Springer Optimization and Its Applications 211

Nathan Gaw
Panos M. Pardalos
Mostafa Reisi Gahrooei *Editors*

Multimodal and Tensor Data Analytics for Industrial Systems Improvement

Springer

Springer Optimization and Its Applications

Volume 211

Aims and Scope

Optimization has continued to expand in all directions at an astonishing rate. New algorithmic and theoretical techniques are continually developing and the diffusion into other disciplines is proceeding at a rapid pace, with a spot light on machine learning, artificial intelligence, and quantum computing. Our knowledge of all aspects of the field has grown even more profound. At the same time, one of the most striking trends in optimization is the constantly increasing emphasis on the interdisciplinary nature of the field. Optimization has been a basic tool in areas not limited to applied mathematics, engineering, medicine, economics, computer science, operations research, and other sciences.

The series **Springer Optimization and Its Applications (SOIA)** aims to publish state-of-the-art expository works (monographs, contributed volumes, textbooks, handbooks) that focus on theory, methods, and applications of optimization. Topics covered include, but are not limited to, nonlinear optimization, combinatorial optimization, continuous optimization, stochastic optimization, Bayesian optimization, optimal control, discrete optimization, multi-objective optimization, and more. New to the series portfolio include Works at the intersection of optimization and machine learning, artificial intelligence, and quantum computing.

Volumes from this series are indexed by Web of Science, zbMATH, Mathematical Reviews, and SCOPUS.

Nathan Gaw • Panos M. Pardalos •
Mostafa Reisi Gahrooei
Editors

Multimodal and Tensor Data Analytics for Industrial Systems Improvement

Springer

Editors
Nathan Gaw
Department of Operational Sciences
Air Force Institute of Technology
Wright-Patterson AFB, OH, USA

Panos M. Pardalos
Industrial and Systems Engineering
University of Florida
Gainesville, FL, USA

Mostafa Reisi Gahrooei
Industrial and Systems Engineering
University of Florida
Gainesville, FL, USA

ISSN 1931-6828 ISSN 1931-6836 (electronic)
Springer Optimization and Its Applications
ISBN 978-3-031-53091-3 ISBN 978-3-031-53092-0 (eBook)
https://doi.org/10.1007/978-3-031-53092-0

Mathematics Subject Classification: 62-02, 62B10, 62G99, 62H12, 62H20, 62H25, 62H30, 62H35, 62H99, 90B06, 62P10, 62-07

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Paper in this product is recyclable.

Preface

Multimodal and Tensor Data Analytics for Industrial Systems Improvement gives a topical overview of various methods found in multimodal data fusion for industrial engineering and operations research applications (e.g., manufacturing, medicine, agriculture, etc.). This book covers the latest methodologies available for using multimodal data fusion and analytics across a variety of applications. Advancements in sensing technologies and the shift toward the Internet of Things (IoT) have transformed and will continue to transform data analytics by producing new requirements and more complex forms of data. For example, (1) in manufacturing systems, multiple high-resolution sensors are available to monitor the condition of interconnected assets; (2) in healthcare, real-time data collected by wearable devices can be integrated with lab tests and images for more accurate patient prognostics; (3) in agriculture, hyperspectral imaging along with biosensors data can generate predictions about the yield and health of farming products. The abundance of multi-perspective data creates an unprecedented opportunity to design more efficient systems and make near-optimal operational decisions. On the other hand, the structural complexity and heterogeneity of the generated data poses significant challenges to extracting useful features and patterns for making use of the data and facilitating decision-making. Therefore, continual research to develop new statistical and analytical methodologies that overcome these data challenges and turn them into opportunities is needed. The purpose of this book is to demonstrate the recent developments and challenges in multimodal data analytics and to create a pathway toward new research development. The editors of this book combined various topics in the multimodal data analytics to provide comprehensive presentation of a variety of multimodal data analytics methods, such as functional analysis, tensor data analysis, spatiotemporal analysis, multimodal deep learning, the fusion of domain knowledge and data analytics, and multimodal federated learning. The chapters composing this book are written by eminent researchers and practitioners who present their research results and ideas based on their expertise. Thus, in this book, a wide spectrum of topics is presented that appear under the following titles, written by the respective authors and author groups.

Introduction to Multimodal and Tensor Data Analytics
Nathan Gaw, Mostafa Reisi Gahrooei, and Panos M. Pardalos

Functional Methods for Multimodal Data Analysis
Minhee Kim

Advanced Data Analytical Techniques for Profile Monitoring
Peiyao Liu and Chen Zhang

Statistical Process Monitoring Methods Based on Functional Data Analysis
Christian Capezza, Fabio Centofanti, Antonio Lepore, Alessandra Menafoglio, Biagio Palumbo, Simone Vantini

Tensor and Multimodal Data Analysis
Jing Zeng and Xin Zhang

Tensor Data Analytics in Advanced Manufacturing Processes
Bo Shen and Ning Wang

Spatiotemporal Data Analysis: A Review of Techniques, Applications, and Emerging Challenges
Imtiaz Ahmed and Ahmed Shoyeb Raihan

Offshore Wind Energy Prediction Using Machine Learning with Multi-resolution Inputs
Feng Ye, Travis Miles, Ahmed Aziz Ezzat

Sparse Decomposition Methods for Spatio-Temporal Anomaly Detection
Hao Yan, Ziyue Li, Xinyu Zhao, Jiuyun Hu

Multimodal Deep Learning
Amireza Shaban and Safoora Yousefi

Multimodal Deep Learning for Manufacturing Systems: Recent Progress and Future Trends
Yinan Wang and Xiaowei Yue

Synergy of Engineering and Statistics: Multimodal Data Fusion for Quality Improvement
Jianjun Shi, Michael Biehler, and Shancong Mou

Manufacturing Data Fusion: A Case Study with Steel Rolling Processes
Andi Wang

AI-Enhanced Fault Detection Using Multi-structured Data in Semiconductor Manufacturing
Linus Kohl, Theresa Madreiter, and Fazel Ansari

A Survey of Advances in Multimodal Federated Learning with Applications
Gregory Barry, Elif Konyar, Brandon Harvill, and Chancellor Johnstone

Bayesian Multimodal Data Analytics: An Introduction
Marco Luigi Giuseppe Grasso and Panagiotis Tsiamyrtzis

Bayesian Approach to Multimodal Data in Human Factors Engineering
Katherina A. Jurewicz and David M. Neyens

Bayesian Multimodal Models for Risk Analyses of Low-Probability High-Consequence Events
O. Arda Vanli

We would like to express our special thanks to all the authors of the chapters contributed to this book. Finally, we would like to acknowledge the superb assistance of the Springer staff during the preparation of this publication.

Wright-Patterson AFB, OH, USA Nathan Gaw
Gainesville, FL, USA Panos M. Pardalos
Gainesville, FL, USA Mostafa Reisi Gahrooei

Contents

Introduction to Multimodal and Tensor Data Analytics

Nathan Gaw, Mostafa Reisi Gahrooei, and Panos M. Pardalos

1 Overview

As data collection instruments have improved in quality and quantity for various applications, there has been an unprecedented increase in the availability of data from multiple sources (known as modalities). Modalities commonly express a large degree of heterogeneity in form, scale, resolution, and accuracy. Thus, determining how to combine the data for prediction and data characterization effectively is becoming increasingly important. Once these fragmented glimpses of each modality are interwoven, a more comprehensive understanding of the system of interest emerges. Several research studies have proposed ways to integrate and analyze multimodality data and discussed the limitations of current methodologies. This textbook gives a topical overview of various methods in multimodal data fusion for industrial engineering and operations research applications (e.g., manufacturing, medicine, and renewable energy). This book will cover the latest methodologies available for using multimodal data fusion and analytics across various applications.

Advancements in sensing technologies and the shift toward the Internet of Things (IoT) have transformed and will continue to transform data analytics by producing new requirements and more complex forms of data. For example, (1) in manufacturing systems, multiple high-resolution sensors are available to monitor the condition of interconnected assets [15], (2) in healthcare, real-time data collected

N. Gaw (✉)
Department of Operational Sciences, Air Force Institute of Technology, Wright-Patterson AFB, OH, USA
e-mail: Nathan.Gaw@afit.edu

M. R. Gahrooei · P. M. Pardalos
Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL, USA
e-mail: mreisigahrooei@ufl.edu; pardalos@ise.ufl.edu

N. Gaw et al. (eds.), *Multimodal and Tensor Data Analytics for Industrial Systems Improvement*, Springer Optimization and Its Applications 211,
https://doi.org/10.1007/978-3-031-53092-0_1

by wearable devices can be integrated with lab tests and images for more accurate patient prognostics [6, 26], (3) in agriculture, hyperspectral imaging along with biosensors data can generate predictions about the yield and health of farming products [3, 19], and (4) in renewable energy, spatiotemporal analytics of energy generation from wind turbines can be utilized to determine optimal placement of windmills to optimize system-level power production efficiency [4]. The abundance of data creates an unprecedented opportunity to design more efficient systems and make near-optimal operational decisions. On the other hand, the structural complexity and heterogeneity of the collected data pose a significant challenge to extracting useful features and patterns to use the data and facilitate decision-making. Therefore, continual research is needed to develop new statistical and analytical methodologies that overcome these data challenges and turn them into opportunities.

Multimodal dataset analysis has seen significant growth since its inception in the revolutionary work of Hotelling in 1936 [10]. Techniques such as multiset canonical correlation analysis, parallel factor analysis (PARAFAC), and tensor decomposition were introduced around the mid-twentieth century [11, 14, 23], but their application was mostly confined to specific areas like chemometrics [9]. With the recent surge of multimodal datasets, industries such as manufacturing, healthcare, agriculture, and renewable energy are increasingly exploring the potential advantages of these datasets. A systematic analysis and fusion of multimodal data facilitate decision-making processes and improvements to systems, thus enhancing their efficiency through effective predictive models, reliable abnormality detection methods, or accurate, interpretable features extracted from data.

Despite the apparent advantages of analyzing multimodal datasets, the nuances of leveraging the commonalities and disparities between modalities remain challenging. Data heterogeneity, differences in scale, resolution, and accuracy, as well as conflicting or redundant modes, pose considerable challenges to the progress of multimodal data analysis.

Within the chapters contained in this book, we have three primary modes of fusion: early fusion (low-level fusion), late fusion (high-level fusion), and intermediate fusion. Each methodology and application area tends to favor one fusion method over others. Early fusion, also referred to as low-level fusion, refers to the fusion of different modalities by leveraging the predictor information (i.e., independent variables) solely. This approach may be used as a preprocessing step before training a model or as an entirely unsupervised task to produce features representing underlying patterns across modalities. The goal of feature preprocessing in early fusion is to merge raw features from different modalities to craft new features that embody the complementary information of the raw features from distinct modalities. These newly created features are then used in a supervised model for a training task. When early fusion is used as an unsupervised task, it aims to combine features across modalities to identify underlying patterns that exist across these different modalities or to generate a visualization that accurately presents information from the different modalities, such as combining various types of medical imaging to create an image that displays complementary information

[8, 16, 17]. A popular instance of early fusion is Principal Component Regression (PCR), where the Principal Component Analysis (PCA) is utilized to extract input features that are then used to predict an output value.

Late fusion, or high-level fusion, is the integration at the decision-making level. After each modality has been processed and modeled individually, the predictions made from each modality can be merged in various ways. These ways depend on the relevance of each modality for the prediction task, the suitability of the modality combination (for instance, whether the fusion should be modeled as an element-wise summation [1], a weighted average [12], bilinear product [2], etc.), the noise level in each modality, and other considerations decided by the practitioner. Ensemble learning [20, 21, 25] and deep late fusion [12, 18, 22, 24] are some common examples of late fusion.

Intermediate fusion happens when features from different modalities are combined during the model training process, using both predictors (independent variables) and response (dependent variables). These methods incorporate fusion directly into the model training process and make fusion decisions that optimize the objective (for example, accuracy, detection rate, etc.). Partial least square (PLS) is an example of intermediate fusion, where the fused features are extracted in a supervised manner to explain the output [27]. Tensor regression is another instance of intermediate fusion, in which it extracts features from tensors containing multiple sources to estimate the output [5]. Deep learning architectures can also be engineered to perform intermediate fusion [13].

To help the reader comprehend early, late, and intermediate fusion concepts, we provide a figure adapted from [7]. Each subfigure demonstrates an example of each fusion type to combine three modalities, A, B, and C, with functions f and g representing model steps to fuse the data (whether early, late, or intermediate). Each method aims to efficiently use available data to predict outcomes (model outputs) and incorporate the information in the available multimodal data. Reference [7] provides additional real-world examples to further explain the multimodal fusion concepts illustrated in this figure (Fig. 1).

This book can be used as a reference for anyone in this general research area and an educational resource for those interested in building their knowledge set and capabilities. Primarily, this book would appeal to researchers in academia and graduate and post-graduate students with an interest in the areas of Industrial Engineering (IE), Operations Research (OR), Machine Learning, and Deep Learning. Multimodal data analytics for system improvement is becoming an increasingly popular field of research that does not yet have a comprehensive review with an industrial engineering/operations research audience in mind. Secondarily, this book appeals to tech industries interested in the applications of multimodal data analytics to advance and improve their data-enabled decision-making processes. Application examples (from manufacturing, healthcare, renewable energy, agriculture, etc.) in this book will guide industries to integrate data integration techniques into their workflow. Having a reference that summarizes the latest advances and how they can be applied to various applications will provide a rich resource to these individuals.

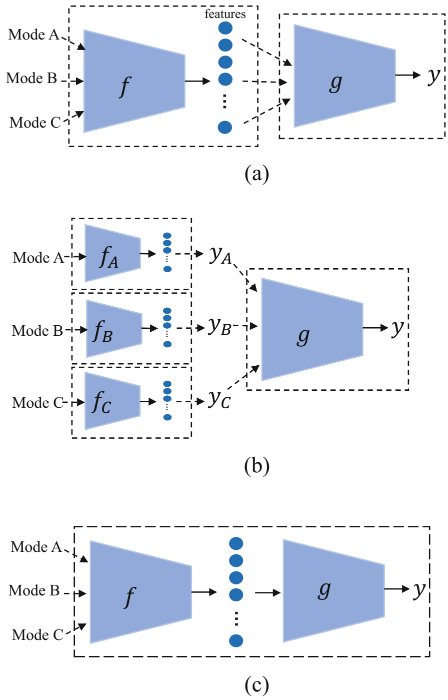

Fig. 1 Illustration of different levels of fusion. (**a**) Early Fusion. (**b**) Late Fusion. (**c**) Intermediate Fusion

In this comprehensive exploration, there are a variety of contributors that explore topic areas separated by parts. This book is structured as follows:

- The *current chapter* introduces the general concepts of multimodal data and tensor analytics and provides a roadmap for this work.
- *Part 1* delves into Functional Data Analysis, detailing its principles and applications, including an exploration of statistical process monitoring (SPM) of functional process variables with an application to multistage profile monitoring of sequential samples obtained from consecutive stages in manufacturing systems. This part contains three chapters.
- *Part 2* discusses Tensor Data Analytics, a key technique for handling multimodal datasets in advanced manufacturing (i.e., Industrial Internet of Things, IIoT) and medical domains using high-dimensional data (e.g., magnetic resonance imaging, electroencephalography, etc.). This part contains two chapters.
- *Part 3* focuses on Spatiotemporal Data Analytics, demonstrating how they are used in various fields, including offshore wind energy, manufacturing, computer vision, and recommender systems. This part contains three chapters.
- *Part 4* introduces the topic of Multimodal Deep Learning, presenting its importance in the current era of newer, more computationally involved methodologies. A variety of application areas are explored, including precision medicine, autonomous driving, and materials science. This part contains two chapters.

- *Part 5* explains how domain knowledge is integrated with multimodal data to generate useful insights and predictive models. Topics covered include fault detection of multi-structured data in semiconductors, quality improvement in manufacturing, and a case study with steel rolling processes. This part contains two chapters.
- *Part 6* explores Federated and Distributed Learning for Multimodal Data, presenting the benefits of these methods and their potential.
- Finally, *Part 7* explores the use of multimodal distributions and Bayesian frameworks in various applications, including human factors.

With a balanced blend of theory and real-world applications, this book aims to advance the understanding of multimodal data and the potential benefits this analysis can bring to diverse domains. The purpose of this book is to demonstrate the recent developments and challenges in multimodal data analytics within the domain of industrial and systems engineering.

References

1. Brentan, B. M., Luvizotto Jr, E., Herrera, M., Izquierdo, J., & Pérez-García, R. (2017). Hybrid regression model for near real-time urban water demand forecasting. *Journal of Computational and Applied Mathematics, 309*, 532–541.
2. Chen, D., & Irwin, D. (2017). Black-box solar performance modeling: Comparing physical, machine learning, and hybrid approaches. *ACM SIGMETRICS Performance Evaluation Review, 45*(2), 79–84.
3. Costa, L., McBreen, J., Ampatzidis, Y., Guo, J., Gahrooei, M. R., & Babar, M. A. (2022). Using UAV-based hyperspectral imaging and functional regression to assist in predicting grain yield and related traits in wheat under heat-related stress environments for the purpose of stable yielding genotypes. *Precision Agriculture, 23*(2), 622–642.
4. Ding, Y., Barber, S., & Hammer, F. (2022). Data-driven wind turbine performance assessment and quantification using Scada data and field measurements. *Frontiers in Energy Research, 10*, 1050342.
5. Gahrooei, M. R., Yan, H., Paynabar, K., & Shi, J. (2021). Multiple tensor-on-tensor regression: An approach for modeling processes with heterogeneous sources of data. *Technometrics, 63*(2), 147–159.
6. Gaw, N., Li, J., & Yoon, H. (2022). A novel semi-supervised learning model for smartphone-based health telemonitoring. *IEEE Transactions on Automation Science and Engineering, 21*(1), 428–441.
7. Gaw, N., Yousefi, S., & Gahrooei, M. R. (2022). Multimodal data fusion for systems improvement: A review. *IISE Transactions, 54*(11), 1098–1116.
8. He, C., Liu, Q., Li, H., & Wang, H. (2010). Multimodal medical image fusion based on IHS and PCA. *Procedia Engineering, 7*, 280–285.
9. Hopke, P. K. (2003). The evolution of chemometrics. *Analytica Chimica Acta, 500*(1–2), 365–377.
10. Hotelling, H. (1936). Simplified calculation of principal components. *Psychometrika, 1*(1), 27–35.
11. Ilarshman, R. (1970). Foundations of the PARAFAC procedure: Models and methods for an "explanatory" multi-mode factor analysis. *UCLA Working Papers in Phonetics, 16*, 1–84.

12. Kahou, S. E., Bouthillier, X., Lamblin, P., Gulcehre, C., Michalski, V., Konda, K., Jean, S., Froumenty, P., Dauphin, Y., Boulanger-Lewandowski, N., et al. (2016). EmoNets: Multimodal deep learning approaches for emotion recognition in video. *Journal on Multimodal User Interfaces, 10*(2), 99–111.
13. Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., & Fei-Fei, L. (2014). Large-scale video classification with convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1725–1732).
14. Kettenring, J. R. (1971). Canonical analysis of several sets of variables. *Biometrika, 58*(3), 433–451.
15. Miao, H., Wang, A., Li, B., & Shi, J. (2022). Structural tensor-on-tensor regression with interaction effects and its application to a hot rolling process. *Journal of Quality Technology, 54*(5), 547–560.
16. Moin, A., Bhateja, V., & Srivastava, A. (2016). Weighted-PCA based multimodal medical image fusion in contourlet domain. In *Proceedings of the International Congress on Information and Communication Technology* (pp. 597–605). Springer.
17. Rajalingam, B., & Priya, R. (2017). Multimodality medical image fusion based on hybrid fusion techniques. *International Journal of Engineering and Manufacturing Science, 7*(1), 22–29.
18. Ramachandram, D., & Taylor, G. W. (2017). Deep multimodal learning: A survey on recent advances and trends. *IEEE Signal Processing Magazine, 34*(6), 96–108.
19. Reisi-Gahrooei, M., Whitehurst, J. A., Ampatzidis, Y., & Pardalos, P. (2021). Dimensionality reduction techniques for high-dimensional data in precision agriculture. In *Modeling for Sustainable Management in Agriculture, Food and the Environment* (pp. 28–39). CRC Press.
20. Sagi, O., & Rokach, L. (2018). Ensemble learning: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8*(4), e1249.
21. Samareh, A., Jin, Y., Wang, Z., Chang, X., & Huang, S. (2018). Detect depression from communication: How computer vision, signal processing, and sentiment analysis join forces. *IISE Transactions on Healthcare Systems Engineering, 8*(3), 196–208.
22. Simonyan, K., & Zisserman, A. (2014). Two-stream convolutional networks for action recognition in videos. *Advances in Neural Information Processing Systems, 27*, 568–576.
23. Tucker, L. R. (1964). The extension of factor analysis to three-dimensional matrices. *Contributions to Mathematical Psychology, 110119*, 110–182.
24. Wu, D., Pigou, L., Kindermans, P.-J., Le, N. D.-H., Shao, L., Dambre, J., & Odobez, J.-M. (2016). Deep dynamic neural networks for multimodal gesture segmentation and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 38*(8), 1583–1597.
25. Yokoya, N., Ghamisi, P., & Xia, J. (2017). Multimodal, multitemporal, and multisource global data fusion for local climate zones classification based on ensemble learning. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (pp. 1197–1200). IEEE.
26. Yoon, H., & Gaw, N. (2021). A novel multi-task linear mixed model for smartphone-based telemonitoring. *Expert Systems with Applications, 164*, 113809.
27. Zhao, Q., Caiafa, C. F., Mandic, D. P., Chao, Z. C., Nagasaka, Y., Fujii, N., Zhang, L., & Cichocki, A. (2012). Higher order partial least squares (HOPLS): A generalized multilinear regression method. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 35*(7), 1660–1673.

Part I
Functional Methods for Multimodal Data

Functional Methods for Multimodal Data Analysis

Minhee Kim

1 What Is Functional Data and FDA?

In this section, we will first discuss the concept of functional data. Examples of functional data will be introduced to cover different data modes. Widely known methods to analyze such functional forms of data will then be briefly reviewed.

1.1 Functional Data

Let us start with simple examples of functional data. Figure 1 represents simulated sensor signal measurements of five units. Although signals are measured at discrete time points and are noisy, the signal trajectory of each unit clearly illustrates underlying increasing trends and can be seen as a function of time. In other words, the dataset consists of five *functional* samples, $Signal_i(t)$, $i = 1, \ldots, 5$. Another example is illustrated in Fig. 2. Compared to the signal example in Fig. 1, the measurements are much more sparse, with each unit having 8–10 measurements, and each measurement is not equally spaced. By considering each trajectory as one smooth function, FDA can easily address such sparse and unevenly distributed observations. We will revisit this sparse dataset example in the following subsections.

While univariate time series data as in Figs. 1 and 2 are the most common examples of functional data, more complicated forms of data such as multivariate or correlated (e.g., images) can also be considered as functional data. For instance,

M. Kim (✉)
University of Florida, Gainesville, FL, USA
e-mail: mkim3@ufl.edu

N. Gaw et al. (eds.), *Multimodal and Tensor Data Analytics for Industrial Systems Improvement*, Springer Optimization and Its Applications 211,
https://doi.org/10.1007/978-3-031-53092-0_2

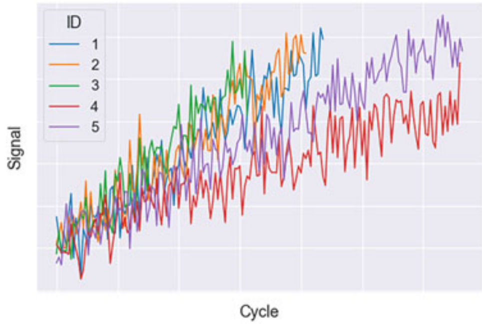

Fig. 1 Simulated noisy sensor signal measurements from five units

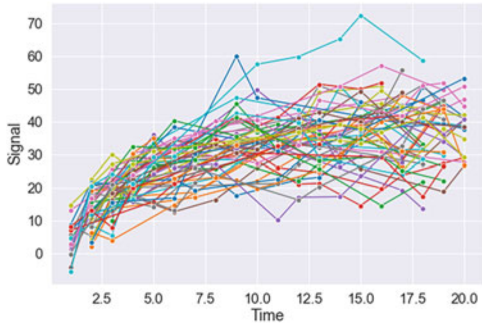

Fig. 2 Simulated sparse longitudinal measurements from 50 units

in the above example, signals from multiple sensors of one unit can be modeled together as one multivariate functional data. Also, variable t of a random function $X(t)$ is not necessarily time and can represent different domains such as location.

What motivates us to view data as a function instead of a multivariate random variable (vector)? The core assumption of functional data is *smoothness*. In other words, two neighboring data points are assumed to be similar to each other. For instance, the signal measurements of a unit at time t and $t - 1$ are assumed to be similar. The underlying smoothness allows us to borrow information from neighboring data points and avoid the curse of dimensionality. Note that functional data are intrinsically infinite-dimensional. Without the underlying smoothness assumption, there would be little difference between treating the data as functional and just as multivariate. In the following subsections, we will delve deeper into popular FDA methodologies to analyze functional data and its applications in various fields.

1.2 Examples of Functional Data Analyses

In this subsection, we will explore two of the most commonly used FDA methods: functional principal component analysis (FPCA) and functional linear model (FLM). Similar to the conventional PCA for multivariate data, FPCA explores major variations of sample curves (trajectories) by finding principal component functions that are orthogonal and maximize the curve variation. FLM extends the notion of a linear regression model to the functional context. In particular, for FLMs, either dependent (response) or independent variables (covariates) or both can be functional.

1.2.1 Functional Principal Component Analysis (FPCA)

Suppose there is a random function $X(t)$ in a compact (time) interval $\mathcal{T}$ with mean function $\mu(t)$ and covariance function $\Sigma(t, t')$. By minor abuse of notation, the covariance operator is defined as $\Sigma(g) = \int_{\mathcal{T}} \Sigma(s, t) g(s) ds$ for any function g satisfying $E(\int_{\mathcal{T}} g^2(s) ds) < \infty$. Under mild assumptions, Mercer's theorem implies that the covariance operator Σ has orthonormal eigenfunctions (principal component functions) $\phi_k(t),\ k = 1, 2, \ldots,$ with nonincreasing eigenvalues λ_k, i.e., $\lambda_1 \geq \lambda_2 \geq \ldots,$ satisfying $\Sigma(\phi_k) = \lambda_k \phi_k$ [9]. This results in the following Karhunen–Loève decomposition of the random function $X(t)$:

$$X(t) = \mu(t) + \sum_{k=1}^{\infty} A_k \phi_k(t) \ . \tag{1}$$

The kth functional principal component (FPC) score is $A_k = \int_{\mathcal{T}} (X(t) - \mu(t)) \phi_k(t) dt$, which are random variables uncorrelated across k, i.e., $Cov(A_k, A_{k'}) = 0$ if $k \neq k'$, $E(A_k) = 0$, and $Var(A_k) = \lambda_k$. Assuming that the top few FPCs explain most of the variability in the random curves, $X(t)$ can be approximated by a linear combination of the top K principal component functions with the corresponding FPC scores as coefficients:

$$X(t) \approx \mu(t) + \sum_{k=1}^{K} A_k \phi_k(t). \tag{2}$$

There are various ways for determining the proper value of K, including using the fraction of explained variance, AIC (Akaike information criterion), or BIC (Bayesian information criterion) [14]. Similar to the conventional PCA for multivariate data, estimating the top principal component functions is equivalent to estimating the covariance function $\Sigma(t, t')$ with low-rank structure.

As an example, Figs. 3 and 4 show the mean function and top two principal component functions, i.e., $\phi_1(t)$ and $\phi_2(t)$, extracted from the sparse longitudinal measurement dataset introduced in Sect. 1.1. These two principal component functions explain 89% of the total variance. This implies that we can represent an infinite-dimensional trajectory with only two values, i.e., $A_1,$ and A_2.

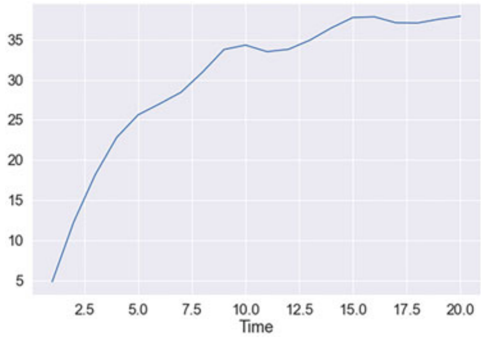

Fig. 3 FPCA results of the simulated sparse dataset: mean function

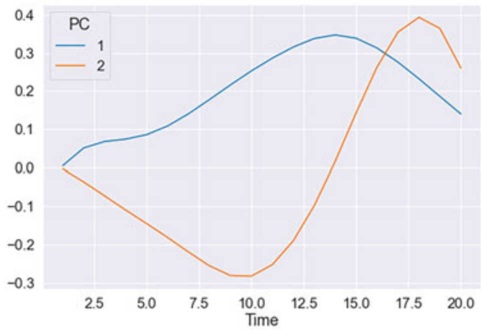

Fig. 4 FPCA results of the simulated sparse dataset: top two principal component functions

1.2.2 Functional Linear Model (FLM)

Another widely studied FDA method is functional linear models (FLMs) [12]. The FLMs have extended the traditional methodologies developed for linear regression and analysis of variance in the functional context. FLMs can be largely grouped into three categories: (i) scalar responses with functional covariates (scalar on function), (ii) functional responses with scalar covariates (function on scalar), and (iii) functional responses with functional covariates (function on function). Let us start with the conventional linear model with a scalar response Y and (multiple) scalar covariates $\boldsymbol{X} = [X_1, \ldots, X_M]$ as follows:

$$Y = \beta_0 + \sum_{m=1}^{M} \beta_m X_m + \varepsilon, \tag{3}$$

where the noise term is denoted as ε, and β_m is a regression coefficient. The FLMs with scalar responses and functional covariates replace the covariates $\boldsymbol{X}$ into a function:

$$Y = \beta_0 + \int_{\mathcal{T}} \beta(t) X(t) + \varepsilon. \tag{4}$$

This results in the coefficients $\beta(t)$ becoming functional as well. Recall that in conventional linear regression models, the number of data points must be larger than the number of covariates, M. The problem in (4) is that now the covariate $X(t)$ is infinite-dimensional, and thus we cannot estimate $\beta(t)$ by simply minimizing the errors. To overcome this, we can represent $\beta(t)$ as a linear combination of basis functions and impose regularization in the basis space, such as smoothness, truncation, or sparsity [13].

The FLMs with functional responses and (multiple) scalar covariates are as follows:

$$Y(t) = \beta_0(t) + \sum_{m=1}^{M} \beta_m(t) X_m + \varepsilon(t). \tag{5}$$

Various estimation methods have been proposed to obtain the estimates of the functional coefficients $\beta_m(t)$ [12]. One simple method is a two-step estimation, where we first obtain an initial estimate $\tilde{\beta}_m(t)$ by locally fitting the model in (5) using ordinary least squares and then smooth these initial estimates across t to obtain the final estimate $\hat{\beta}_m(\cdot)$ [3]. There are also other one-step approaches [5, 6].

Lastly, the FLMs with functional responses and functional covariates are as follows:

$$Y(t) = \beta_0(t) + \beta(t) X(t) + \varepsilon(t). \tag{6}$$

The above model is also referred to as functional concurrent models since both $Y(t)$ and $X(t)$ should be in the same domain $\mathcal{T}$ and $Y(t)$ depends only on concurrently observed $X(t)$. A more general extension of (6) is a model with unconstrained surface coefficient $\beta(s, t)$:

$$Y(t) = \beta_0(t) + \int \beta(s, t) X(s) ds + \varepsilon(t). \tag{7}$$

For the model in (7), we usually impose regularization on the coefficient surface $\beta(s, t)$ in each dimension separately to obtain more interpretable and efficient results. Several methods have been proposed on the function-on-function FLMs with different choices of basis functions or regularization approaches [1, 7, 13].

In the following section, we will introduce several multimodal data analysis applications using such FDA methodologies in detail.

2 Why Is FDA Useful for Multimodal Data Analysis?

This section highlights and summarizes why FDA is particularly useful in multi-modal data fusion. First of all, FDA can provide a *consistent* summary of different

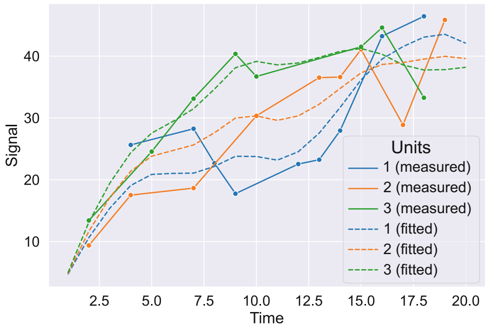

Fig. 5 Simulated sparse measurements collected from three randomly selected units (solid) and their fitted functions using FPCA (dashed)

data modes for multimodal data fusion. In particular, FDA enables us to efficiently handle multiple time series with different acquisition intervals, lengths, or sparsity. For instance, consider sparse longitudinal measurements from three randomly selected units illustrated in Fig. 5 (see Sect. 1.1 for more information on the dataset). Our goal can be to study the variations between units or to combine the information from all units. Yet, this can be challenging as all units have different measurement time points and numbers of measurements. By using FDA, each time series can be viewed as one function, and by applying FPCA discussed in Sect. 1.2, we can estimate the underlying evolution of each unit as dashed lines in Fig. 5. In this way, each unit is now summarized into a set of FPC scores A_k, which will make it much easier to conduct further data fusion.

It is also possible to apply FDA in a broader context of multimodal data fusion. For instance, we may fuse time series data and image data or analyze measurements collected from three experiments that used different instruments. The consistent summary of multimodal data can lead to dimensionality reduction, visualization, integration, and other benefits. The next subsection will go into further detail on how to leverage the FDA results (a summary of different data modes) to facilitate such multimodal data analysis.

In addition to providing a summary of multiple data modes, different methodologies in the FDA can also directly model both static (scalar) data modes and dynamic or spatial (functional) data modes. The FLMs introduced in Sect. 1.2 are one representative example. Compared to the conventional linear models whose responses and predictors are (multiple) scalar values, the FLMs allow us to incorporate functional data as either responses or predictors or both into the model. Here are some specific application examples of FLMs in various fields: Scalar-on-function models in (4) can be used in environmental engineering to model the precipitation (scalar response) based on the temperature over time (functional covariate). Also, in quality control, the quality of a final product (scalar response) can be modeled based on the production sensor signal over time (functional covariate). Using the function-on-scalar model in (5), we can establish a model with the voltage over time as a functional response and the current as a scalar covariate. In the field of economics, a function-on-scalar linear model can be used to predict the GDP over time based on the inflation rate. The function-on-function model in (6) can be applied to spatial multimodal data and model the temperature of different locations and the corresponding air pressure. As functional data become more widely available in various fields, FDA is becoming a more powerful tool to conduct multimodal data analysis.

3 How Can FDA Be Used in Multimodal Data Analysis?

In the earlier sections, we briefly explored how FDA can be employed to analyze multimodal data. This section will review recent work on FDA to help a deeper understanding of how the FDA can be used for multimodal data, using a covariate-dependent sparse FDA as an example.

3.1 Problem Setting

Suppose we have multimodal data consisting of multiple units. For each unit, we have sparse and noisy measurements of its process (functional data) and (multiple) static covariates. One example is a manufacturing unit which consists of several sparse measurements of its maintenance status (dynamic functional data) and its design specifications (static covariates). Our goal is to characterize the effects of these covariates on functional data and to predict a functional trajectory of a new unit of interest based on its covariates.

While existing FDA methodologies such as FLMs in Sect. 1.2 can be used, several significant challenges need to be overcome. First, the informative covariates may have complicated effects on functional data which cannot be modeled linearly or additively. Second, many existing models may not provide accurate and reliable results given only sparse and irregularly spaced measurements. To address these issues, we establish the following covariate-dependent sparse FDA model [8]. Figure 6 illustrates the overall framework of the covariate-dependent sparse functional data analysis. We will first model the variation within each unit through the FPCA and then model the variation between different units. Through these two types of variations, we will be able to make predictions of the unit of interest and identify important covariates as well.

3.2 Variations Within Each Unit: FPCA

Suppose there are I units, and the ith unit has n_i noisy and sparse measurements. Let Y_{ij} be the jth measurement of the ith unit at a random point t_{ij}. Using the FPCA in (2), we can decompose the functional measurements as follows:

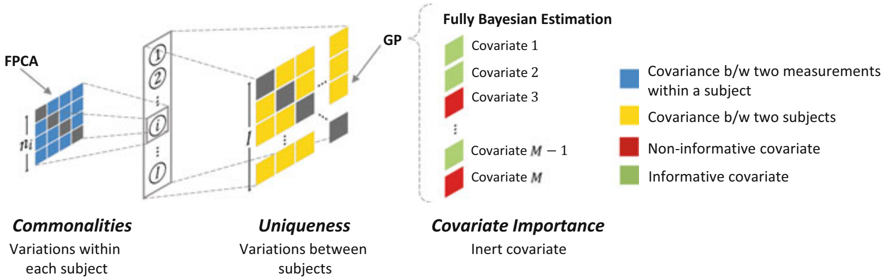

Fig. 6 Illustration of the covariate-dependent sparse FDA framework [8]

$$Y_{ij} \approx \mu(t_{ij}) + \sum_{k=1}^{K} A_{ik}\phi_k(t_{ij}) + \varepsilon_{ij} = \mu(t_{ij}) + \boldsymbol{A}_{(i)}^T \boldsymbol{\phi}(t_{ij}) + \varepsilon_{ij}, \tag{8}$$

where $E(\varepsilon_{ij}) = 0$, $Var(\varepsilon_{ij}) = \sigma_\varepsilon^2$, $\boldsymbol{\phi}(t) = [\phi_1(t), \ldots, \phi_K(t)]^T$ is a set of vectors derived from the eigenfunctions, and $\boldsymbol{A}_{(i)} = [A_{i1}, \ldots, A_{iK}]^T$ is a vector of the FPC scores, unique to the ith unit.

Similar to Yao et al. [15], we can pool all measurements from all units to overcome the sparsity issue and apply smoothing methods such as local linear smoothing [2] to estimate the mean function $\hat{\mu}(t)$ and covariance surfaces $\hat{\Sigma}(t, t')$. Once the mean and covariance of the functional data have been estimated, the estimation of the eigencomponents (λ_k and $\phi_k(t)$) can be done straightforwardly by solving the eigen-equations, i.e.,

$$\int_{\mathcal{T}} \widehat{\Sigma}(t, t')\widehat{\phi}_k(t)dt = \widehat{\lambda}_k \widehat{\phi}_k(t), \tag{9}$$

where $\widehat{\phi}_k$ is unit to $\int_{\mathcal{T}} \widehat{\phi}_k(t)^2 dt = 1$ and $\int_{\mathcal{T}} \widehat{\phi}_k(t) \times \widehat{\phi}_{k'}(t)dt = 0$ for $k < k'$.

After obtaining the estimation λ_k and $\phi_k(t)$, each unit's trajectory (function) can be summarized into a set of FPC scores A_k, similar to Fig. 5. In general, the FPC scores of each unit are estimated by numerical integration based on its definition, $\hat{A}_k = \int_{\mathcal{T}} (X(t) - \hat{\mu}(t))\hat{\phi}_k(t)dt$. Yet, this approach does not work well with sparse measurements and does not take into account the covariate information. Next, we will introduce a novel approach to overcome these limitations.

3.3 Variations Between Units: Kernel Method

In general, it is assumed that the FPC scores A_{ik} and A_{rk} for the ith and rth units are independent for $i \neq r$. Nevertheless, as discussed in the previous subsections, there may be an underlying relationship between the FPC scores $\{A_{ik}\}_{i=1}^{I}$ and the corresponding units' covariates $\boldsymbol{X}_i$ and $\boldsymbol{X}_r$. In such models, it is reasonable to assume that $Cov(A_{ik}, A_{rk})$ depends on the covariate difference between the ith and rth units, $\|\boldsymbol{X}_i - \boldsymbol{X}_r\|$, where $\|\cdot\|$ is a distance measure such as the Euclidean norm. Such a priori knowledge is not taken into account in the conventional FPC score estimation described in Sect. 3.2.

Different models can be applied to encode such covariate information into the estimation of FPC scores. In the following, we will use the Gaussian processes (GPs) due to its great flexibility to establish the nonparametric relation, the interpolation capability at any covariate $\boldsymbol{X}$, and the ability to quantify uncertainties. In particular, we impose K independent zero-mean GPs on A_k, $k = 1, \ldots, K$:

$$A_k(\boldsymbol{Z}) \sim \mathcal{GP}(0, K_k(\boldsymbol{X}, \boldsymbol{X}')), \tag{10}$$

where $K_k(\boldsymbol{X}, \boldsymbol{X}')$ is the kernel function of the GP on A_k. For the ith unit with covariates $\boldsymbol{X}_i$, the prior distribution of A_{ik} follows a Gaussian distribution with variance $K_k(\boldsymbol{X}_i, \boldsymbol{X}_i)$. For any two units $i, r \in \{1, \cdots, I\}$, the covariance between A_{ik} and A_{rk} is now quantified by their covariate similarities, i.e., $Cov(A_{ik}, A_{rk}) = K_k(\boldsymbol{X}_i, \boldsymbol{X}_r)$.

As suggested in Kim et al. [8], the kernel $K_k(\boldsymbol{X}, \boldsymbol{X}')$ is specified as the squared exponential covariance function with a separate scale parameter ρ_{km} for each covariate Z_{im}:

$$K_k(\boldsymbol{X}_i, \boldsymbol{X}_r) = \lambda_k \exp\left(-\frac{1}{2} \sum_{m=1}^{M} \frac{1}{\rho_{km}} (X_{im} - X_{rm})^2 \right). \tag{11}$$

This kernel design is also well known as the automatic relevance determination (ARD) kernel [10], where the characteristic length scale for the mth covariate is given by $\rho_{km}^{1/2}$. Ideally, if the mth covariate is irrelevant, the estimation of ρ_{km} should be large enough in order for the model to ignore this covariate, i.e., the difference between Z_{im} and Z_{rm} has negligible effects on the covariance between A_{ik} and A_{rk}. On the other hand, if ρ_{km} is small, A_k will vary rapidly along the corresponding covariate, implying the high "relevance" of the mth covariate. For more detailed information about the informative covariate identification in this framework, see Kim et al. [8]. Another key part of the kernel design in (11) is that the kth largest eigenvalue λ_k derived from FPCA acts as a scale factor. In this way, for the ith unit with covariates $\boldsymbol{X}_i$, the variance of the prior distribution of the kth FPC score is reduced to λ_k, i.e., $Var(A_{ik}) = K_k(\boldsymbol{Z}_i, \boldsymbol{X}_i) = \lambda_k$, and thus the proposed design bridges the gap between FPCA and GP modeling. This further resolves the unidentifiability issue in the ARD kernel [16].

To summarize, we first pool the covariance within a unit at the level of Y_{ij} to characterize the commonalities shared by all units, i.e., estimating the parameters of FPCA: $\mu(t)$, $\phi_k(t)$, λ_k, and σ_ε^2. We then consider the covariance between units at the level of $\boldsymbol{X}_i$ to characterize the uniqueness of each unit through the FPC scores $\boldsymbol{A}_{(i)}$ and model the covariate importance through $\boldsymbol{\rho}_k$. Note that while these two procedures are conducted separately, they are under an *integrated* structure where we use the variation of measurements Y_{ij} to derive the variation from covariates $\boldsymbol{Z}$ (between units) and the variation left conditioned on covariates $\boldsymbol{X}$ (within a unit). This is possible because each unit's trajectory is summarized through a set of FPC scores, and the covariate information only contributes to the between-units covariance of these scores. Also, note that the proposed model handles functional responses and (multiple) scalar covariates similar to the FLMs in (5). Yet, the main difference lies in that instead of assuming linear and additive effects of covariates, we model the nonlinear and multiplicative effects of covariates.

3.4 Estimation and Prediction

In this subsection, we will discuss how to estimate the set of length scale parameters $\boldsymbol{\rho}$ and make predictions of the functional trajectories of the units of interest, i.e., estimate new units' FPC scores based on their covariates and measurements. The most widely used approach to estimate the parameters of the mean and covariance functions in GPs is through Type II maximum likelihood (maximizing the marginal likelihood). Although this method has several advantages including traceability for Gaussian noise models and analytical solutions for predictions, the resulting point estimation tends to be unstable and overfit, especially in the cases of a small number of units, highly sparse measurements per unit, or high-dimensional covariates [11]. These issues worsen in our problem since there are no direct realizations of the FPC scores $\boldsymbol{A}_{(i)}$. Instead, only indirect observations of the linear combination of FPC scores of each unit are available through the sparse and noisy measurements.

To address these issues and obtain more robust results, we may establish a fully Bayesian hierarchical scheme.

Prior over hyperparameters	$\rho_{km} \sim p(\rho_{km}),\ \ k = 1, \ldots, K$ and $m = 1, \ldots, M$
Prior over parameters	$\boldsymbol{A}_k \vert \boldsymbol{Z}, \boldsymbol{\rho}_k \sim N(0, \boldsymbol{K}_k),\ \ k = 1, \ldots, K$
Likelihood	$\boldsymbol{Y} \vert \boldsymbol{A} \sim N(\boldsymbol{\mu} + \boldsymbol{\Phi}\boldsymbol{A}, \sigma_\varepsilon^2 \mathbb{I})$

Under the Bayesian hierarchical scheme, numerical methods like Hamiltonian Monte Carlo (HMC) or No-U-Turn-Sampler (NUTS) can be applied to estimate the posterior distribution of the unknown parameters $\boldsymbol{\rho}$ [4]. Based on these estimation results, it is straightforward to make the predictions for a new unit with covariate information $\boldsymbol{Z}^\star$. In particular, we integrate over the joint posterior:

$$p(\boldsymbol{A}^\star | \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{Z}^\star) = \int \int p(\boldsymbol{A}^\star | \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{Z}^\star, \boldsymbol{A}, \boldsymbol{\rho}) p(\boldsymbol{A} | \boldsymbol{\rho}, \boldsymbol{Y}, \boldsymbol{Z}) p(\boldsymbol{\rho} | \boldsymbol{Y}, \boldsymbol{Z}) d\boldsymbol{A} d\boldsymbol{\rho}.$$

Using the fact that the joint prior distribution of $\boldsymbol{A}^\star$ and $\boldsymbol{A}$ is Gaussian distribution and $p(\boldsymbol{A}^\star | \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{Z}^\star, \boldsymbol{A}, \boldsymbol{\rho})$ and $p(\boldsymbol{A} | \boldsymbol{\rho}, \boldsymbol{Y}, \boldsymbol{Z})$ are Gaussian distributions as well, the above can be further simplified by integrating out $\boldsymbol{A}$:

$$p(\boldsymbol{A}^{\star}|\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{Z}^{\star}) = \int p(\boldsymbol{A}^{\star}|\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{Z}^{\star}, \boldsymbol{\rho}) p(\boldsymbol{\rho}|\boldsymbol{Y}, \boldsymbol{Z}) d\boldsymbol{\rho}, \tag{12}$$

where $\boldsymbol{A}^{\star}|\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{Z}^{\star}, \boldsymbol{\rho} \sim N(\boldsymbol{\mu}^{\star}, \boldsymbol{\Sigma}^{\star})$ with parameters,

$$\begin{aligned} \boldsymbol{\mu}^{\star} &= \boldsymbol{K}^{\star}\boldsymbol{K}^{-1}\left(\frac{1}{\sigma_{\varepsilon}^{2}}\boldsymbol{\Phi}^{T}\boldsymbol{\Phi} + \boldsymbol{K}^{-1}\right)^{-1}\frac{1}{\sigma_{\varepsilon}^{2}}\boldsymbol{\Phi}^{T}(\boldsymbol{Y} - \boldsymbol{\mu}), \text{ and} \\ \boldsymbol{\Sigma}^{\star} &= \boldsymbol{K}^{\star}\boldsymbol{K}^{-1}\left(\frac{1}{\sigma_{\varepsilon}^{2}}\boldsymbol{\Phi}^{T}\boldsymbol{\Phi} + \boldsymbol{K}^{-1}\right)^{-1}\boldsymbol{K}^{-1}\boldsymbol{K}^{\star T} + \boldsymbol{K}^{\star\star} - \boldsymbol{K}^{\star}\boldsymbol{K}^{-1}\boldsymbol{K}^{\star T}. \end{aligned} \tag{13}$$

$\boldsymbol{K}^{\star\star} = \boldsymbol{K}(\boldsymbol{Z}^{\star}, \boldsymbol{Z}^{\star})$ denotes the covariance matrix of $\boldsymbol{Z}^{\star}$ and $\boldsymbol{K}^{\star}$ denotes that between $\boldsymbol{Z}^{\star}$ and $\boldsymbol{Z}$, where $\boldsymbol{K}(\boldsymbol{Z}, \boldsymbol{Z}')$ is a diagonal matrix with diagonal entries $K_k(\boldsymbol{Z}, \boldsymbol{Z}')$, $k = 1, \ldots, K$. The proof of (13) can be found in Kim et al. [8]. As a result, the predictive distribution in (12) can be approximated with Monte Carlo integration:

$$\begin{aligned} p(\boldsymbol{A}^{\star}|\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{Z}^{\star}) &= \int p(\boldsymbol{A}^{\star}|\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{Z}^{\star}, \boldsymbol{\rho}) p(\boldsymbol{\rho}|\boldsymbol{Y}, \boldsymbol{Z}) d\boldsymbol{\rho} \\ &\approx \frac{1}{H}\sum_{h=1}^{H} p(\boldsymbol{A}^{\star}|\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{Z}^{\star}, \boldsymbol{\rho}_{(h)}), \quad \boldsymbol{\rho}_{(h)} \sim p(\boldsymbol{\rho}|\boldsymbol{Y}, \boldsymbol{Z}), \end{aligned} \tag{14}$$

where $\boldsymbol{\rho}_{(h)}$ is a random draw from the hyperparameter posterior obtained through the numerical method. We can then obtain the approximate posterior distribution of the response $Y^{\star}$ using $Y^{\star} = \widehat{\mu}(t^{\star}) + \sum_{k=1}^{K} A_k^{\star}\widehat{\phi}_k(t^{\star})$.

Note that this is just one of many examples that illustrate in detail how functional methods can be utilized for multimodal data analysis. For more detailed information about the practical implementation and case study results of this particular methodology, please refer to [8].

4 Concluding Remarks

In this chapter, we discussed the importance and usefulness of the FDA focusing on multimodal data analysis. The FDA offers effective statistical approaches to handle various functional data, such as time series data, spatial data, and imaging data. Particularly, for multimodal data analysis, the FDA provides several unique advantages, such as providing a cohesive summary of different data modes, facilitating further multimodal data fusion, and directly incorporating both static and dynamic data modes into a unified model. As functional data has become more and more widely available in different domains, the FDA is emerging as one of the most potent tools for multimodal data analysis. However, there remain many exciting open questions in the FDA that need to be addressed to further advance its capabilities in multimodal data analysis, such as scalability and computational efficiency.

References

1. Besse, P. C., & Cardot, H. (1996). Approximation spline de la prévision d'un processus fonctionnel autorégressif d'ordre 1. *Canadian Journal of Statistics, 24*(4), 467–487.
2. Fan, J., & Gijbels, I. (1996). *Local polynomial modelling and its applications*. Routledge.
3. Fan, J., & Zhang, W. (1999). Statistical estimation in varying coefficient models. *The Annals of Statistics, 27*(5), 1491–1518.
4. Hoffman, M. D., & Gelman, A. (2014). The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research, 15*(1), 1593–1623.
5. Hoover, D. R., Rice, J. A., Wu, C. O., & Yang, L.-P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika, 85*(4), 809–822.
6. Huang, J. Z., Wu, C. O., & Zhou, L. (2002). Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika, 89*(1), 111–128.
7. Ivanescu, A. E., Staicu, A.-M., Scheipl, F., & Greven, S. (2015). Penalized function-on-function regression. *Computational Statistics, 30*(2), 539–568.
8. Kim, M., Allen, T., & Liu, K. (2023). Covariate dependent sparse functional data analysis. *INFORMS Journal on Data Science, 2*(1), 81–98.
9. Mercer, J. (1909). Functions of positive and negative type, and their connection with the theory of integral equations. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character, 83*(559), 69–70.
10. Neal, R. M. (1995). *Bayesian learning for neural networks*. Ph.D. thesis, University of Toronto.
11. Neal, R. M. (1997). Monte Carlo implementation of Gaussian process models for Bayesian regression and classification. arXiv preprint physics/9701026.
12. Ramsay, J. O., & Dalzell, C. J. (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society: Series B (Methodological), 53*(3), 539–561.
13. Ramsay, J.O., & Silverman, B. W. (2008). *Functional data analysis*. Springer.
14. Wang, J.-L., Chiou, J.-M., Müller, H.-G., & Mueller, H.-G. (2016). Review of Functional Data Analysis. *Annual Review of Statistics and Its Application, 3*(1), 257–295.
15. Yao, F., Müller, H. G., & Wang, J. L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association, 100*(470), 577–590.
16. Zhang, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association, 99*(465), 250–261.

Advanced Data Analytical Techniques for Profile Monitoring

Peiyao Liu and Chen Zhang

1 Introduction

As sensing technology advances, in situ sensing is being increasingly deployed in current manufacturing systems, where high-resolution in-process data are continuously recorded with high frequency. These data, known as *profile* or *functional data*, are very common in the current advanced manufacturing systems. Examples include tonnage signals [17] from a forging process, turning signals from a pipe-casing tightening process [13], surface profiles from a lathe-turning process [4], etc. Compared to traditional quality characteristics, profile data provide more detailed features of the system, which facilitate better in-process monitoring. For example, Fig. 1 illustrates the three-channel profiles over one in-control (IC) and one out-of-control (OC) sample from a pipe-casing tightening process. Developing effective methods to accurately model IC profiles and efficiently detect changes in OC profiles is essential for improved statistical process control (SPC) and quality management.

We highlight three main challenges in developing advanced data analytical methods for profile modeling and monitoring. First, as the sensing technology advances nowadays, a group of sensors are often installed in a manufacturing system to collect profile data of different process variables simultaneously, from which the collected data are called multi-channel profiles. These multi-channel profile data may have complex correlation structures. When the number of profiles is ultra-high, the number of parameters to model the between-profile correlations will be very high and leads to the curse of dimensionality problem. Furthermore, when a fault occurs, it may only affect small segments of very few profiles and increase

P. Liu · C. Zhang (✉)
Department of Industrial Engineering, Tsinghua University, Beijing, P.R. China
e-mail: liupy20@mails.tsinghua.edu.cn; zhangchen01@tsinghua.edu.cn

N. Gaw et al. (eds.), *Multimodal and Tensor Data Analytics for Industrial Systems Improvement*, Springer Optimization and Its Applications 211,
https://doi.org/10.1007/978-3-031-53092-0_3