &= \sum_{n=0}^{\infty}\left(\left\langle \sqrt[n]{\frac{\boldsymbol{\zeta}(i)}{n!}}\mathbf{x}(i), \sqrt[n]{\frac{\boldsymbol{\zeta}(j)}{n!}}\mathbf{x}(j)\right\rangle\right)^n \\ &= \sum_{n=0}^{\infty} \sum_{\sum_{k=1}^{K} n_k = n} \left(\sqrt[n]{\frac{\boldsymbol{\zeta}(i)}{n!}}\binom{n}{n_1, \ldots, n_K}^{1/2}\right. \\ &\quad \times (x_1(i))^{n_1} \ldots (x_K(i))^{n_K} \\ &\quad \left. \cdot \sqrt[n]{\frac{\boldsymbol{\zeta}(j)}{n!}}\binom{n}{n_1, \ldots, n_K}^{1/2} (x_1(j))^{n_1} \ldots (x_K(j))^{n_K}\right) \\ &= \Big\langle \boldsymbol{\Psi}(i), \boldsymbol{\Psi}(j)\Big\rangle\end{aligned}$$

where

$$\boldsymbol{\zeta}(i) = \exp\left(-\frac{1}{2}\big\|\mathbf{x}(i)\big\|_2^2\right)$$

$$\boldsymbol{\Psi}(i) = \left[\zeta(i)\binom{n}{n_1, \ldots, n_K}^{1/2} (x_1(i))^{n_1} \ldots (x_K(i))^{n_K}\right]'_{\substack{n=0,\ldots,\infty, \\ \sum_{k=1}^{K} n_k = n}}.$$

Hence, there exists a mapping $\boldsymbol{\Psi}$ corresponding to (10). Therefore, $\bar{h}_l(i, j, \boldsymbol{\theta}_k)$ is a valid kernel. ■

B. Proof of Proposition 2

Proof: Let us define $\boldsymbol{Y}(\mathbf{t}) = \boldsymbol{X}_r(\mathbf{t}) - \boldsymbol{\mu}(\mathbf{t})$. Using Bayes' theorem, we can derive the posterior distribution of the FPC score $\boldsymbol{\xi} = [\xi_{r1}, \ldots, \xi_{rK}]'$ as follows: $P(\boldsymbol{\xi}\big|\boldsymbol{Y}(\mathbf{t})) = \frac{P(\boldsymbol{\xi}, \boldsymbol{Y}(\mathbf{t}))}{P(\boldsymbol{Y}(\mathbf{t}))} = \frac{P(\boldsymbol{Y}(\mathbf{t})\big|\boldsymbol{\xi})P(\boldsymbol{\xi})}{P(\boldsymbol{Y}(\mathbf{t}))} \propto P(\boldsymbol{Y}(\mathbf{t})\big|\boldsymbol{\xi})P(\boldsymbol{\xi})$. The distribution $P(\boldsymbol{Y}(\mathbf{t})\big|\boldsymbol{\xi})$ is represented as

$$P\left(\boldsymbol{Y}(\mathbf{t})|\boldsymbol{\xi}\right) = P\left(Y(t_{r1}), \ldots, Y(t_{rp_r})|\boldsymbol{\xi}\right)$$

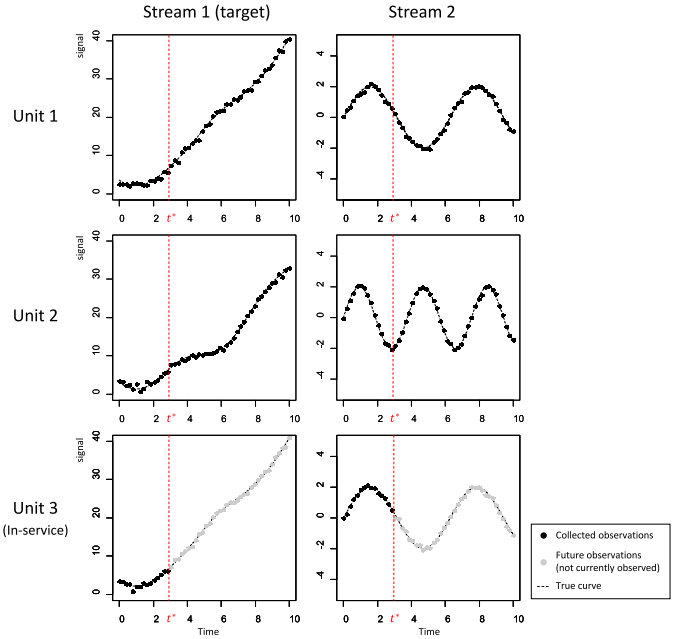

Fig. A.1. Simple example of multistream longitudinal data.

$$= \prod_{u=1}^{p_r} P\left(Y(t_{ru})|\boldsymbol{\xi}\right)$$
$$= \mathcal{N}\left(\boldsymbol{Y}(\mathbf{t})|\boldsymbol{\Phi}(\mathbf{t})\boldsymbol{\xi}, '\sigma_\epsilon^2 \mathbf{I}_{p_r}\right). \quad (13)$$

Given the prior distribution $P(\boldsymbol{\xi}) = \mathcal{N}(\boldsymbol{\xi}|\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ and using normal distribution theory, the posterior distribution $P(\boldsymbol{\xi}|\boldsymbol{Y}(\mathbf{t}))$ is a normal distribution derived as

$$P\left(\boldsymbol{\xi}|\boldsymbol{Y}(\mathbf{t})\right) = \mathcal{N}(\hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\Sigma}})$$

where

$$\hat{\boldsymbol{\xi}} = \hat{\boldsymbol{\Sigma}}\left(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0 + \frac{1}{\sigma_\epsilon^2}\boldsymbol{\Phi}(\mathbf{t})'\boldsymbol{Y}(\mathbf{t})\right)$$
$$= \hat{\boldsymbol{\Sigma}}\left(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0 + \frac{1}{\sigma_\epsilon^2}\boldsymbol{\Phi}(\mathbf{t})'\left(\boldsymbol{X}_r(\mathbf{t}) - \boldsymbol{\mu}(\mathbf{t})\right)\right)$$
$$\hat{\boldsymbol{\Sigma}} = \left(\frac{1}{\sigma_\epsilon^2}\boldsymbol{\Phi}(\mathbf{t})'\boldsymbol{\Phi}(\mathbf{t}) + \boldsymbol{\Sigma}_0^{-1}\right)^{-1}.$$

Note that $\hat{\boldsymbol{\xi}} = \boldsymbol{\xi}^*$, $\hat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}^*$, and $P(\boldsymbol{\xi}|\boldsymbol{Y}(\mathbf{t})) = P(\boldsymbol{\xi}|\boldsymbol{X}_r(\mathbf{t}))$ since $\boldsymbol{\mu}(\mathbf{t})$ is fixed. ■

C. Comparison to Functional Clustering Methods

In this section, we discuss the differences between our approach and functional clustering methods for longitudinal data [57]–[61]. The main differences are 1) whether knowledge transfer across streams occurs and 2) whether the extrapolation of incomplete signals is conducted.

To be detailed, we provide a simple hypothetical example of three (two historical and one in-service) units with two sensors, illustrated in Fig. A.1. Assuming that each signal has been collected for $t \in [0, 10]$, we are operating the in-service unit 3 and now the current time is t^*. The goal is to extrapolate the evolution of the signal of the target stream 1 for the in-service unit 3. In this setting, a single-stream functional clustering method [57]–[59] would not perform well as every signal in stream 1 for $t \in [0, t^*]$ exhibits a similar trend, resulting in clustering every target signal into the same class. On the other hand, our proposed model can effectively infer the difference by leveraging knowledge obtained from stream 2; in stream 2, the signal from unit 1 is more similar to that of unit 3 rather than 2, hence it is reasonable to expect that the future evolution of the target signal from the unit 1 will be similar to that of unit 3. That is, our model transfers information across streams.

Furthermore, the functional clustering methods do not provide extrapolation of an incomplete signal, which is one of the significant components of our model. Remark that there are several studies on functional clustering for multivariate longitudinal data (e.g., [60], [61]) accounting for similarities across streams. However, they mainly focus on clustering and it is not straightforward to incorporate the extrapolation procedure into their framework, while our approach seamlessly integrates the empirical Bayesian updating for computationally efficient extrapolation.

D. Computer Specifications

We utilize Intel(R) Core(TM) i7-8700 CPU @ 3.40 GHz (8 CPUs) and 32 GB RAM. We implemented all algorithms in **R** version 3.5.0.

REFERENCES

[1] M. Caldara, C. Colleoni, E. Guido, V. Re, G. Rosace, and A. Vitali, "A wearable sweat pH and body temperature sensor platform for health, fitness, and wellness applications," in *Sensors and Microsystems*. Berlin, Germany: Springer, 268, 2014, pp. 431–434.
[2] M. Magno *et al.*, "Infinitime: Multi-sensor wearable bracelet with human body harvesting," *Sustain. Comput.: Informat. Syst.*, vol. 11, pp. 38–49, 2016.
[3] W. Q. Meeker and Y. Hong, "Reliability meets big data: Opportunities and challenges," *Qual. Eng.*, vol. 26, no. 1, pp. 102–116, 2014.
[4] S. M. Salamati, C. S. Huang, B. Balagopal, and M.-Y. Chow, "Experimental battery monitoring system design for electric vehicle applications," in *Proc. IEEE Int. Conf. Ind. Electron. Sustain. Energy Syst.* 2018, pp. 38–43.
[5] Y.-L. Hsu *et al.*, "Design and implementation of a smart home system using multisensor data fusion technology," *Sensors*, vol. 17, no. 7, 2017, Art. no. 1631.
[6] J. O. Ramsay and B. W. Silverman, *Functional Data Analysis*, 2nd ed. Berlin, Germany: Springer, 2005.
[7] J. Peng and D. Paul, "A geometric approach to maximum likelihood estimation of the functional principal components from sparse longitudinal data," *J. Comput. Graph. Statist.*, vol. 18, no. 4, pp. 995–1015, 2009.
[8] C.-Z. Di, C. M. Crainiceanu, B. Caffo, and N. M. Punjabi, "Multilevel functional principal component analysis," *Ann. Appl. Statist. Ann. Appl. Statist.*, vol. 3, no. 1, pp. 458–488, 2009.
[9] J.-L. Wang, J. M. Chiou, and H.-G. Müller, "Functional data analysis," *Annu. Rev. Statist. Appl.*, vol. 3, pp. 257–295, 2016.
[10] L. Xiao, L. Cai, W. Checkley, and C. Crainiceanu, "Fast covariance estimation for sparse functional data," *Statist. Comput.*, vol. 28, no. 3, pp. 511–522, 2018.
[11] R. Kontar, J. Son, S. Zhou, C. Sankavaram, Y. Zhang, and X. Du, "Remaining useful life prediction based on the mixed effects model with mixture prior distribution," *IISE Trans.*, vol. 49, no. 7, pp. 682–697, Mar. 2017.
[12] J.-B. Schiratti, S. Allassonniere, O. Colliot, and S. Durrleman, "Learning spatiotemporal trajectories from manifold-valued longitudinal data," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 2404–2412.

[13] S. Allassonnière, J. Chevallier, and S. Oudard, "Learning spatiotemporal piecewise-geodesic trajectories from longitudinal manifold-valued data," in *Adv. Neural Inf. Process. Syst.*, 2017, pp. 1152–1160.

[14] J.-B. Schiratti, S. Allassonnière, O. Colliot, and S. Durrleman, "A bayesian mixed-effects model to learn trajectories of changes from repeated manifold-valued observations," *J. Mach. Learn. Res.*, vol. 18, no. 1, pp. 4840–4872, 2017.

[15] R. R. Zhou, N. Serban, and N. Gebraeel, "Degradation modeling applied to residual lifetime prediction using functional data analysis," *Ann. Appl. Statist.*, vol. 5, no. 2B, pp. 1586–1610, 2011.

[16] R. Zhou, N. Gebraeel, and N. Serban, "Degradation modeling and monitoring of truncated degradation signals," *IIE Trans.*, vol. 44, no. 9, pp. 793–803, Jun. 2012.

[17] X. Fang, R. Zhou, and N. Z. Gebraeel, "An adaptive functional regression-based prognostic model for applications with missing data," *Rel. Eng. Syst. Saf.*, vol. 133, pp. 266–274, 2015.

[18] M. A. Álvarez and L. D. Lawrence, "Computationally efficient convolved multiple output Gaussian processes," *J. Mach. Learn. Res.*, vol. 12, no. 41, pp. 1459–1500, May 2011.

[19] A. D. Saul, J. Hensman, A. Vehtari, and N. D. Lawrence, "Chained Gaussian processes," in *Proc. Artif. Intell. Statist.*, 2016, pp. 1431–1440.

[20] R. Kontar, S. Zhou, C. Sankavaram, X. Du, and Y. Zhang, "Nonparametric modeling and prognosis of condition monitoring signals using multivariate Gaussian convolution processes," *Technometrics*, vol. 60, no. 4, pp. 484–496, 2018.

[21] H. Liu, J. Cai, and Y.-S. Ong, "Remarks on multi-output gaussian process regression," *Knowl.-Based Syst.*, vol. 144, pp. 102–121, 2018.

[22] R. Kontar, G. Raskutti, and S. Zhou, "Minimizing negative transfer of knowledge in multivariate Gaussian processes: A scalable and regularized approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, to be published, doi: 10.1109/TPAMI.2020.2987482.

[23] W. Wang, R. Tuo, and C. Jeff Wu, "On prediction properties of kriging: Uniform error bounds and robustness," *J. Amer. Stat. Assoc.*, vol. 115, pp. 1–27, 2019.

[24] S. Fieuws and G. Verbeke, "Pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles," *Biometrics*, vol. 62, no. 2, pp. 424–431, 2006.

[25] Y.-M. Xia and N.-S. Tang, "Bayesian analysis for mixture of latent variable hidden Markov models with multivariate longitudinal data," *Comput. Statist. Data Anal.*, vol. 132, pp. 190–211, 2019.

[26] J. A. Dubin and H.-G. Müller, "Dynamical correlation for multivariate longitudinal data," *J. Amer. Stat. Assoc.*, vol. 100, no. 471, pp. 872–881, Jan. 2012.

[27] D. Xiang, P. Qiu, and X. Pu, "Nonparametric regression analysis of multivariate longitudinal data," *Stat. Sinica*, pp. 769–789, 2013.

[28] G. L. Hickey, P. Philipson, A. Jorgensen, and R. Kolamunnage-Dona, "Joint modelling of time-to-event and multivariate longitudinal outcomes: Recent developments and issues," *BMC Med. Res. Methodol.*, vol. 16, no. 1, 2016, Art. no. 117.

[29] S. Deng, K.-y. Liu, and X. Zhao, "Semiparametric regression analysis of multivariate longitudinal data with informative observation times," *Comput. Statist. Data Anal.*, vol. 107, pp. 120–130, 2017.

[30] J. Proudfoot, W. Faig, L. Natarajan, and R. Xu, "A joint marginal-conditional model for multivariate longitudinal data," *Statist. Med.*, vol. 37, no. 5, pp. 813–828, 2018.

[31] G. Verbeke, S. Fieuws, G. Molenberghs, and M. Davidian, "The analysis of multivariate longitudinal data: A review," *Stat. Methods Med. Res.*, vol. 23, no. 1, pp. 42–59, 2014.

[32] K. Liu, N. Z. Gebraeel, and J. Shi, "A data-level fusion model for developing composite health indices for degradation modeling and prognostic analysis," *IEEE Trans. Automat. Sci. Eng.*, vol. 10, no. 3, pp. 652–664, Jul. 2013.

[33] C. Song and K. Liu, "Statistical degradation modeling and prognostics of multiple sensor signals via data fusion: A composite health index approach," *IISE Trans.*, vol. 50, no. 10, pp. 853–867, May 2018.

[34] X. Fang, N. Z. Gebraeel, and K. Paynabar, "Scalable prognostic models for large-scale condition monitoring applications," *IISE Trans.*, vol. 49, no. 7, pp. 698–710, Mar. 2017.

[35] X. Fang, K. Paynabar, and N. Z. Gebraeel, "Multistream sensor fusion-based prognostics model for systems with single failure modes," *Rel. Eng. Syst. Saf.*, vol. 159, pp. 322–331, 2017.

[36] J.-M. Chiou, Y.-T. Chen, and Y.-F. Yang, "Multivariate functional principal component analysis: A normalization approach," *Stat Sinica*, pp. 1571–1596, 2014.

[37] C. Happ and S. Greven, "Multivariate functional principal component analysis for data observed on different (dimensional) domains," *J. Amer. Stat. Assoc.*, vol. 113, no. 522, pp. 649–659, 2018.

[38] C. R. Rao, "Some statistical methods for comparison of growth curves," *Biometrics*, vol. 14, pp. 1–17, 1958.

[39] P. E. Castro, W. H. Lawton, and E. A. Sylvestre, "Principal modes of variation for processes with continuous sample curves," *Technometrics*, vol. 28, no. 4, pp. 329–337, 1986.

[40] P. Besse and J. Ramsay, "Principal components analysis of sampled functions," *Psychometrika*, vol. 51, no. 2, pp. 285–311, 1986.

[41] B. Silverman, "Smoothed functional principal components analysis by choice of norm," *Ann. Statist.*, vol. 24, no. 1, pp. 1–24, 1996.

[42] G. Boente and R. Fraiman, "Kernel-based functional principal components," *Statist. Probab. Lett.*, vol. 48, no. 4, pp. 335–345, 2000.

[43] A. Kneip and K. Utikal, "Inference for density families using functional principal component analysis," *J. Amer. Stat. Assoc.*, vol. 96, no. 454, pp. 519–532, 2001.

[44] F. Yao, H.-G. Müller, and J.-L. Wang, "Functional data analysis for sparse longitudinal data," *J. Amer. Stat. Assoc.*, vol. 100, no. 470, pp. 577–590, 2005.

[45] J. Huang, H. Shen, and A. Buja, "Functional principal components analysis via penalized rank one approximation," *Electron. J. Statist.*, vol. 2, pp. 678–695, 2008.

[46] J. Goldsmith, S. Greven, and C. Crainiceanu, "Corrected confidence bands for functional data using principal components," *Biometrics*, vol. 69, no. 1, pp. 41–51, 2013.

[47] L. Xiao, V. Zipunnikov, D. Ruppert, and C. Crainiceanu, "Fast covariance estimation for high-dimensional functional data," *Statist. Comput.*, vol. 26, no. 1–2, pp. 409–421, 2016.

[48] J. Goldsmith *et al.*, *Refund: Regression With Funct. Data*, 2018, *r package version 0.1-17*. [Online]. Available: https://CRAN.R-project.org/package=refund

[49] J. Fan and I. Gijbels, *Local Polynomial Modelling and Its Applications*. London, U.K.: Chapman and Hall, 1996.

[50] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: The MIT Press, 2005.

[51] D. Rizopoulos, "Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data," *Biometrics*, vol. 67, no. 3, pp. 819–829, 2011.

[52] J. Son, Q. Zhou, S. Zhou, X. Mao, and M. Salman, "Evaluation and comparison of mixed effects model based prognosis for hard failure," *IEEE Trans. Rel.*, vol. 62, no. 2, pp. 379–394, Jun. 2013.

[53] A. Saxena and D. Simon, "Damage propagation modeling for aircraft engine run-to-failure simulation," in *Int. Conf. Prognostics Health Manage.*, Denvor, CO, 2008, pp. 1–9.

[54] A. Saxena and K. Goebel, "PHM08 challenge data set," Moffett Field, CA, 2008. [Online]. Available: https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/

[55] Y. Liu, D. K. Frederick, J. A. DeCastro, J. S. Litt, and W. W. Chan, "Users guide for the commercial modular aero-propulsion system simulation (C-MAPSS)," National Aeronautics and Space Administration (NASA), Cleveland, OH, Tech. Rep., 2012.

[56] T. Hofmann, B. Schölkopf, and A. J. Smola, "Kernel methods in machine learning," *Ann. Statist.*, pp. 1171–1220, 2008.

[57] J.-M. Chiou and P.-L. Li, "Functional clustering and identifying substructures of longitudinal data," *J. Roy. Stat. Soc., Ser. B (Stat. Methodol.*, vol. 69, no. 4, pp. 679–699, 2007.

[58] H.-g. Müller, "Functional modelling and classification of longitudinal data," *Scand. J. Stat.*, vol. 32, no. 2, pp. 223–240, 2005.

[59] J.-M. Chiou and P.-L. Li, "Functional clustering of longitudinal data," in *Funct. Oper. Statist*. Springer, 2008, pp. 103–107.

[60] J. Jacques and C. Preda, "Model-based clustering for multivariate functional data," *Comput. Statist. Data Anal.*, vol. 71, pp. 92–106, 2014.

[61] A. Maruotti and A. Punzo, "Model-based time-varying clustering of multivariate longitudinal data with covariates and outliers," *Comput. Statist. Data Anal.*, vol. 113, pp. 475–496, 2017.