UNIVERSITY OF CALIFORNIA SAN DIEGO

Applications of Longitudinal and Functional Data Analysis

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Biostatistics

by

Wenyi Lin

Committee in charge:

Professor Armin Schwartzman, Chair
Professor Michael Donohue
Professor Loki Natarajan
Professor Cheryl Lee Rock
Professor Wesley Kurt Thompson
Professor Jingjing Zou

2022

The Dissertation of Wenyi Lin is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2022

DEDICATION

To my parents: For your selfless love and support.

EPIGRAPH

It is a capital mistake to theorize before one has data.
Insensibly one begins to twist facts to suit theories,
instead of theories to suit facts.

Arthur Conan Doyle

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

It has been an amazing 5-year journey to obtain doctoral training in the division of Biostatistics at University of California San Diego (UCSD). Starting from here, I became aware of the novelty of biostatistics research. I have met so many great people, whom I sincerely wish to acknowledge.

First and foremost, I would like to express my deepest appreciation to my mentor and committee chair, Professor Armin schwartzman. I am grateful for your dedication in helping me become a good biostatistician and a good researcher. You have been my academic advisor since I started my Ph.D. study in 2017 and always inspired me with your enthusiasm for research and work. I would like to specifically thank you for shaping my analytical skills and scientific intuition. Those discussions we had about research works, career plans or daily issues are valuable treasure for my whole life. Thank you for your tremendous encouragement and support throughout my Ph.D. study.

I would also like to extend my deep gratitude to my committee members, Professor Loki Natarajan and Professor Wesley Kurt Thompson, for providing me with valuable guidance to complete this dissertation. Thank you Professor Loki Natarajan for bringing me into the exciting research field of wearable devices and guiding me through the study procedure. Thank you Professor Wesley Kurt Thompson for providing me the opportunity to work on the field Alzheimer's disease and introducing me to Professor Michael Donohue. It has been a great experience collaborating with both of you on these exciting projects, through which, I gained valuable understanding of the related research fields. Thank you for providing scholarly and hands-on advice, which inspired me to become a better researcher.

Moreover, I would like to thank my committee members, Professor Michael Donohue, Professor Cheryl Lee Rock and Professor Jingjing Zou, for your valuable feedback and comments on my dissertation. Your experiences and knowledge on both scientific and statistical background helped me shape research ideas and refine analytical results. I am truly grateful for all your suggestions on my research.

Furthermore, I would like to thank Professors Karen Messer, Florin Vaida, Xin Tu Sonia Jain, Charles Berry, Lin Liu, Xinlian Zhang, Lily Xu and Steven Edland for teaching me and helping me in this program. And thank you all my friends and classmates at UCSD, especially Ruifeng Chen, Jinyuan Liu, Serene Jiang, Kristen Hansen, Yuqi Qiu, Yiran Zhang and Tuo Lin, and I am grateful to meet you all in San Diego. Thank you Junrui Di, Ding Ding, Jiawei Bai, Haoyu Zhang, Zhiyu Xia from Johns Hopkins for always supporting me. Thank you Bingru Weng. Even though we are thousands of miles away from each other, you are always my best friend. Special thank to Bumeng Zhuo, for your patience and consistent help.

Lastly, I would like to dedicate this dissertation to my mother, Liping Jiang, and my father, Feng Lin. I always think I am the luckiest daughter in the world to have you as my parents. Thank you for supporting me mentally and financially for my education and my life. Wish you safe and sound and hope to see you soon.

Chapter 3, in full, has been prepared for submission for publication as ”Wenyi Lin, Michael C. Donohue, Philip Insel, Armin Schwartzman, and Wesley K. Thompson. Bayesian Multivariate Growth Mixture Modeling of Longitudinal Data: An Application to Alzheimer’s Disease Study.”. The dissertation author was the primary author of this paper.

Chapter 4, in full, has been submitted for publication as ”Wenyi Lin, Jingjing Zou, Chongzhi Di, Dorothy D. Sears, Cheryl L Rock and Loki Natarajan. Longitudinal associations between timing of physical activity accumulation and health: Application of functional data methods, submitted to Statistics in Biosciences”. The dissertation author was the primary author of this paper.

Chapter 5, in full, has been prepared for submission for publication as ”Wenyi Lin and Loki Natarajan. Multilevel longitudinal functional principal component analysis”. The dissertation author was the primary author of this paper.

Chapter 6, in full, has been prepared for submission for publication as ”Wenyi Lin and Armin Schwartzman. Spatial-temporal modeling and spatial inference using NA-CORDEX climate data”. The dissertation author was the primary author of this paper.

VITA

2015	Bachelor of Science in Information and Computing Science Zhejiang University, Hangzhou, China
2017	Master of Science in Biostatistics Johns Hopkins University, Baltimore, U.S.
2017–2022	Graduate Student Researcher University of California San Diego
2022	Doctor of Philosophy in Biostatistics University of California San Diego, U.S.

PUBLICATIONS

Wenyi Lin, Kyle Hasenstab, Guilherme Moura Cunha, and Armin Schwartzman. Comparison of handcrafted features and convolutional neural networks for liver MR image adequacy assessment. Scientific Reports 10, no. 1 (2020): 1-11.

Wenyi Lin, Michael C. Donohue, Philip Insel, Armin Schwartzman, and Wesley K. Thompson. ”Bayesian Multivariate Growth Mixture Modeling of Longitudinal Data: An Application to Alzheimer’s Disease Study.” bioRxiv (2021).

Wenyi Lin, Jingjing Zou, Chongzhi Di, Dorothy D. Sears, Cheryl L Rock and Loki Natarajan. Longitudinal associations between timing of physical activity accumulation and health: Application of functional data methods, Statistical Methods in Medical Research. (Submitted)

Donohue, Michael C., Gopalan Sethuraman, Oliver Langford, **Wenyi Lin**, Philip Insel, Wesley K. Thompson, Rema Raman, Reisa A. Sperling, and Paul S. Aisen. ”Alternatives to MMRM for preclinical Alzheimer’s clinical trials: Clinical trial design and implementation.” Alzheimer’s & Dementia 16 (2020): e044915.

Datta, Abhirup, **Wenyi Lin**, Amrita Rao, Daouda Diouf, Abo Kouame, Jessie K. Edwards, Le Bao, Thomas A. Louis, and Stefan Baral. ”Bayesian estimation of MSM population size in Côte d’Ivoire.” Statistics and Public Policy 6, no. 1 (2019): 1-13.

Youssef, Fady, Lin Liu, **Wenyi Lin**, Ranier Bustamante, Ashley Earles, Santhi Swaroop Vege, Thomas J. Savides et al. ”Sa1645-Comparison of the Fukuoka Consensus Guidelines and the American Gastroenterological Association Guidelines as Predictors for Developing Pancreas Cancer Among Patients with Pancreatic Cysts.” Gastroenterology 154, no. 6 (2018): S-340.

ABSTRACT OF THE DISSERTATION

Applications of Longitudinal and Functional Data Analysis

by

Wenyi Lin

Doctor of Philosophy in Biostatistics

University of California San Diego, 2022

Professor Armin Schwartzman, Chair

The objective of this thesis is to utilize statistical methods for longitudinal and functional data analysis, where repeated measures are observed. This dissertation is comprised of three main applications. In the first study, we aimed to model developing trajectories of multiple biomarker outcomes simultaneously and predict latent disease stages of Alzheimer's disease, using data from the multicohort longitudinal Alzheimer's Disease Neuroimaging Initiative (ADNI) study. For sparsely observed outcomes over a relatively short period of time, we proposed a flexible Bayesian multivariate growth mixture model to identify distinct longitudinal patterns of disease progression and three latent trajectory patterns among ADNI participants that overlap with but do not correspond one-to-one with diagnostic status.

In the second study, we observed densely sampled physical activity (PA) data acquired from accelerometers, which are widely used for tracking human movement and provide minute-level PA records, and intended to explore its association with health outcomes related with obesity. We developed multiple multilevel functional models, based on functional principal component analysis (FPCA) approaches, to study the hierarchical structure and temporal patterns of daily PA data from 245 overweight/obese women at three visits over a one-year period. We found that the health outcomes are strongly associated with PA variation and revealed that the timing of PA during the day can also impact changes in outcomes.

We further extended the implementation to densely sampled data in both spatial and temporal domains, focusing on modeling and testing climate change effects using regional climate data from the North American Coordinated Regional Downscaling Experiment (NA-CORDEX). We constructed spatial-temporal models which incorporate geographically weight regression strategies, and performed spatial inference on trend parameters regarding temperature and precipitation change. As a result, we identified regions with significant climate change in California, Colorado and Kansas, and compared similarities and differences of global warming effects in local scales.

Chapter 1

Introduction to Longitudinal and Functional Data Analysis

1.1 Introduction

Study designs with repeated measurements, where records are measured on the same subject at different times or under different conditions, are consistently gaining interests currently. These measurements can be collected sparsely or densely, temporally or spatially, regularly or irregularly, thus posing challenges and also great possibility for statistical modeling and inference. In this thesis, we aimed to derive analysis strategies for studies with varying types of repeated measured data structure.

Both longitudinal data analysis (LDA) and functional data analysis (FDA) deal with data consisting of repeated measurements of objects over time. In FDA, the analysis and theory of data can be further derived to images or other general objects. The major difference between them comes from the format of data points distribution, or in other words, the type of sampled 'grids'. Specifically, 'grids' treated in the LDA studies are typically more sparse, and often irregularly spaced, whereas FDA typically models 'grids' which are densely recorded in an equally spaced domain. Rice (2004) [152] provided an interesting overview for comparing the perspectives and methods of LDA and FDA. It was demonstrated that though with differences, there are many common intentions among them, including estimation of individual curves or functions from noisy measures, characterizing both homogeneity and patterns of variability and assessment of the associations between curves and covariates. These similarities could be reflected from smoothing individual curves, in both LDA and FDA.

The LDA is aimed to characterize the change in response over time and the factors that affect change, which plays a key role in epidemiology, clinical research, and therapeutic evaluation [44]. Compared with a cross-sectional study, where the response is measured once per subject and only the between-individual differences can be obtained, LDA can capture within-individual change from repeated measures on individuals. Therefore, one main characteristic of LDA is to estimate the correlation within the observations for each subject for drawing valid inferences. In addition, the number of repeated observations and their timing, are commonly not the same for each subject, which is known as unbalanced data. Let y_{ij} denote response variable for subject i at time j, where $i = 1,\ldots,n$ and $j = 1,\ldots n_i$. Usually, n_i is a relatively small number in LDA, especially when compared with the number in FDA modeling.

Functional data, on the contrary, normally have high or even infinite dimensional structure. A one-dimensional stochastic process, referred to as the first-generation functional data in Wang et al. (2016) [181], is the most general form of functional data. It can be represented as a random sample of independent real valued functions $\{X_1(t),...,X_n(t)\}$ over a compact interval $I = [0,T]$ on the real line and assumed to be in a Hilbert space such as the space of square-integrable functions of $L^2(I)$. Specifically, a stochastic process $X(t)$ is called $L^2(I)$ if and only if it satisfies $E(\int_I X^2(t)dt) < \infty$. The time interval where observations are recorded can be dense or sparse, whereas a sparse functional data normally arise in longitudinal studies. The conventional functional data are considered densely sampled when the number of observations converges to infinity, the mean function $EX(t)$ can achieve the parametric $\sqrt{n}$ convergence rate for standard metrics, such as the L^2 norm. For all n subjects, functional data are usually recorded on the same dense time grid of ordered and regular time grid $t_1,\ldots,t_p$ and $p \to \infty$ applies to typical functional data. Next-generation functional data are part of complex data objects, which can be multivariate, be correlated, or involve images or shapes.

In reality, measurement errors are commonly observed in both LDA and FDA, such as random fluctuations around a smooth trajectory or actual errors in the measurements. These random noises are often assumed to be independent across and within subjects. Because repeated measurements were recorded for each subject, LDA and FDA naturally better accommodate measurement errors than cross-sectional studies. In addition, smoothing the raw curves using expansions in basis functions are commonly seen in both LDA and FDA, but from different points of view. In FDA, the number of basis functions are chosen for good approximation properties and thus can be very large. However, the dimension of subspace is typically small in LDA, which may only include linear or quadratic functions.

Another difference in LDA and FDA is about the estimation of covariance matrix. In FDA, because the data are sampled on a regular grid, the covariance structure can be estimated directly from observations using the sample covariance. Therefore, estimating covariance is of direct interest rather than viewed as a nuisance parameter in FDA. However, since the grid points are sparse and perhaps irregularly spaced in LDA, this covariance structure is not of primary interest and is usually with some specific formats, such as independent and autoregressive correlations.

Currently, there is no significant gap between LDA and FDA studies. FDA-based approaches, such as the functional principal component analysis, are commonly implemented in LDA. Meanwhile, the methods taken in LDA, in particular ideas borrowed from mixed and random effects modeling, are widely utilized by recent FDA studies. Hence, the criteria of constructing statistical models and performing inference for LDA or FDA intrinsically rely on study objectives and method performances.

1.2 Objectives and Research Questions

The main objective of this thesis is to investigate statistical modelling and inference of varying types of data related with LDA and FDA. Specifically, methods and algorithms were proposed to respect the original data structure and extract features for answering corresponding scientific questions. Data sources were from Alzheimer's Disease Neuroimaging Initiative (ADNI) study, the MENU study and North American Coordinated Regional Downscaling Experiment (NA-CORDEX) Program. Their related research questions are presented briefly in this section.

1.2.1 Alzheimer's Disease Biomarker Data

The main objective of this project is to model Alzheimer's disease developing trajectories and predict latent disease stages. Data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) study (http://adni.loni.usc.edu/) were used as an illustration example for understanding the Alzheimer's disease pathological progression. This multicohort longitudinal study started in 2004, recording over 1900 volunteers at varying disease progression stages. Difficulties arise in capturing both between- and within-subject variation, as well as estimating overall and individual trajectories across various biomarker domains and disease stages. We aimed to model the trajectories of multiple biomarker outcomes by means of Bayesian multivariate growth mixture model and simultaneously predict individual's disease stage.

1.2.2 Accelerometer data

In this study, we focused on statistically modeling the physical activity data and exploring its association between health outcomes related with obesity. The data were collected from the MENU weight-loss study conducted in UCSD, which contains 245 non-diabetic and overweight women [97, 154]. Specifically, physical activity data were measured via accelerometer sensors, based on activity counts derived from high-resolution acceleration signals in minute-level. Challenges mainly arise in deriving solid models to extract features from these densely sampled data, together with accounting for the unbalanced hierarchical structure in predictors and outcomes. In our analysis, we treated PA as functional data with multiple levels and implemented the multilevel/longitudinal functional principal component analysis and functional longitudinal regression models to investigate the complex data structure.

1.2.3 Climate data

The main objective of this project is to model climate change effects in regional scales incorporating both temporal-spatial modeling and inference strategies. Climate data were downloaded from the North American Coordinated Regional Downscaling Experiment (NA-CORDEX) [117], including temperature and precipitation data. The dataset provided both observed historical and simulated future time series data over North America from 1950 to 2100, which were generated from Regional Climate Models. Challenges typically come from the structure of climate, for which we need perform the model construction and inference in both densely-sampled temporal and spatial domains. We considered to build spatial-temporal models for both temperature and precipitation data and further provided statistical inference on estimated parameters via the Coverage Probability Excursion (CoPE) sets approach.

1.3 Organization Structure

This dissertation is structured as follows.

Chapter 1 provided an overview of longitudinal and functional data analysis. An outline of this dissertation was laid out.

Chapter 2 introduced both the scientific and methodological background related with each research question included in this thesis.

In Chapter 3, we introduced a longitudinal study using ADNI data and derived a Bayesian multivariate growth mixture model. The proposed methods simultaneously model the trajectories of multiple Alzheimer's Disease biomarkers and provide predictions of latent states of disease progression.

In Chapter 4, methods related with functional principal component analysis were provided to analyze functional physical activity and its association with related health outcomes, using data from the MENU study. Chapter 5 further developed the proposed models in Chapter 4 to more general cases and performed a series of simulation studies.

In Chapter 5, we proposed strategies to construct spatial-temporal models, using climate data from NA-CORDEX project, and implemented spatial inference on climate changing effects.

Chapter 2

Scientific and Methodological Background

2.1 Multivariate Growth Mixture Models and the Application with Longitudinal Alzheimer's Disease Biomarker Data

2.1.1 Introduction

Alzheimer's disease (AD) is a slowly progressive neurodegenerative disease and the most common cause of dementia - a continuous decline in thinking, behavioral and social skills that may cause final loss of cognitive functions. Currently, among approximately 50 million people worldwide with dementia, between 60% to 70% of all cases are estimated to have AD [113]. In the United States, estimated 6.2 million people aged 65 and older are living with AD. Furthermore, because of the increasing number of Americans aged 65 and older, the annual number of new cases of AD and other dementias is about to double by 2050 [187]. Meanwhile, there's still no cure for AD, but some treatments may change disease progression. In 2021, the Food and Drug Administration (FDA) has granted accelerated approval for the first putative disease-modifying therapy [43, 135]. Therefore, AD is recognized as a major epidemic and poses great medical challenges [71].

A diagnosis of definite AD requires direct and accurate analysis of brain tissue samples, which can be obtained either at autopsy or from a brain biopsy. In 2007, new knowledge and guidelines about the prodromal symptomatic stage of AD were incorporated by an International

Working Group (IWG) [37]. And in 2011, Alzheimer's Association and the National Institute on Aging (NIA) issued four diagnostic criteria and guidelines for AD that focus on three stages of AD, inclduing dementia due to Alzheimer's, mild cognitive impairment (MCI) due to Alzheimer's and preclinical Alzheimer's [81, 115, 3, 167]. Harmonized diagnostic criteria for AD were later proposed to reach consensus between two sets of diagnostic criteria from IWG and NIA [126].

Biomarkers are key indicators of specific changes that characterise AD progression. Specifically, the pathological confirmation of AD requires presence of amyloid beta ($A\beta$) peptides and the evidence of neurofibrillary tangles (NFTs) of protein tau [3], which also differentiates AD from other forms of dementia. Meanwhile, dementia also represents the end stage of a long time of accumulation of related pathological changes [83]. Besides $A\beta$ and tau, atrophy measured by structural magnetic resonance imaging (MRI) is regarded as a powerful biomarker of AD state and progression [179, 52]. It provides rich information in estimating tissue damage or loss in vulnerable brain regions, such as the hippocampus and entorhinal cortex, which are predictive of progression of MCI to AD. Even though the exact mechanisms related with cognitive impairments are still largely unknown, Jack et al. (2010, 2013) proposed that major AD biomarkers become abnormal in a temporally ordered manner, with a shape of sigmoidal function of time [83, 82]. It starts with abnormality in amyloid biomarkers, followed by biomarkers of neurodegeneration related with tau, then clinical symptoms. Figure 2.1 presents the patterns of these biomarkers changing from cognitive normal to dementia over the course of AD.

In addition to biomarkers, the Apolipoprotein E (APOE) gene, especially the $\varepsilon4$ allele, is believed to be the most common genetic risk factor for both early-onset and late-onset AD [142, 102]. Comprehensive evaluations suggest that the gene is closely associated with decreased CSF $A\beta$ levels and higher tau-related levels. Meanwhile, carriers of APOE $\varepsilon4$ are at increased risk for numerous structural and functional brain changes associated with AD, before clinical symptoms become evident [102, 103].

In this project, we focused on developing and applying innovative analytical tools to determine the multivariate dynamics of long-term trajectories of AD-related biomarkers.

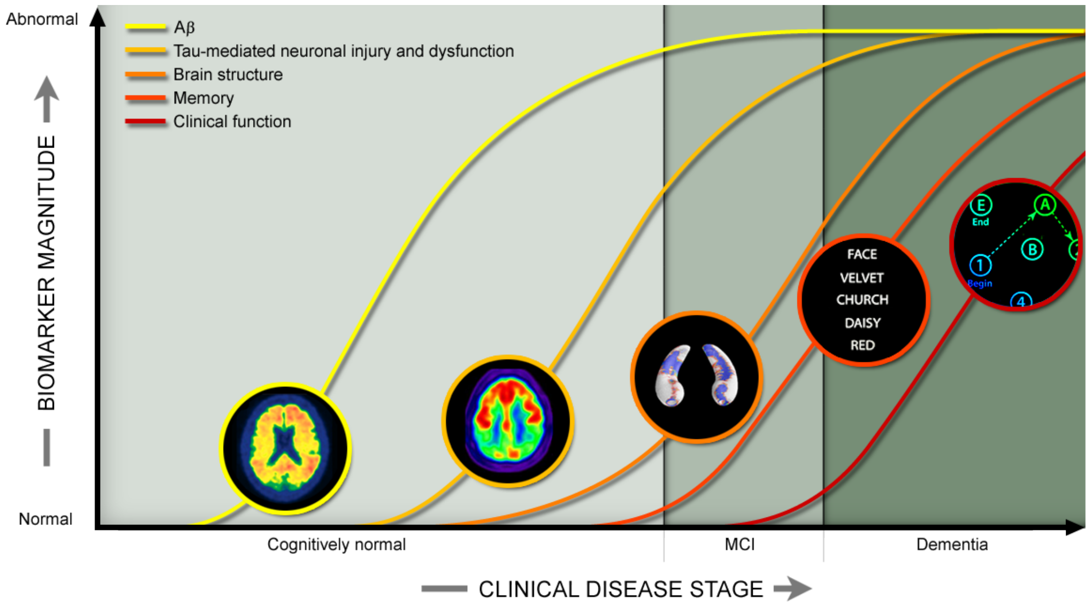


Figure 2.1. The temporal changes of biomarkers along the cognitive continuum from normal to abnormal. Illustration adapted from ADNI (http://adni.loni.usc.edu/study-design/).

2.1.2 Motivation: ADNI

The data used for this study came from the Alzheimer's Disease Neuroimaging Initiative (ADNI), one of the world's leading studies designed to develop clinical, imaging, genetic, and biochemical biomarkers for the early detection and tracking of AD (https://adni.loni.usc.edu/, [128]). This ongoing longitudinal study was initiated in 2004 and has already implemented four phases of the study, including ADNI-1, ADNI-GO, ADNI-2 and ADNI-3. To date, the study has recruited over 1900 participants at different AD stages, including normal cognition (CN), early or late MCI and early AD. In addition, many of the participants enrolled in ADNI 1 are also followed in subsequent studies, and new participants have been enrolled in each of the subsequent phases. It prolonged follow-up times of some participants and enriched the database in the ADNI study, while on the other hand, it also brought challenges of missing data in the longitudinal study.

Certain inclusion criteria need to be met for all participants to be enrolled in the ADNI study, including age between 55 and 90, Modified Hachinski score ≤ 4, Geriatric Depression

Scale less than 6, permitted medications stable for at least 4 weeks prior to screening, etc. Furthermore, there were specific criteria for MCI and AD patients, such as memory complaint from patient or study partner and the disease stage identified with multiple objective scores and ratings.

2.1.3 Linear Mixed Effects Model

Developing analytical models of longitudinal disease trajectories was underscored in ADNI study, where the outcome variables were measured repeatedly for each study participant. Linear mixed effects models (LMEs) are commonly implemented to investigate changes over time in longitudinal studies. In this section, we provide a brief overview of the LME modeling based on works in Laird and Ware (1982) [95].

Let y_{ij} denote a biomarker response for subject $i, i = 1, 2, \ldots, n$ at visit $j, j = 1, 2, \ldots, m_i$ and $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{im_i})'$ be the vector of response. The LME model has the form,

$$\boldsymbol{y}_i = X_i\boldsymbol{\beta} + Z_i\boldsymbol{a}_i + \boldsymbol{\varepsilon}_i, \tag{2.1}$$

where X_i is the $m_i \times q$ fixed effects covariate matrix and $\boldsymbol{\beta}$ is the corresponding coefficient vector with length p. Z_i is the random effects model matrix with dimension $m_i \times p$. $\boldsymbol{a}_i$ is the $p \times 1$ vector of subject-specific random effects and follows $\boldsymbol{a}_i \sim N(0_p, \Sigma)$. $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \ldots, \varepsilon_{im_i})$, with $\boldsymbol{\varepsilon}_i \sim N(0_{m_i}, \sigma_\varepsilon^2 I_{m_i})$ and is independent of $\boldsymbol{a}_i$'s. Conditional on $\boldsymbol{a}_i$, observations y_{ij} are independent with each other with

$$E(\boldsymbol{y}_i|\boldsymbol{a}_i) = X_i\boldsymbol{\beta} + Z_i\boldsymbol{a}_i \quad \text{and} \quad Var(\boldsymbol{y}_i|\boldsymbol{a}_i) = \sigma_\varepsilon^2 I_{m_i}$$

If the random effect $\boldsymbol{a}_i$ is considered as a part of error term in $(y_{ij}$, i.e. $e_{ij} = Z'_{ij}\boldsymbol{a}_i + \boldsymbol{\varepsilon}_{ij}$, the marginal distribution of $\boldsymbol{y}_i$ is given as $\boldsymbol{y}_i \sim N(X_i\boldsymbol{\beta}, G_i)$, where $G_i = Z_i\Sigma Z'_i + \sigma_\varepsilon^2 I_{m_i}$. The marginal representation in fact reduces the linear mixed effects model to a general linear model. Denote all covariance parameters in G_i by a parameter vector $\boldsymbol{\theta}$.

Given covariance parameters $\boldsymbol{\theta}$ are known, the maximum likelihood estimator (MLE) of fixed effects $\boldsymbol{\beta}$ and the best linear unbiased predictors (BLUP) of random effects $\boldsymbol{a}_i$ are provided as,

$$\hat{\boldsymbol{\beta}} = (\sum_{i=1}^{n} X_i' G_i^{-1} X_i)^{-1} \sum_{i=1}^{n} X_i' G_i^{-1} \boldsymbol{y_i}, \tag{2.2}$$

and

$$\hat{\boldsymbol{a}}_i = \Sigma Z_i' G_i^{-1} (\boldsymbol{y}_i - X_i \hat{\boldsymbol{\beta}}). \tag{2.3}$$

It can be seen that both $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{a}}_i$ are linear functions of $\boldsymbol{y_i}$'s, therefore, their corresponding standard errors are derived as,

$$\hat{V}_\beta = Var(\hat{\boldsymbol{\beta}}) = (\sum_{i=1}^{n} X_i' G_i^{-1} X_i)^{-1}, \tag{2.4}$$

and

$$Var(\hat{\boldsymbol{a}}_i) = \Sigma Z_i' (G_i^{-1} - G_i^{-1} X_i (\sum_{i=1}^{n} X_i' G_i^{-1} X_i)^{-1} X_i' G_i) Z_i \Sigma. \tag{2.5}$$

However, Equation 2.5 only assesses the error of estimation and underestimates the variation in $\hat{\boldsymbol{a}}_i - \boldsymbol{a}_i$, because it ignores the variation of $\boldsymbol{a}_i$. In practice, the standard errors of prediction for $\hat{\boldsymbol{a}}_i$ are based on

$$\hat{V}_a = Var(\hat{\boldsymbol{a}}_i - \boldsymbol{a}_i) = \Sigma - \Sigma Z_i'(G_i^{-1} - G_i^{-1}X_i(\sum_{i=1}^{n} X_i'G_i^{-1}X_i)^{-1}X_i'G_i)Z_i\Sigma. \tag{2.6}$$

When the covariance parameters $\boldsymbol{\theta}$ are unknown, we use the empirical MLE for $\hat{\boldsymbol{\beta}}$ and empirical BLUP for $\hat{\boldsymbol{a}}_i$, by replacing the G_i with $\hat{G}_i$. $\boldsymbol{\theta}$ are estimated using restricted maximum likelihood (REML), which are generally preferable to maximum likelihood estimates. In fact, the REML estimate maximize based on any full-rank set of error contrasts $\boldsymbol{u'y}$, such that $E(\boldsymbol{u'y}) = 0$. The full details of REML estimates derivation can be found in original Laird & Ware (1982) paper, which is involved with the Bayesian approach with unified treatment of estimation and computation.

These estimates are denoted as $\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}})$ and $\hat{\boldsymbol{a}}_i(\hat{\boldsymbol{\theta}})$. Meanwhile, estimates of the standard errors of $\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}})$ and $\hat{\boldsymbol{a}}_i(\hat{\boldsymbol{\theta}})$ are obtained by substituting $\hat{\boldsymbol{\theta}}$ in Equation 2.5 and 2.6.

In addition, unlike a general linear model, for which we are often interested in estimating the mean response profile $E(\boldsymbol{y}_i) = X_i\boldsymbol{\beta}$, with LMEs we can predict individual response profile $\hat{\boldsymbol{y}}_i = X_i\hat{\boldsymbol{\beta}} + Z_i\hat{\boldsymbol{a}}_i$. It can be found from this equation that the individual prediction $\hat{\boldsymbol{y}}_i$ is in fact a weighted average of estimated population mean $X_i\hat{\boldsymbol{\beta}}$ and observed individual profile $\boldsymbol{y}_i$, expressed as $\hat{\boldsymbol{y}}_i = (\hat{\sigma_\varepsilon}^2\hat{G}_i^{-1})X_i\hat{\boldsymbol{\beta}} + (1 - \hat{\sigma_\varepsilon}^2\hat{G}_i^{-1})\boldsymbol{y}_i$. These weights can be interpreted as the ”proportion” of variability in $\boldsymbol{y}_i$ within subjects and the ”proportion” of variability between subjects.

Therefore, advantages of using LMEs modeling include, a) it allows partitioning of variation into between-subject and within-subject variation b) it provides prediction of response at individual-level, c) it can also handle unbalanced designs, such as different observation times.

2.1.4 Growth Mixture Modeling

The LMEs modeling, which models the longitudinal developmental trajectories of biomarkers for individuals in the ADNI study, is also known as the growth curve modeling. Conventional growth modeling applications are able to model within-person change and between-person differences in change [133], while they normally assume that the sample is drawn from a single population. Therefore, a latent class or growth mixture modeling (GMM) approach [130] was introduced to capture information about between-person differences in within-person change, by incorporating unobserved heterogeneity within the larger population in the model.

Instead of assuming that the growth trajectories of all individuals are generated from common parameters, GMM allows for differences in growth parameters across unobserved sub-populations. Specifically, it provides a framework for identifying post-hoc stages and describes the group differences in change. Extending Equation 2.1, a GMM can be written as,

$$f(\boldsymbol{y}_i|\boldsymbol{X}_i, \boldsymbol{Z}_i) = \sum_{k=1}^{K} P(C_i = k) f(\boldsymbol{y}_i|\boldsymbol{X}_i, \boldsymbol{Z}_i, C_i = k), \tag{2.7}$$

where $f(\boldsymbol{y}_i|\boldsymbol{X}_i, \boldsymbol{Z}_i, C_i = k)$ has the same distribution as in Equation 2.1, conditional on $C_i = k$. $P(C_i = k)$ is the probability that individual i belongs to class k and satisfies $\sum_{k=1}^{K} P(C_i = k) = 1$. It can be estimated using either the maximum likelihood [99] or Bayesian methods [145]. In essence, both approaches implement iterative procedures to obtain parameter estimates, as well as the posterior estimates of the $P(C_i = k)$ for identifying individuals' sub-groups.

One typical challenge for constructing a GMM is to determine the number of classes. In Ram & Grimm (2009) [145], they provided detailed instructions for performing the model selection. The first criterion is to check the validity and interpretability of all potential models. Secondly, the remaining models can be compared using relative fit information criteria, such as finding models with the smallest Bayesian information criteria (BIC) value or a significant likelihood ratio test statistic [105]. With a Bayesian approach, we may also consider the Widely Applicable Information Criterion (WAIC) [184]. Furthermore, models can be evaluated with respect to the classification accuracy, if ground truths are known for some classifications.

The basic implementation of the GMM was extended to accommodate with our data structure and analysis goals in Chapter 3. In particular, we constructed a Bayesian multivariate growth mixture model (BMGMM), which can efficiently fit longitudinal trajectories and identify latent classes via Bayesian methods.

2.2 Functional principal component analysis and its application in accelerometer-measured physical activity data

2.2.1 Introduction

Physical inactivity is a major public health problem in modern society, with national reports from CDC showing that only 1 in 4 US adults meet recommended amounts of physical activity (PA) [24]. Combined with sedentary behavior, they are believed to be strongly related with an increased risk of chronic health conditions, such as obesity, type 2 diabetes, heart disease, and certain types of cancers and all-cause mortality [151]. Regular PA, such as walking, cycling or doing sports, provides significant benefits for health and is recommended by the majority of health-related facilities [136].

Both subjective and objective assessment tools have been developed to measure PA levels, which is critical for providing any forms of intervention. Subjective methods, such as diaries, questionnaires and surveys, can be easily obtained and are time-efficient. However, these methods normally depend on individual‘s own observation and subjective evaluation, which can make the reports inconsistent, biased or with inadequate reliability [144, 68].

On the other hand, objective approaches use wearable, or body-fixed motion sensors, to provide PA estimates. For instance, accelerometry-based sensors can measure the accelerations of objects in motion along reference axes. Therefore, because acceleration is correlated with external force, it can reflect intensity and frequency of human movement [194]. Currently, accelerometers have been widely utilized as useful and practical sensors for measuring and assessing PA in not only clinical/laboratory practices but also free-living environments [111].

In this project, we focused on developing and implementing multiple statistical models to explore the associations between accelerometer measured PA and health-related outcomes.

2.2.2 Motivation: The MENU weight-loss study

The motivation of this project came from the MENU study. It was conducted in the NIH-funded Transdisciplinary Research on Energetics and Cancer Center (TREC) at UCSD (2011-2017) and was a one-year behavioral intervention study to investigate the role of dietary macronutrient composition on weight loss and metabolic, hormonal and inflammatory factors in overweight/obese women [97, 154]. Clinical measurements were taken at three time points, the baseline, 6 and 12 months, when participants were instrumented with the accelerometer for 7 days during all waking hours. 245 overweight non-diabetic women were recruited in this study.

Accelerometry-based PA data was collected using the GT3X Actigraph (ActiGraph, LLC; Pensacola, FL). The device provides second-by-second estimates of activity that can be categorized into minutes spent in sedentary, light, moderate, and vigorous activity using calibration thresholds [158]. Non-wear time was defined using pre-defined algorithms, identifying intervals of 90 minutes or longer in which all count values were 0. Days with more than 10 hours of wear time were saved.

At each data collection clinic visit, body mass index (weight in kilograms/ height in m^2) was calculated. Other related health outcomes, including glucose, total cholesterol, triglycerides, high-density lipoprotein cholesterol (HDL-C) levels, low-density lipoprotein cholesterol (LDL-C) levels, high-sensitivity C-reactive protein (CRP), insulin values, etc., were acquired at each visit.

Therefore, the major goal is to derive statistical analysis models to investigate the patterns of PA in the MENU study using data collected from accelerometers, as well as the relation with health outcomes which quantify overweight and obese status.

2.2.3 Statistical analyses on accelerometry-based physical activity

Currently, many health studies use wearable accelerometry-based devices, for they are fairly inexpensive and convenient for participants to wear for a period of time [47]. As it was summarized in Zhang et al. [198], two main categories of statistical analyses were considered for utilizing accelerometer data.

The first approach is to derive algorithms for translating the acceleration signal into estimates of metrics. For instance, the raw acceleration data acquired from ActiGraph devices can be transformed to activity counts [19, 73] or a more complex vector magnitude using count data from 3-axes [11]. Bai et al. [8] further summarize a general workflow to translate the acceleration signal for identifying different activity types. Meanwhile, related software implements were developed to handle the increasing need of more accurate and precise estimates of acceleration signal, such as R packages `PhysicalActivity` and `pawacc` [27, 56].

The second interest is to find out association patterns regarding the obtained summary estimates or behaviors. Since the devices can save the PA data for multiple days or weeks, longitudinal data analysis was always incorporated. Carson V et al. [23], for example, explored the longitudinal association between time spent in vigorous-intensity PA and cardiometabolic risk factors in youth. Harding et al. [63] described the longitudinal changes of sedentary time and physical activity through adolescence. Furthermore, more complex longitudinal data analysis models were generated to handle multivariate outcomes and different variable types, however, most of these studies mainly utilized summary metrics of PA data.

2.2.4 Functional principal component analysis

With respect to the form of measured count data from accelerometers, which are often continuous curves or functions, functional data analysis (FDA), especially the functional principal components analysis (FPCA), is considered as a potent statistical approach [148]. It is based on principal component analysis (PCA) [87], which is an essential unsupervised dimension reduction tool for multivariate data analysis. PCA aims to find a set of orthogonal components, or principal components (PCs), which are linear combinations of original features and preserves the maximum amount of variation from them. The idea of PCA was then extended to functional data, by replacing vectors by functions, as well as a series of corresponding notations [162]. Therefore, the major goal of FPCA is to decompose the space of curves into principal directions of variation.

For one observed PA curve $X_(t)$ at time $t \in \mathscr{D}$, it is assumed to be a squared integrable random function with mean $E\{X(t)\} = \mu(t)$ and covariance function $cov\{X(s), X(t)\} = K_X(s,t)$. Mercer's theorem implies the spectral decomposition of $K_X(s,t)$:

$$K_X(s,t) = \sum_{k=1}^{\infty} \lambda_k \phi_k(s) \phi_k(t),$$

where λ_k are the nonnegative eigenvalues with descending order and ϕ_k are the corresponding orthogonal eigenfunctions. The Karhunen and Loéve (KL) [88, 107] expansion of $X(t)$ is provided as $X(t) = \mu(t) + \sum_{k=1}^{\infty} \xi_k \phi_k(t)$, where $\xi_k = \int_{t \in \mathscr{D}} \{X(t) - \mu(t)\} \phi_k(t) dt$. Here ξ_k's are referred as principal component scores, with mean zero and variance λ_k and are mutually uncorrelated. The KL expansion is the fundamental platform for FPCA and facilitates dimension reduction for in practice, only the first few eigenvalues and eigenfunctions are needed to provide a good approximation to the infinite sum.

For estimating the mean function, $\mu(t)$, and covariance function, $K(s,t)$, the method of moments (MoM) [89] approach was employed based on symmetric sums. With the empirical covariance function, the estimation of the eigenfunctions and eigenvalues based on spectral decomposition is straightforward, while the tricky part is to select optimal number of components for the approximation of the full KL expansion. Practical solutions include using cross validation [153] or Akaike information criterion (AIC) and Bayesian information criterion (BIC) [195]. A simpler way is to keep the first few PC components, which fulfill a cumulative fraction of variance explained threshold, such as 0.9. With the estimated $\mu(t)$ and $\phi_k(t)$, the principal component scores ξ_k's can then be estimated by direct numerical integration.

Another issue with the implementation of FPCA is the existence of measurement error in functional observations, which is commonly seen in many applications. It can be addressed by smoothing the raw data before applying FPCA [147, 164]. An alternative approach is to smooth the covariance function [195] and furthermore, Xiao et al. [192] provided a computationally faster sandwich smoother on the empirical covariance matrix, which is in particular applicable in high- dimensional settings.

The FPCA model provides the fundamental technique to functionally model data with more than one level, based on the framework of functional linear mixed model (FLMM) [124]. Extensions include but not limit to Multilevel FPCA for extracting core intra- and inter-subject geometric components in the two-level nested model [33]; Longitudinal FPCA for modeling longitudinal dynamics [60]; High Dimensional Multilevel FPCA for accommodating high-dimensional setting [199]; and Structured FPCA for handling both nested and cross designs. Details and comparisons of these models will be further discussed in Chapter 4 and 5.

2.2.5 Scalar-on-function regression models

Scalar-on-function regression models are extensions of the traditional multivariate linear model that associates functional predictors with scalar outcomes, and in our study, we are interested in exploring the relationship between PA effects on overweight-related health outcomes. The topic has been investigated in a rich literature.

For outcomes with normal distribution, the functional linear model (FLM) was reviewed in Ramsay and Silverman [148]. A linear model with scalar outcome $Y \in \mathscr{R}$ and vector predictor $\mathbf{X}$ with length p is provided as

$$Y = \alpha_0 + \mathbf{X}\boldsymbol{\beta} + \varepsilon, \tag{2.8}$$

where α_0 and $\boldsymbol{\beta}$ are the regression coefficients. ε is the random noise with mean zero and finite variance. Replacing the vector predictor $\mathbf{X}$ with the centered functional predictor $\tilde{X}(t) = X(t) - \mu(t)$ introduced in Section 2.2.4, the coefficient vector $\boldsymbol{\beta}$ now becomes a coefficient function $\beta(t)$ and the corresponding FLM is expressed as,

$$Y = \alpha_0 + \int_{t \in \mathscr{D}} X(t)\beta(t)dt + \varepsilon. \tag{2.9}$$

By expanding $X(t)$ and $\beta(t)$ in the same functional basis, such as the eigenbasis, it leads to $X(t) = \sum_{k=1}^{\infty} \xi_k \phi_k(t)$ and $\beta(t) = \sum_{k=1}^{\infty} \beta_k \phi_k(t)$. The resulting model can be seen as equivalent to the traditional linear model with the form,

$$Y = \alpha_0 + \sum_{k=1}^{\infty} \beta_k \xi_k + \varepsilon, \tag{2.10}$$

where the infinite sum on the right side can be replaced with a finite sum of the first few terms as discussed in Section 2.2.4. Since the scores ξ_k's are treated as the predictors in the model, it is called a functional principal components regression (FPCR) model [150]. Model 2.9 can be extended to include multiple functional predictors $X_1(t), \ldots, X_p(t)$, as well as other vector predictors $\mathbf{Z}_1, \ldots, \mathbf{Z}_q$.

Equation 2.9 and 2.10 provide the basis to settings where repeated functional observations or scalar outcomes are collected for each subject in a sample. Crainiceanu et al. [31] derived the Generalized Multilevel FLM for modeling data with only multilevel functional structure in predictors, based on the data-driven basis derived from Multilevel FPCA [33]. Goldsmith et al. [59] proposed longitudinal penalized functional regression to data include both repeated measures in scalar outcomes and functional predictors. Furthermore, based on the Longitudinal FPCA setup [60], Gertheiss et al. [57] constructed two versions of Longitudinal FPCR by either using scores or spline-based curves as functional predictors when data are observed at multiple visits.

2.3 Spatial-Temporal Modeling and Spatial Inference Using Regional Climate Data

2.3.1 Introduction

Essentially, climate can be viewed as the statistics of weather over an arbitrarily defined time span [29]. The most common statistic is the average of climate conditions, such as temperature, precipitation, the frequency and intensity of extreme events. Other statistics include variance, no trend, etc. In the past, climate models were built under a key assumption termed stationarity, which indicates that these statistics of climate conditions do not vary with a sufficiently long time period and will be similar to the recent past in the future [91]. However, this stationarity assumption is violated as a result of human-induced climate change. Currently, the effects of global climate change are apparent across a wide range of extreme weather events, such as prolonged periods of heat, heavy rainfall, and severe droughts or floods in some regions [118]. The global climate is changing because of human emissions of heat-trapping gases. Overall, a global warming of approximately 0.85 $^\circ C$ has occurred over the period 1880 to 2012 [119]. The situation is even worse in Arctic areas. Alaskan temperatures have increased nearly twice as fast as those in the contiguous United States, and is expected to continue in the future [1]. Meanwhile, as the temperature increases, more water evaporates from aquatic systems, i.e. oceans, lakes, etc., which have already caused more heavy rainfall and precipitation events over the past 50 years.

Climate models are an extension of weather forecasting in long time spans, which are used to project the possible future evolution and to understand the climate system itself [70]. They are built to divide the atmosphere, ocean, and land surface up into a large amount of discrete cells to solve numerical equations representing the physical, biological, and chemical phenomena.

Projections of the past and future climate on a global scale, calculated by global climate models (GCMs), have been extensively studied. In particular, GCMs incorporate measured values as forcing data to simulate the past climate, whereas for future projections, information from particular emission scenarios needs to be employed. The Intergovernmental Panel on Climate Change (IPCC) provided two sets of emission scenarios used in GCM simulations, including SRES and RCP scenarios [80, 79, 1]. Currently, GCMs have a spatial resolution of horizontal mesh ranging from 100-500 km and provide output with a 6-hour temporal frequency. Due to this relatively coarse spatial resolution and temporal scale, GCMs often fail to capture many features of regional and local scale estimates of climate change. The limitations are particularly critical in estimating extreme climate and research questions involved with local or regional topography and geographical features.

Downscaling techniques are needed to describe the local consequences, which extract high-resolution information from GCM output into projections that are more representative of regional-scale climate changes [70]. There are two principal ways to combine the information on local conditions with global change, including the empirical-statistical downscaling (ESD) and dynamical downscaling by means of regional climate models (RCMs). RCMs use the GCM output data as lateral boundary conditions at a much higher resolution and cover only selected portions of the globe. Typically, RCM integrations are run at 10-50 km horizontal resolution, thus they are able to provide more detailed characteristics of climate in a small area. They are different from statistical models, such as ESD models, which establish statistical relationships to translate GCM output into high-resolution future projections. Compared with RCMs, ESD models are relatively easy to implement and interpret but heavily rely on historical climate data and observed relationships. Though RCM integrations generation can be costly and computationally intensive, they are believed to be better representative of the current climate, as well as a more accurate projection of future climate. Therefore, many studies, including the North American Regional Climate Change Assessment Program (NARCCAP) [116] and North American Coordinated Regional Downscaling Experiment (NA-CORDEX) [117], have devoted many efforts to derive dynamically downscaled output and their results have been extensively used in varying impacts research.

Numerous statistical methods and models are used to analyse climate change. Because climate is a complex system with many variables acting nonlinearly on a wide range of spatial–temporal domains, both statistical modeling and uncertainty measurements are important in analysis. Hennemuth et al. [69] summarized multiple statistical methods for exploring various properties of the climate system, including general summary statistics of fundamental concepts, extreme value analysis, time series analysis, significance test, spatial-temporal methods, etc. Some of these methods will be introduced in more detail in Chapter 6.

In this project, we focused on deriving novel multivariate spatial-temporal models and performing corresponding spatial inference on climate data generated from regional climate models in NA-CORDEX program.

2.3.2 NA-CORDEX Program

NA-CORDEX is the North American part of the international Coordinated Regional Downscaling Experiment (CORDEX) program sponsored by the World Climate Research Program [116]. The general aim of the CORDEX program [58] is to downscale a number of GCM climate scenarios/predictions derived from the Coupled Model Intercomparison Program Phase 5 (CMIP5) archives, in a range of limited-area regions. Within the CMIP5 data archive, the following variables: temperature, eastward and northward wind velocity, specific humidity and surface pressure, were covered in 6 hourly global models. Additional measures, including daily and/or monthly values of sea surface temperature, sea-ice fraction, soil moisture and soil temperature, were also requested. According to the specifications of the international CORDEX program, the NA-CORDEX program produced downscaled simulations and data for North America using multiple statistical and dynamical downscaling models driven by an ensemble of GCMs.

In NA-CORDEX program, the RCM simulations were completed across an approximately 150-year time span, including the historical simulations driven by CMIP5 GCMs for 1950–2006, and the RCP 8.5 and 4.5 scenarios for 2006–2100. A balanced matrix of GCM-RCM combinations focused on 25 km and 50 km resolutions was incorporated for the dynamical downscaling. In total, the NA-CORDEX currently has 27 simulations, coming from seven different RCMs, forced by seven different CMIP5 GCMs, at two different resolutions (25 km and 50 km).

2.3.3 Spatial-Temporal Modeling

Modeling climate data simulated in the NA-CORDEX program normally involves with temporal and spatial data structures. Firstly, climate change in general refers to change over time, which makes time series modeling as an important field for climate analysis. It is obvious that the stationarity assumption is violated under current climate changing scenario, especially the global warming situation, therefore, the major task of the time series analysis is to model the data for estimating the parameters describing the trend, variability and other effects [34]. In climate modeling, a trend represents the gradual change of some variable, such as temperature and precipitation, over a period of time. Trend estimation can be realized via simple linear regression, using climate variables as outcomes and the time as the predictor [92]. Since the climate is a complex system, the trend analysis can be further extended to incorporate nonlinear or nonparametric terms [127].

Spatial interpolation is commonly used to transform one grid resolution to a different grid resolution, in both GCMs and RCMs. In climate models, the field typically refers to a two-dimensional geographical space (longitude–latitude). The two-dimensional bilinear interpolation is the simplest interpolation method, determining values by means of linear interpolation first in one direction, and then going into the other direction [177]. Additional information, such as the distance between points, can be incorporated to calculate values at unknown points based on a weighted average value at known points. The inverse distance weighting (IDW) approach was motivated by the weighted average, which assigns weights as inverse of the distance to each known point [64]. In addition, the spline interpolation, such as thin plate smoothing splines [171] and B-splines [98], further constructed a surface with minimal curvature and preserve small-scale properties.

The trend analysis and spatial interpolation provide the basis for spatial-temporal modeling of climate data, separately in the time domain and space domain. Interpretation of correlations of time series values from climate records and estimated parameters in a spatial domain can be provided via correlation maps [160]. Currently, more advanced methods, such as spatial-temporal Gaussian random field [14] and Generalized Additive Models [67], are widely implemented to explore the spatial-temporal data structure.

2.3.4 Spatial Inference

Hypothesis testing or significance testing is another important aspect of climate research. For instance, statistical significance of those linear slopes estimated from the trend analysis needs further assessment, with the null hypothesis set as no trend. At each location, the normally distributed parameters can be tested pointwisely via a Z-test or Student-t test. Furthermore, the bootstrapping provides a resampling method, which repeats the test parametrically or non-parametrically from resamples and determines a distribution of the test statistic [42]. These approaches are commonly seen in climate science due to the simplicity and efficiency [32], however, pointwise inference can be weakened because the familywise error rate (FWER) is not controlled.

A better approach is to account for the problem of multiple testing problems or adjust the spatial correlation between tests. The traditional way is to perform the Bonferroni correction for controlling the FWER, but it is too conservative when many tests are done in the whole spatial domain. Other approaches, such as FWER-controlling methods for spatial signal detection based on Gaussian random fields, were derived to implement simultaneous inference [51, 49]. Recently, Sommerfeld et al. [166] proposed a new approach for addressing these problems based on constructing Coverage Probability Excursion (CoPE) sets, which accounts for the simultaneous inference problem caused by multiple comparison in a space. Meanwhile, because the confidence sets are constructed by means of a multiplier bootstrap method, the CoPE method only requires mild assumptions about the data and is also very fast to apply.

Chapter 3

Bayesian Multivariate Growth Mixture Modeling of Longitudinal Alzheimer's Disease Biomarker Data

3.1 Introduction

Alzheimer's disease (AD), an irreversible neurodegenerative disorder, is the most common cause of dementia, affecting millions across the world. Biomarkers, such as those derived from neuroimaging, are commonly collected to characterize progression of AD pathology. Tracking the temporal evolution of multiple AD biomarkers may improve understanding of disease mechanisms [83, 82]. Several authors have also emphasized the importance of using biomarkers to predict risk of decline from mild cognitive impairment to dementia [114, 17]. Prediction of AD progression is potentially valuable for designing clinical trials and in clinical practice, as preventive measures can be more effective prior to onset of dementia, i.e. when patients are cognitively normal (CN) or have mild cognitive impairments (MCI) [139, 28].

Linear mixed-effects models (LMEs) provide a flexible and powerful statistical framework for analyzing longitudinal biomarkers that enable characterization of between-patient variation in within-patient AD progression. LMEs have been widely implemented in prior AD research, including applications to magnetic resonance imaging (MRI) data [16], positron emission tomography (PET) [96], and clinical and neuropsychological assessments [178]. However, disease progression may be better captured by evolution of multiple biomarkers simultaneously [83, 82]. Additionally, most longitudinal studies of AD are of much shorter duration than the length of time involved in progression of the illness [35]. Analytic approaches are thus needed to account for multivariate biomarker dynamics (and their heterogeneity across subjects) using available short-term longitudinal data.

Growth mixture models (GMM) [131], which are an extension of mixed-effects models that incorporate latent classes for random effects distributions, can be usefully applied to address these issues. GMMs have been implemented in multiple longitudinal studies related with Alzheimer's disease. For example, Pietrzak et al. [141] applied GMMs to reveal three predominant trajectories of a composite score of episodic memory change. Lin et al. [101] fit two separate GMMs to examine the potentially heterogeneous longitudinal trajectories of episodic memory and executive function for identifying the existence of successful cognitive agers. Leoutsakos et al. [100] implemented parallel-process growth mixture models on both cognitive and functional measures and a follow-up multinomial logistic regression to predict class membership. These applications either focused on univariate modeling or *post hoc* analyses of the effects of multiple measures based on GMM results. Lai et al. [94] constructed a multivariate finite mixture latent trajectory model, which can identify subgroups of patients. However, the specification of the model requires constraints on covariance matrices; moroever, the expectation–maximization (EM) algorithm was applied for parameter estimation, for which the performance often depends on the choice of starting values.

To address these limitations, we propose a Bayesian multivariate growth mixture model (BMGMM) incorporating latent states for prediction of AD progression. Our proposed methodology has several advantages compared with previous work. First, our model enables statistical inference for modeling multivariate longitudinal growth trajectories and simultaneously predicts latent classes, incorporating random effects for both latent classes and outcomes. Second, the proposed model accommodates covariates for both outcome trajectory modeling and latent class modeling. Third, the proposed model estimates probability of latent class membership predicting future transitions in disease progression. Finally, as a fully Bayesian hierarchical model, the BMGMM incorporates data from longitudinal trajectories of multiple outcomes in a unified modeling framework, allowing flexible and rigorous interrogation of the posterior distribution, model selection and checking, and inference about long-term disease dynamics from short-term multivariate longitudinal biomarker data.

3.2 Case Study: Alzheimer's Disease Neuroimaging Initiative (ADNI) Study

The proposed approach was applied to the Alzheimer's Disease Neuroimaging Initiative (ADNI) data. ADNI is a multicohort longitudinal study started in 2004, tracking the progression of AD with clinical, imaging, genetic and biospecimen biomarkers. Over 1900 volunteers between the ages of 55 and 90 have been recruited in four waves or phases of study. The first phase, referred as ADNI-1, consists of 800 individuals: 200 cognitvely normal (CN), 400 mild cognitive impairment (MCI), and 200 with mild dementia. Following phases enrolled additional individuals at different stages as well as keeping participants from the prior cohorts. More information about the details of the study can be found at http://adni.loni.usc.edu/.

The classification of stages of disease progression in ADNI was given as follows [140]. CN individuals have a Clinical Dementia Rating (CDR) score of 0, Mini Mental State Exam (MMSE) between 24-30, no memory complaints, and paragraph recall scores meeting education adjusted cutoffs for CN. In addition, CN individuals could not have any significant impairment in cognitive functions or activities of daily living. The Significant Memory Concern (SMC) cohort was added to the second phase of the study, i.e. ADNI-2. Individuals identified as in the stage of SMC have a self-report significant memory concern, quantified by using the Cognitive Change Index. MCI participants have a CDR of 0.5, MMSE between 24-30, a subjective memory complaint and could not qualify for the diagnosis of dementia, and paragraph recall scores meeting education adjusted cutoffs for MCI. Dementia participants had CDR of 0.5 or 1, MMSE of 20 to 26, memory complaints and meet criteria for probable AD. At the screening visit, all subjects were required to provide demographics, family history, medical history and physical examinations and neurological examinations were given to record crucial signs.

3.3 Methods

Here, we present our Bayesian multivariate growth mixture model (BMGMM) for analyzing multiple biomarkers from multivariate longitudinal data. Section 3.3.1 presents the BMGMM model formulation. Section 3.3.2 and Appendix A.1 describe our model inference and our efficient Gibbs sampling algorithm. Sections 3.3.2, 3.3.2 and 3.3.2 describe prior specification, derivation of conditional posteriors, and model selection.

3.3.1 Model

Let y_{ijl} denote the response of biomarker l, $l = 1,2,...,L$, for subject i, $i = 1,2,...,n$ at jth time point t_{ijl}, $j = 1,...,m_{il}$. We assume the distribution of y_{ijl} is characterized by a mixed-effects model,

$$y_{ijl} = \boldsymbol{x}_{ijl}^T \boldsymbol{\beta}_l + a_{0il} + a_{1il} t_{ijl} + \varepsilon_{ijl},\ \ i = 1,\dots n,\ j = 1,\dots,m_{il},\ l = 1,\dots,L, \tag{3.1}$$

where $\boldsymbol{x}_{ijl}$ is Q_l-dimensional vector of fixed effects (which may include time-varying covariates); $\boldsymbol{\beta}_l$ is the corresponding Q_l-dimensional vector of regression coefficients, and $\varepsilon_{ijl} \overset{\text{i.i.d.}}{\sim} N(0, \sigma_\varepsilon^2)$. The parameters a_{0il} and a_{1il} are biomarker-specific subject random intercepts and slopes, respectively. We assume that the joint distribution of these random effects depends on subject membership in one of K latent subgroups in the data, capturing underlying differences in disease states not necessarily captured by clinical diagnostic assessments.

Let $C_i \in \{1,\dots,K\}$ be a random variable denoting the (unknown) latent class membership of the ith subject. Further, let $\boldsymbol{a}_i = (a_{0i1}, a_{1i1}, ..., a_{0iL}, a_{1iL})^T$ denote the $2L$-dimensional vector of random coefficients for the ith subject. We assume $\boldsymbol{a_i}$ follows a multivariate normal distribution conditional on C_i, so that $\boldsymbol{a}_i|C_i = k \sim N(\boldsymbol{\alpha}_k, \boldsymbol{\Sigma}_k)$. Here, $\boldsymbol{\alpha}_k = (\alpha_{0k1}, \alpha_{1k1}, ..., \alpha_{0kL}, \alpha_{1kL})^T$ are class-specific random effect means and $\boldsymbol{\Sigma}_k$ are class-specific $2L \times 2L$ random effect covariance matrices, $k = 1, \dots, K$.

Subject-specific probabilities of latent class memberships $\boldsymbol{\pi}_i(\boldsymbol{z}_i) = (\pi_{i1}(\boldsymbol{z}_i), ..., \pi_{iK}(\boldsymbol{z}_i))^T$ are modeled using multinomial logistic regression:

$$P(C_i = k|\boldsymbol{z}_i) = \pi_{ik}(\boldsymbol{z}_i) = \frac{\exp(\boldsymbol{z}_i^T \boldsymbol{\gamma}_k)}{\sum_{k'=1}^K \exp(\boldsymbol{z}_i^T \boldsymbol{\gamma}_{k'}))}, \tag{3.2}$$

where $\boldsymbol{z}_i$ is a Q_z-dimensional vector of time-invariant covariates related with latent class and $\boldsymbol{\gamma}_k$ is the corresponding class-specific coefficient vector. To ensure identifiability of the model, $\boldsymbol{\gamma}_K \equiv \mathbf{0}$ and hence the last class K is the reference category for the multinomial logistic regression. Therefore, by incorporating the latent class setting, we implement a mixture model of the form

$$f(y_{ijl}|\boldsymbol{x}_{ijl}, t_{ijl}, \boldsymbol{z}_i) = \sum_{k=1}^K P(C_i = k|\boldsymbol{z}_i) f(y_{ijl}|\boldsymbol{x}_{ijl}, t_{ijl}, C_i = k), \tag{3.3}$$

where $f(y_{ijl}|\boldsymbol{x}_{ijl}, t_{ijl}, C_i = k)$ denotes the mixed-effects model format in Equation 3.1, conditional on $C_i = k$, and $P(C_i = k|\boldsymbol{z}_i)$ denotes the multinomial logistic model in Equation 3.2.

3.3.2 Model Inference

In this section we describe the Bayesian prior specification, model fitting algorithm, and model selection metrics. A detailed description of the Gibbs sampling algorithm is given in Appendix A.1.

Prior Specification

Here, we describe the prior specification for the BMGMM. We adopt weakly informative prior distributions to all model parameters. For $\boldsymbol{\beta}_l, l = 1, \dots, L$ and $\boldsymbol{\alpha}_k, k = 1, \dots, K$, independent conjugate normal distributions $N_{Q_x}(0, c\mathbf{I})$ and $N_{2L}(0, c\mathbf{I})$ are assigned, respectively. In our implementation, we set $c = 100$. We assume a conjugate *Inverse-Gamma*(δ_1, δ_2) prior for σ_ε^2 and specify $\delta_1 = \delta_2 = 1$.

For class-specific coefficients $\boldsymbol{\gamma}_k, k = 1, \ldots, K-1$, we adopt Polya-Gamma data augmentation approach, to avoid the the need for complicated approximation or numerical integration[143]. Each $\boldsymbol{\gamma}_k$ is assigned a prior under the Pólya–Gamma sampling scheme. A Detailed description of this algorithm, with corresponding priors, can be found in A.1.1.

For the class-specific covariance matrix $\boldsymbol{\Sigma}_k$, the prior is specified as *Inverse-Wishart*$(\nu + 2L - 1, 2\nu\Delta)$, where Δ is a diagonal matrix with elements λ_l, which are assumed to be independently distributed with *Gamma*$(\frac{1}{2}, \frac{1}{\psi_l^2})$. This prior is referred as *Half-t*$(\nu, \boldsymbol{\psi})$ for it generates a Half-t distribution with ν degrees of freedom and scale parameter ψ_l for standard deviations. This prior is implemented to reduce potential impact of misestimation of the correlation coefficients [75], considering the complex covariance matrices designed for multivariate biomarkers. We specify $\nu = 2$ to obtain a uniform prior for the correlation coefficients [55] and $\phi_l = 1$.

Posterior Computation

The prior specification above provides full conditional distributions for all model parameters which can be efficiently updated via a Gibbs sampler. Let $\boldsymbol{C} = (C_1, \ldots, C_n)^T$ denote the vector of latent class memberships and let $\boldsymbol{a} = (\boldsymbol{a}_1^T, \ldots, \boldsymbol{a}_n^T)^T$ denote the stacked vector of random effects for all n subjects. Assuming prior independence of the model parameters, the joint posterior is given by

$$\begin{aligned} P(\boldsymbol{C}, \boldsymbol{a}, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_L, \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_K, \boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_K, \boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_{K-1}, \sigma_\varepsilon^2 | \mathbf{y}) \propto \\ \prod_{k=1}^{K} \prod_{i=1}^{n} \{\pi_{ik}(\boldsymbol{z}_i) [\prod_{l=1}^{L} \prod_{j=1}^{m_{il}} N(y_{ijl} | \xi_{ijl}, \sigma_\varepsilon^2)] N_{2L}(\boldsymbol{a}_i | \boldsymbol{\alpha}_k, \boldsymbol{\Sigma}_k)\}^{\mathbf{1}(C_i = k)} \\ \cdot \pi(\boldsymbol{\beta}_l) \pi(\sigma_\varepsilon^2) \pi(\boldsymbol{\Sigma}_k) \pi(\boldsymbol{\alpha}_k) \pi(\boldsymbol{\gamma}_k), \end{aligned} \tag{3.4}$$

where $\xi_{ijl} = \boldsymbol{x}_{ijl} \boldsymbol{\beta}_l + a_{0il} + a_{1il} t_{ijl}$ and $N(y_{ijl} | \xi_{ijl}, \sigma_\varepsilon^2)$ denotes the normal distribution for y_{ijl} as described in Equation 3.1; $N_{2L}(\boldsymbol{a}_i | \boldsymbol{\alpha}_k, \boldsymbol{\Sigma}_k)$ denotes the normal likelihood for random effects $\boldsymbol{a}_i$

with mean $\boldsymbol{\alpha}_k$ and covariance matrix $\boldsymbol{\Sigma}_k$; $\mathbf{1}(C_i = k)$ is an indicator if subject i belongs to class k; $\pi(\cdot) = \{\pi(\boldsymbol{\beta}_l), \pi(\sigma_\varepsilon^2), \pi(\boldsymbol{\Sigma}_k), \pi(\boldsymbol{\alpha}_k), \pi(\boldsymbol{\gamma}_k)\}$ denote the prior distributions for corresponding parameters as described above. At each iteration s, the sampling scheme consists of the following steps:

1. Sample $\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_{K-1}$ by incorporating Polya-Gamma auxiliary variable as desribed in A.1.1.

2. Sample the class indicators $C_i (i = 1, \dots, n)$ from a discrete categorical distribution with probability $\boldsymbol{\pi}_i = (\pi_{i1}, \dots, \pi_{iK})$ described in A.1.2.

3. For $k = 1, \dots, K$, sample the class-specific parameters $\boldsymbol{\alpha_k}$ and $\boldsymbol{\Sigma}_k$ from their full conditionals.

4. Given $C_i = k$, sample $\boldsymbol{a}_i$ from their full conditionals.

5. Sample σ_ε^2 and sample $\boldsymbol{\beta}_l, l = 1, \dots, L$ from their full conditionals.

The detailed Gibbs sampling algorithm can be found in Appendix A.1. A corresponding R script is provided on Github (https://github.com/wendylin23/BMGMM).

Model comparison

For model comparisons, we implement the Widely Applicable Information Criterion (WAIC) [184]. The WAIC is computed using the log-likelihood evaluated at the posterior draws of the parameter values. Let $\boldsymbol{\theta}$ denote all parameters in the model. The log-likelihood for individual observation vector $\boldsymbol{y}_i = (y_{i11}, \dots, y_{im_{i1}1}, \dots, y_{i1L}, \dots, y_{im_{iL}L})$ is given by

$$\begin{aligned} l(\boldsymbol{y}_i|\boldsymbol{\theta}) = \sum_{k=1}^{K} \mathbf{1}(C_i = k)\{\sum_{l=1}^{L} [\log(\mathbf{I}_{m_{il}}\sigma_\varepsilon^2)^{-\frac{1}{2}} - \frac{1}{2}(\boldsymbol{y}_{il} - \boldsymbol{\xi}_{il})^T (\mathbf{I}_{m_{il}}\sigma_\varepsilon^2)^{-1}(\boldsymbol{y}_{il} - \boldsymbol{\xi}_{il})] - \\ \frac{1}{2}\log|\boldsymbol{\Sigma}_k| - \frac{1}{2}(\boldsymbol{a}_i - \boldsymbol{\alpha}_k)^T \boldsymbol{\Sigma}_k^{-1}(\boldsymbol{a}_i - \boldsymbol{\alpha}_k)\}, \end{aligned} \tag{3.5}$$

where $\boldsymbol{y}_{il} = (y_{i1l}, \dots, y_{im_{il}l})$ and $\boldsymbol{\xi}_{il} = (\xi_{i1l}, \dots, \xi_{im_{il}l})$. Then the WAIC is defined as[156],

$$WAIC = -2\sum_{i=1}^{n} \log\{\sum_{s=1}^{S} \exp(l(\boldsymbol{y}_i|\boldsymbol{\theta}^{(s)}))\} + 2\sum_{i=1}^{n} \mathtt{Var}_{s=1}^{S} l(\boldsymbol{y}_i|\boldsymbol{\theta}^{(s)}), \tag{3.6}$$

where S is the number of MCMC iterations and $\boldsymbol{\theta}^{(s)}$ is the sth draw from the posterior distribution. Lower values of WAIC indicates better model fit. R code providing an efficient computation of WAIC for our BMGMM can be found at https://github.com/wendylin23/BMGMM.

3.4 Simulation study

In these simulation studies, we conducted Monte Carlo experiments to examine the performance of the proposed algorithm and how model misspecification affects performance. For each example, we simulated $n = 200$ individuals and $t = 6$ time points. We kept only 75% of the generated data to create a sparser datasets to better mimic the data in ADNI study. For each individual, the observation times were sampled from a uniform distribution $t_{ijl} \sim \text{Unif}(0,5)$ and the baseline time was sampled from a normal distribution $T_{i0} \sim N(0,5)$ for indicating varying starting time points. Setting $\sigma_\varepsilon = 0.1$ for all simulation scenarios. For each model fitting, we ran two parallel Markov chains and each chain was run 2000 iterations, with the first 1000 iterations discarded, yielding a total of 2000 samples for posterior analysis. Each model was simulated $M = 100$ times.

3.4.1 Model fitting with varying numbers of outcomes and latent states

In the first scenario, we simulated examples from $P = 1,2,3$ outcomes and $K = 2,3$ latent states. Let $\hat{\theta}_s$ be the posterior estimate of a model parameter θ_s in the s-th simulation. The following quantities were considered for assessing the model performance: (i) average bias, Bias $= \frac{1}{S}\sum_{s=1}^{S}(\hat{\theta}_s - \theta_s)$, (ii) total mean squared error, MSE $= \frac{1}{S}\sum_{s=1}^{S}(\hat{\theta}_s - \theta_s)^2$ and (iii) the coverage rate of the 95% credible intervals, $C_{95\%}$. Prediction accuracy of latent classes, p_{acc} for

Table 3.1. Simulation study results of $P = 2$ and $K = 3$. The model was fit to all 200 samples.

Parameter (true values)	**Bias**	**MSE**	**C$_{95}$**	**Bias**	**MSE**	**C$_{95}$**
		$l = 1$			$l = 2$	
β_{l1} (1,3)	0.0104	0.0030	0.92	0.0062	0.0025	0.94
β_{l2} (0.5,-0.1)	0.0018	0.0003	0.86	0.0023	0.0003	0.88
α_{01l} (0.2,0.5)	-0.0189	0.0173	0.96	-0.0236	0.0270	0.96
α_{02l} (2,1)	-0.0041	0.0291	0.90	0.0185	0.0227	0.96
α_{03l} (10,2)	-0.0345	0.0582	0.90	-0.0066	0.0651	0.94
α_{11l} (3,1)	-0.0119	0.0070	0.93	-0.0100	0.0072	0.96
α_{12l} (1,2)	0.0007	0.0033	0.98	0.0050	0.0065	0.93
α_{13l} (-0.5,-0.5)	-0.0085	0.0969	0.97	0.0049	0.0418	0.98
		Covariate 1			**Covariate 2**	
$\boldsymbol{\gamma}_1$ (1,0.5)	0.1180	0.1453	0.90	-0.0227	0.0973	0.94
$\boldsymbol{\gamma}_2$ (2,-0.5)	0.1240	0.1227	0.91	-0.0631	0.0901	0.93
p_{acc}	0.987					

* $K = 3$ is the reference class with $\boldsymbol{\gamma}_3 = 0$.

200 samples, is also included in the table.

Table 3.1 presents the average bias, MSE and C_{95} from the simulation setting with $P = 2$ outcomes and $K = 3$ latent classes. Simulation parameters were set to fixed effects coefficients $\boldsymbol{\beta}_1 = (1,3)'$, $\boldsymbol{\beta}_2 = (0.5,-0.1)'$ and class-specific coefficients $\boldsymbol{\gamma}_1 = (1,0.5)'$, $\boldsymbol{\gamma}_2 = (2,-0.5)'$. Random effects are to with intercepts $\boldsymbol{\alpha}_{01} = (0.2,0.5)'$, $\boldsymbol{\alpha}_{02} = (2,1)'$, $\boldsymbol{\alpha}_{03} = (10,2)'$ and slopes $\boldsymbol{\alpha}_{11} = (3,1)'$, $\boldsymbol{\alpha}_{12} = (1,2)'$, $\boldsymbol{\alpha}_{13} = (-0.5,-0.5)'$, where the random intercepts in the first outcome are more separable compared with the second outcome. For each latent class k, the covariance matrix of random effects are generated from Wishart distribution with $2P$ degrees of freedom. Figure 3.1 plots the coverage rate of the 95% credible intervals of each element in the covariance matrix. Both summary table and plot illustrate that parameters are estimated accurately and coverage of posterior credible intervals is close to the nominal 95%. Other simulation results from varying combinations of number of latent classes and outcomes, included in Appendix A.2, show similarly good performance.

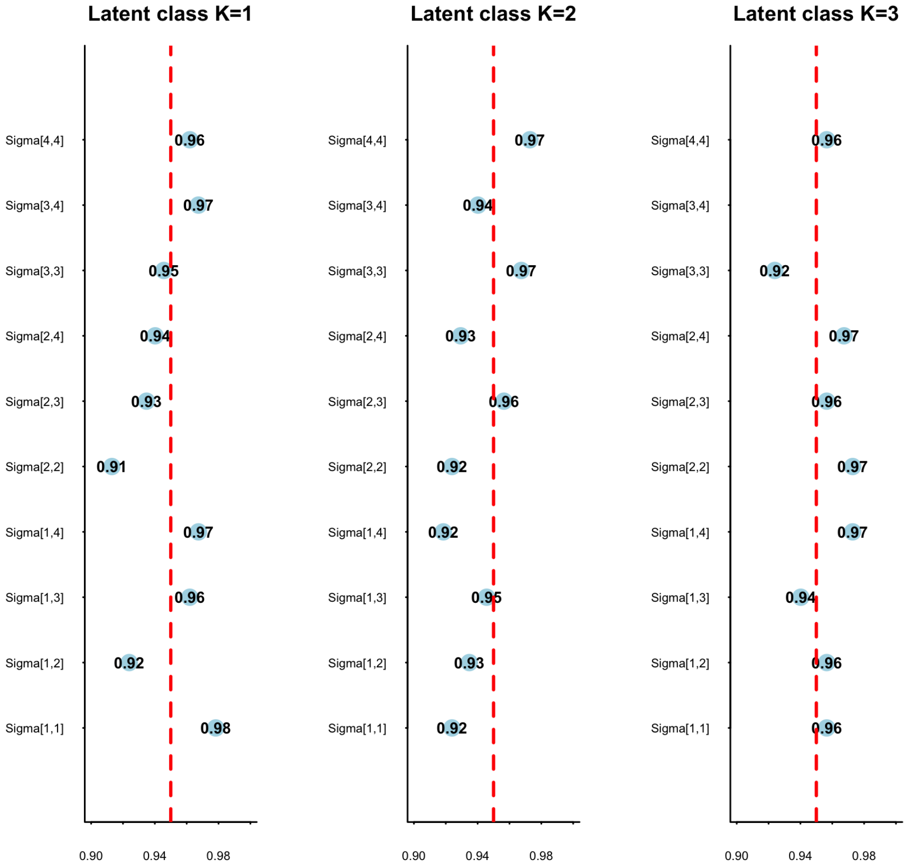


Figure 3.1. Coverage rate of the 95% credible intervals for estimated random effects covariance matrix for three latent classes.

3.4.2 Model misspecification performance

In the second simulation study, we explored the model performance when it is misspecified. The true model was simulated with $P = 2$ outcomes and $K = 3$ latent classes, using the same parameter setting as in section 3.4.1. The covariance matrix was assumed to be varied between classes. In addition to fitting the model with the true setting, we included five misspecification scenarios, (i) assuming $K = 3$ but fitting a univariate model for each outcome (ii) assuming $K = 1$ and fitting multivariate linear mixed models on both outcomes (iii) assuming $K = 2$ latent classes (iv) assuming $K = 4$ latent classes and (v) assuming $K = 3$ but fitting model with homogeneous convariance matrices for all latent classes. Estimated WAICs are computed by averaging over 100 simulation samples and compared between the true model and each misspecified model.

Specifically, for two univariate models, we computed a combined WAIC by assuming an identity covariance matrix between two outcomes. The simulation results are presented in Table 3.2, shows that the model specified with $P = 2$ outcomes, $K = 3$ or $K = 4$ latent classes and heterogeneous covariance matrices is preferred with lower values of WAIC's. This is expected since $K = 3$ conforms with the original model setting and $K = 4$ increases the complexity of model, thus improves the model performance. However, among all 100 simulation replicates specified with $K = 4$ latent classes, 78 of them are predicted with 3 latent classes, indicating the stability of our proposed algorithm even with misspecified parameters.

Table 3.2. Simulation study results from multiple misspecified scenarios and the true model, including case (i): fitting a univariate model for each outcome, case (ii): assuming $K = 1$ and $P = 2$, case (iii): assuming $K = 2$ and $P = 2$, case (iv): assuming $K = 4$ and $P = 2$, and case (v): assuming equal Σ_k for all k's. WAIC is an averaged value from 100 simulation replicates, along with its corresponding standard deviation.

Scenario	**WAIC (SD)**
True model	2930.32(151.36)
Case (i)	3626.53 (299.09)
Case (ii)	3900.75 (103.86)
Case (iii)	3302.77 (193.95)
Case (iv)	2914.95(128.30)
Case (v)	3421.25 (111.41)

3.5 Real Data Analysis

3.5.1 ADNI Data

We used the BMGMM on the ADNI data, focusing on the Alzheimer's Disease Assessment Scale Cognitive Subscale (ADAS-Cog) and structural Magnetic Resonance Imaging (MRI) volumes. The ADAS-Cog [155] is a cognitive assessment, with higher scores indicating more severe cognitive dysfunction, and is frequently used in clinical trials in populations with dementia. In addition, MRI volumes from multiple regions of interest (ROIs), including hippocampus, middle temporal lobe (Mid-Temp), fusiform gyrus, and entorhinal cortex [106, 84], to track the pathophysiology of AD progression. Trajectories for each outcome are displayed in Figure 3.2, with diagnostic status (CN, MCI, AD) indicated by color (blue, yellow, red, respectively). It can be seen from this figure that there is some clustering of trajectories by diagnostic status, most pronounced in ADAS-Cog and least pronounced in fusiform volume.

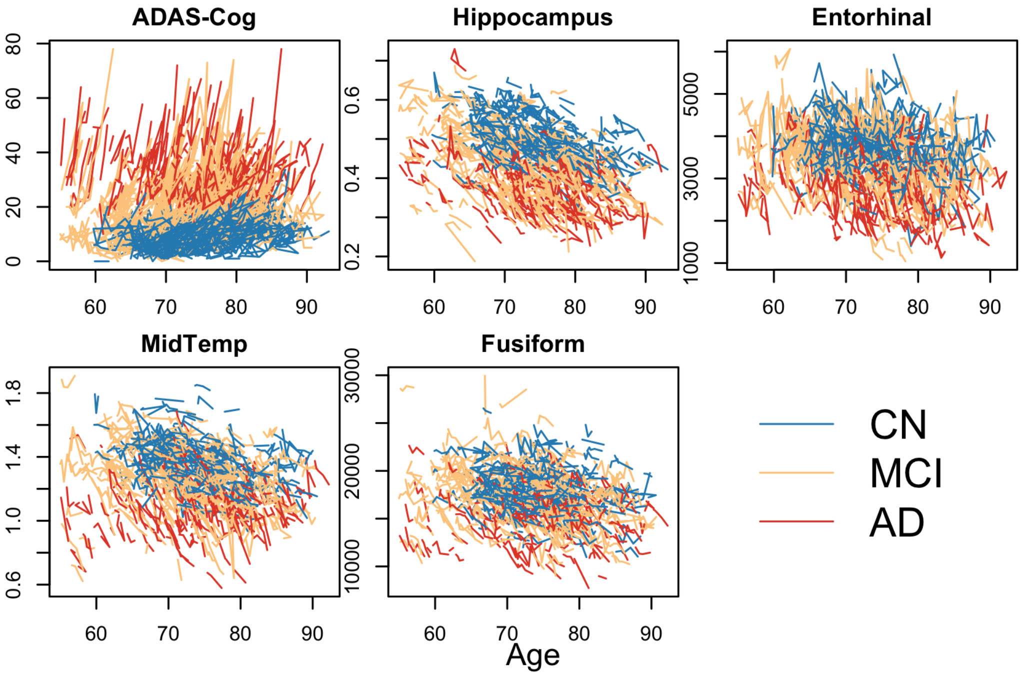


Figure 3.2. Spaghetti plots of the observed biomarker values of ADAS-Cog, hippocampus volume, middle temporal lobe volume, fusiform volume, and entorhinal thickness from 745 participants in the Alzheimer's Disease Neuroimaging Initiative study with respect to their age over time. Colors indicate diagnostic stage at entry, including cognitively normal (blue), mild cognitive impaired (yellow) and dementia (red).

3.5.2 Data analysis

We fitted the proposed BMGMM on $n = 745$ ADNI participants to determine if there are latent classes of disease progression. Note, clinical diagnosis was not included as a predictor in the BMGMM as one goal of this analysis was to examine the degree to which predicted latent class memberships from biomarker and cognitive data mapped on to clinical assessments. Thus, we evaluated the concordance of our predicted latent classes based on these $P = 5$ outcome trajectories with baseline and with follow-up diagnoses. In addition, we compared the predicted latent classes with the progression of the Clinical Dementia Rating Sum of Boxes (CDRSB) [77, 125], which is a global assessment tool for both cognitive and functional impairment. It is used widely as a single primary endpoint for trials studying individuals at earlier stages of AD, when the use of ADAS-Cog is more limited[185]. The comparison between our predicted latent classes and the progression of CDRSB can further imply the role of MRI measures plays in the model fitting.

Since outcome measures have varying scales (as showed in Fig 3.2), they were normalized to be within range 0 and 1 by subtracting the minimum value and divided by the difference between maximum and minimum values, to ensure the stability of model fitting. Presence of apolipoprotein E (APOE) $\varepsilon 4$ allele is a strong genetic risk factor for the development of AD [142] and was used as the fixed effect covariate. High elevated amyloid existence is a binary indicator, defined on high amyloid PET standardized uptake value ratio (SUVR) 11 or low cerebrospinal fluid β-amyloid peptide (CSF Aβ_{42}), and is believed to be significantly associated with worse cognitive measures [36, 62]. Thus we included it as a covariate for class parameters. We ran two parallel Markov chains for 3000 iterations and the first 1500 warm-up iterations were discarded. We compared the WAIC's from models with $K = 1,2,3,4$ latent classes and $K = 3$ was found to be the optimal one. Detailed results of model selection were included in Appendix A.3. We constrained the order of the random intercepts of ADAS-Cog scores to be increasing from latent class 1 to latent class 3, to fix the orders of predicted latent classes. The posterior mean and the 95% credible intervals were computed using the obtained samples for all parameters.

Figure 3.3 shows both population- and subject-level fits of longitudinal trajectories, colored with predicted latent states (left) and baseline diagnosis in ADNI (right). The predicted latent classes give a clearer clustering of trajectories compared with baseline diagnosis, especially for the MRI measures. The posterior means (95% credible intervals) for model parameters are given in Table 3.3. For ADAS-Cog and MidTemp volumes, parameters of both class-specific random intercepts and slopes illustrate a significant separation between the three predicted latent states. For the hippocampal volume and entorhinal thickness, the credible intervals of the estimated random intercepts and slopes between class 2 and class 3 are more overlapping. Fusiform volumes show a similar overlapping pattern in intercepts but more separation in slopes in the three latent classes.

Figure 3.4 displays the predicted individual-level progression intercepts and slopes for ADAS-Cog and four MRI measures, colored by latent classes and symbolized with baseline diagnostic states. These figures support our assumptions of the non-constant convariance matrices among varying latent classes and concur with the results of class-specific parameters in Table 3.3, i.e. the three latent classes are better separated in ADAS-cog and MidTemp volumes. The predicted latent classes display varying patterns of AD development regarding the baseline measurement values and progressive slopes, indicating potentially different rates of AD development in the future within each diagnostic disease state.

The top of Table 3.4 provides the classification summary between the baseline diagnoses and predicted latent classes. The posterior proportions indicate that almost all patients diagnosed with CN at baseline are classified as L1 (86%) while those with dementia at baseline were mostly classified as either L2 (55%) or L3 (43.2%). MCI patients present comparable classification proportions for each latent class. The bottom of Table 3.4, shows that nearly all MCI patients who are predicted with L1 (96.3%) stay in MCI in follow-up diagnosis while for those in L2 and L3, they are more likely to develop dementia. This illustrates that sub-groups within MCI populations can be identified with the predicted latent classes with respect to varying rates of disease progression.

Figure 3.5 further explores the varying patterns of disease progression using the CDRSB as a clinically meaningful marker. The slopes and intercepts were calculated with univariate linear mixed regression models without considering latent states, colored with corresponding predicted latent states and differently shaped to reflect status of changing diagnostics in the future. For MCI patients, almost all patients classified as L1 did not progress to dementia while those who were classified as L2 or L3 were more likely to have progressed. Furthermore, MCI subjects with higher intercepts or slopes in L2 and L3 are more likely to develop to AD dementia.

Table 3.5 concludes the estimated random effect parameters of ADAS-Cog, entorhinal volumes and CDRSB within different disease stages and changing diagnostic states, which conforms with our findings in Figure 3.5, i.e. both intercepts and slopes can inform future direction of AD progression.

3.6 Discussion

We developed a Bayesian multivariate growth mixture model for multivariate longitudinal data. The model is appealing in its ability to incorporate several novel hierarchical approaches, including Half-t priors for standard deviations and Pólya–Gamma data augmentation for class-specific coefficients in multinomial regression. Compared with previous works with growth mixture models, these improvements promote the stability of model fitting with multiple outcomes and latent classes simultaneously in a fully Bayesian hierarchical model, which is efficient and flexible.

The advantages of the proposed model have been supported with our simulation results. We showed that our model can accommodate varying combinations of latent classes and outcomes. For comparison, we also implemented the simulation with commonly-used Inverse-Wishart priors for covariance matrix, which failed to converge in 50% simulation samples. In addition, another binomial and multinomial sampling approach introduced by Holmes and Held (2006) [72], the Bayesian auxiliary variable algorithm, was examined and compared with our algorithm. The Pólya–Gamma algorithm outperforms it both in efficiency and prediction accuracy, especially for multinomial cases. Furthermore, the second part of simulation experiment demonstrated that the misspecification of the model can significantly impair the model performance, validated with WAIC.

We applied our method to ADNI data to characterize the trajectories of ADAS-Cog and four MRI measures simultaneously and predict the latent classes of individuals. We identified three classes and compared them with the baseline diagnosis. We found that CN participants are more likely to be included in latent class 1 and 2 while participants in MCI and dementia tend to be in latent class 2 and 3. In addition, CN and MCI participants have more meaningful variability with respected to predicted states. From milder latent classes (L1/L2) to more severe classes (L2/L3), potential sub-classes within mild-to-moderate baseline diagnosis can be identified. Furthermore, we showed that the predicted latent classes indicate the future disease progression direction for individuals with different baseline diagnoses, which is helpful in understanding the heterogeneity of disease progression, and designing future clinical trials.

Further works are required to extend current study. Firstly, more types of biomarker outcomes may be incorporated and compared for finding optimal combinations in predicting the latent disease classes and future disease progression. Additionally, instead of using the linear model and normality assumption, other methods can be implemented to capture more variability within the data, such as splines and skew-normal distributions. Another potential direction is to include functional data analysis for modeling the random effects, which is flexible and able to capture more complex associations between random intercepts and slopes.

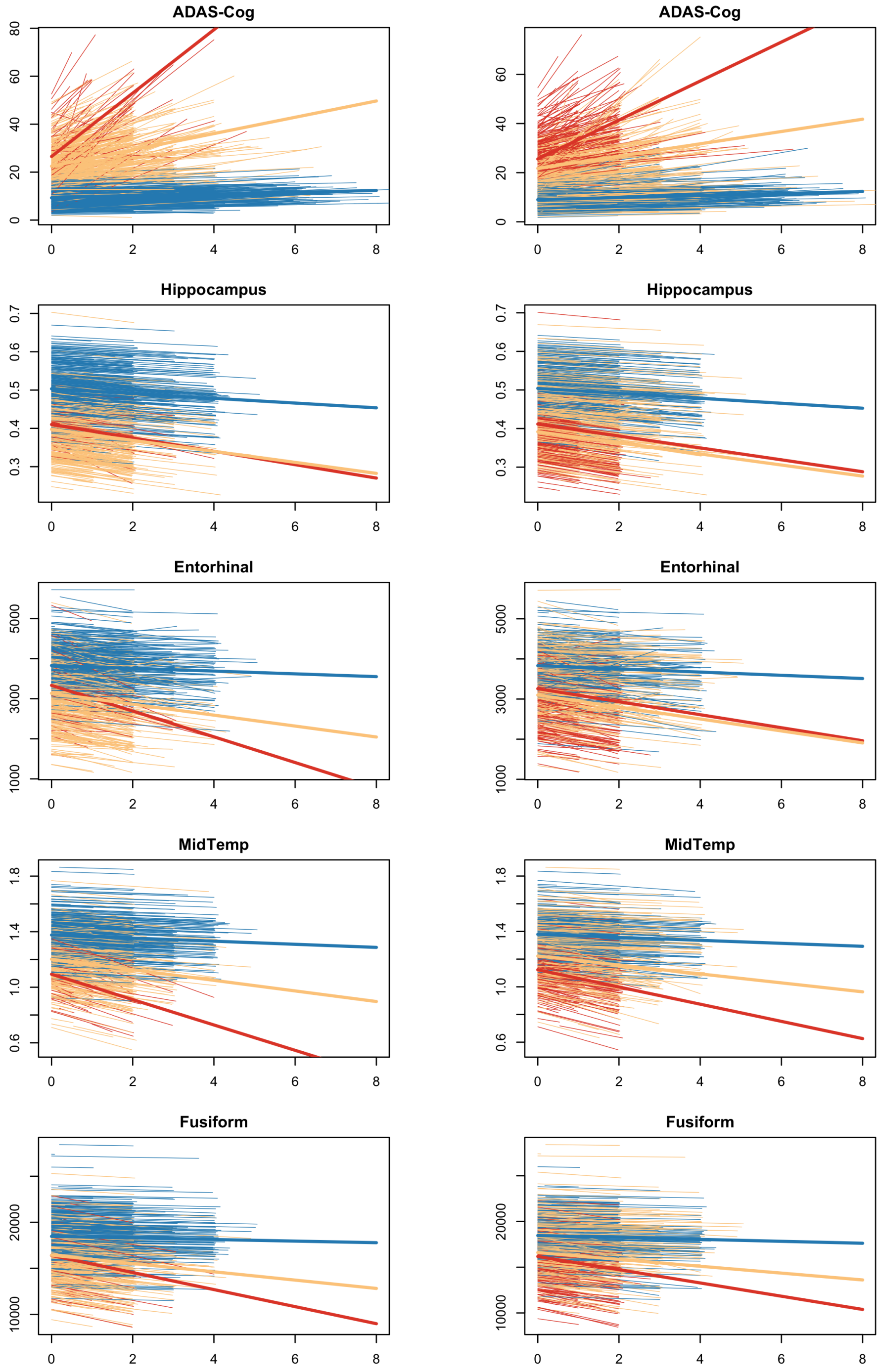


Figure 3.3. The modeled population and individual trajectories of ADAS-Cog, hippocampus, middle temporal lobe, fusiform, and entorhinal volumes (top to bottom). The colors indicate predicted latent classes (left panels) or diagnostic stages (right panels), including cognitively normal or latent class 1 (blue), mild cognitive impaired or latent class 2 (yellow) and dementia or latent class 3 (red).

Table 3.3. Results of posterior estimates of parameters for the proposed BMGMM fit to ADAS-Cog and four MRI measures. Colors indicate the severity of latent classes of the corresponding parameters (blue:1, yellow:2, red:3).

Parameter	Posterior Mean	95% Credible Interval	Parameter	Posterior Mean	95% Credible Interval
ADAS-Cog			**Hippocampus**		
β_1	0.008	(-0.004, 0.017)	β_1	-0.001	(-0.018, 0.017)
α_{01}	0.112	(0.104, 0.119)	α_{01}	0.584	(0.569, 0.600)
α_{11}	0.005	(0.004, 0.007)	α_{11}	-0.012	(-0.013, -0.011)
α_{02}	0.275	(0.259, 0.292)	α_{02}	0.377	(0.354, 0.398)
α_{12}	0.032	(0.027, 0.037)	α_{12}	-0.026	(-0.030,-0.024)
α_{03}	0.324	(0.298, 0.351)	α_{03}	0.413	(0.388, 0.439)
α_{13}	0.102	(0.084, 0.121)	α_{13}	-0.028	(-0.033, -0.021)
Entorhinal			**MidTemp**		
β_1	-0.017	(-0.041, 0.007)	β_1	0.005	(-0.016, 0.024)
α_{01}	0.564	(0.549, 0.579)	α_{01}	0.599	(0.586, 0.613)
α_{11}	-0.008	(-0.011, -0.005)	α_{11}	-0.008	(-0.009, -0.006)
α_{02}	0.418	(0.391, 0.444)	α_{02}	0.482	(0.457, 0.506)
α_{12}	-0.030	(-0.034, -0.025)	α_{12}	-0.024	(-0.027, -0.022)
α_{03}	0.450	(0.417, 0.479)	α_{03}	0.407	(0.376, 0.435)
α_{13}	-0.032	(-0.045, -0.020)	α_{13}	-0.028	(-0.054, -0.040)
Fusiform			**Class coefficients**		
β_1	0.0003	(-0.020, 0.020)	γ_1	-2.405	(-2.886,-1.96)
α_{01}	0.483	(0.468, 0.498)	γ_2	1.368	(0.268,2.739)
α_{11}	-0.005	(-0.006, -0.002)			
α_{02}	0.399	(0.373, 0.423)			
α_{12}	-0.017	(-0.020, -0.013)			
α_{03}	0.380	(0.352, 0.408)			
α_{13}	-0.033	(-0.040, -0.026)			

* β_1 is the estimated coefficient for APOE.

** γ_1 and γ_2 are estimated class coefficients for binary variable of elevated amyloid in the first two latent classes, when $K = 3$ is the reference group.

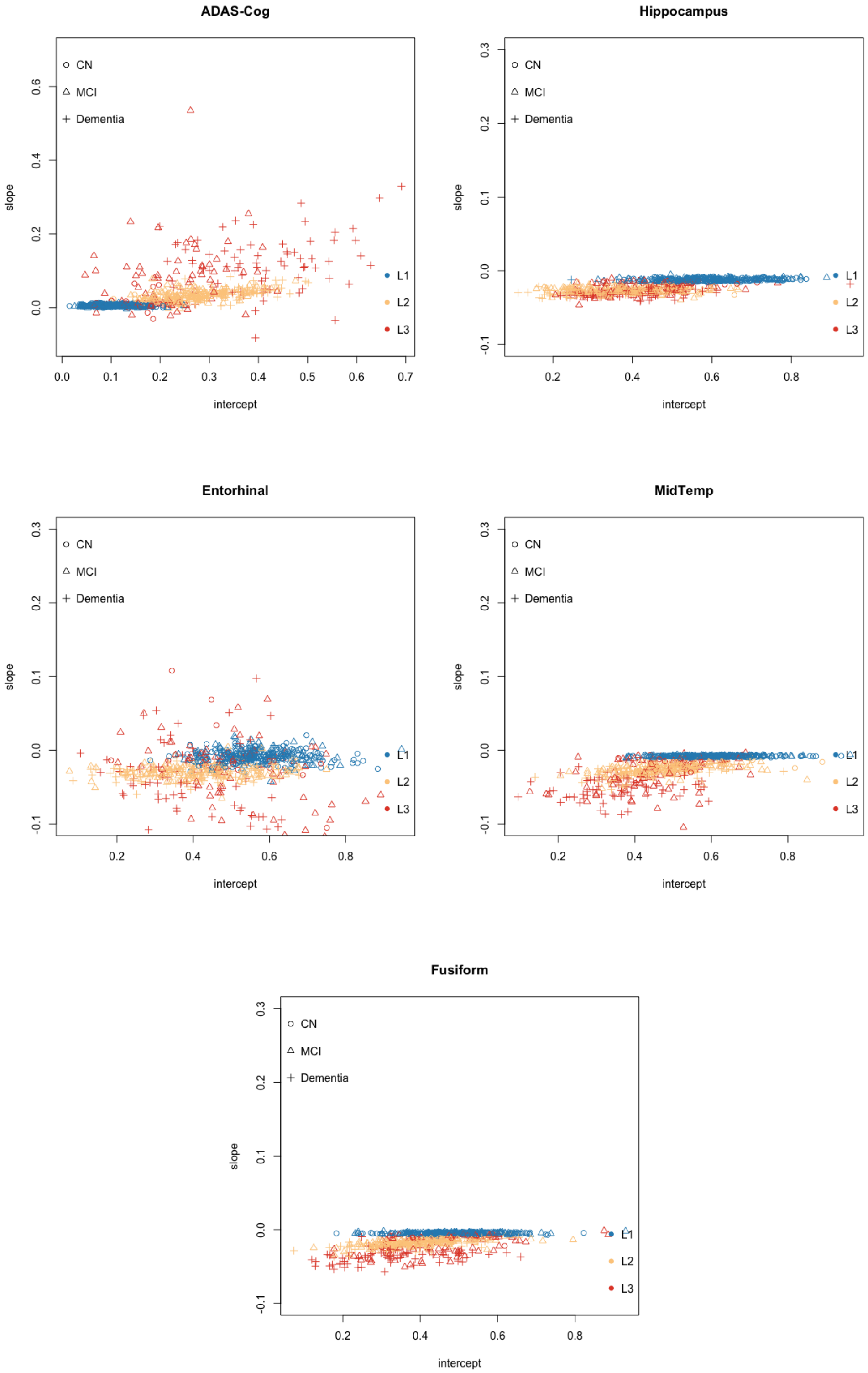


Figure 3.4. The distribution of estimated individual progression intercepts and slopes from multivariate GMM for ADAS-Cog, hippocampus, middle temporal lobe, fusiform, and entorhinal volumes. The dots are colored with predicted latent states and symbolized with diagnostic states.

Table 3.4. Classification summary table for MGMM fitted with ADAS-Cog and four MRI measures. Top table presents the marginal proportions of each baseline diagnostic state being classified as latent state classes. For MCI patients at baseline, the bottom table shows the percentage of future disease development within predicted latent classes.

	Predicted Latent Class		
Baseline Diagnosis	L1	L2	L3
CN	0.860	0.061	0.079
MCI	0.349	0.424	0.550
Dementia	0.018	0.550	0.432

Progression	L1	L2	L3
Stable MCI	0.963	0.388	0.526
MCI to Dementia	0.037	0.612	0.474

Table 3.5. Disease progression parameter summary table for ADAS-Cog, ventricular volumes and CDRSB. 'Change' is a binary variable for indicating whether an individual deteriorate in the follow-up diagnosis, i.e. CN to MCI/AD or MCI to AD. The intercepts and slopes are average estimates for corresponding sub-classes.

	ADAS-Cog			**Entorhinal**			**CDRSB**		
Parameter	**CN**	**MCI**	**Dementia**	**CN**	**MCI**	**Dementia**	**CN**	**MCI**	**Dementia**
Change = 1									
intercept	0.151	0.256	0.337	0.541	0.451	0.418	0.001	0.099	0.143
slope	0.019	0.050	0.088	−0.019	-0.030	-0.040	0.030	0.082	0.064
Change = 0									
intercept	0.104	0.171	0.366	0.559	0.525	0.382	0.003	0.071	0.234
slope	0.006	0.020	0.738	−0.009	-0.016	-0.034	0.002	0.013	0.103

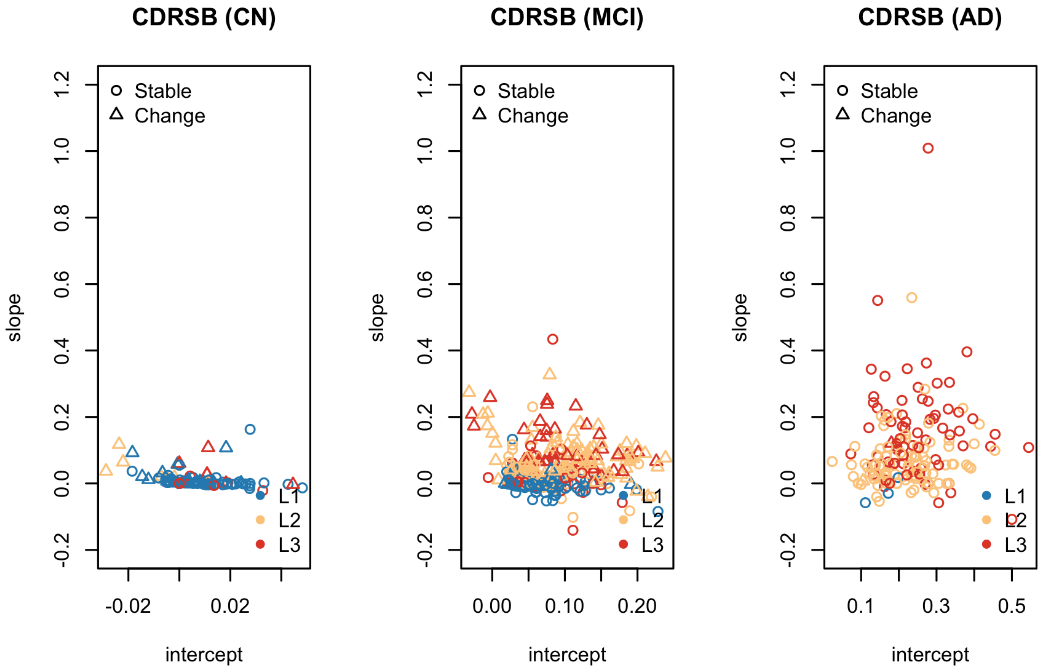


Figure 3.5. The individual-level progression intercept and slope of CDRSB for CN (left), MCI (middle) and dementia (right), colored with predicted latent states. The symbol types represent whether the individual has developed to a severer stage during the study, i.e. ’stable’ means no change and ’change’ means either progression from CN to MCI/Dementia or from MCI to Dementia.

Chapter 4

Longitudinal Associations Between Timing of Physical Activity Accumulation and Health: Application of Functional Data Methods

4.1 Introduction

Physical inactivity and sedentary behavior are known risk factors for cardiovascular disease (CVD), cancer and mortality [108, 18, 9, 137]. Increased physical activity has been demonstrated to improve cardiopulmonary fitness and promote healthy weight management [110]. Current CDC guidelines recommend engaging in 150 minutes/week or more of moderate-vigorous-activity in order to maintain a healthy weight, and for reducing the risk of hypertension, diabetes, heart attacks, and stroke, as well as, osteoporosis, risk of falls, and depression [25]. Given the multitude of health benefits, it is important to develop robust and informative statistical models for exploring the relationship between all aspects of Physical activity (PA) and health outcomes.

Traditional approaches for collecting information about an individual's PA have relied heavily on self-reported questionnaires, sleep-logs, and daily diaries [173]. However, these methods require an individual to recall their PA over a previous period, and hence are often inaccurate and/or biased. Further, these methods do not usually obtain daily PA level or elicit information regarding PA accumulation patterns throughout the day [4]. Because accurate and consistent measurement of PA is critical for designing and assessing interventions, devices such as accelerometers, are increasingly used for recording objective estimates of PA [85]. These devices are self-worn sensors and measure PA based on activity counts derived from high-resolution acceleration signals obtained at the minute-level, or even second-level.

Most studies utilizing accelerometers have focused on aggregate or summary statistics such as daily total or weekly average activity counts or minutes of moderate-vigorous physical activity (MVPA) [182]. While such summary measures of activity are easy to understand and implement using standard statistical techniques, aggregating activity records to daily or weekly averages results in a loss of information. In particular, summarizing precludes the evaluation of temporal variation (e.g., the timing during the day) of PA, which may provide additional insight into associations between diurnal variation in PA and health outcomes.

Functional data analysis is a powerful and well-studied statistical method [146, 181] for modeling curves or functions that are continuous. In the context of PA, functional data methods, and functional principal components analysis (FPCA) in particular, can better elucidate patterns of the full spectrum of accelerometer data. In essence, this approach treats each participant's activity profile as a single functional datum, rather than reducing it to a scalar summary. Various models have been developed to explore the minute-level information, extrapolating from densely sampled accelerometer inputs rather than simply implementing daily or weekly summaries [146, 123, 20, 33, 60, 164]. The review paper by Ramsay et al. [146], provides an overview of methods and applications in FPCA. The main idea is to decompose the dense signal inputs and to extract the principal variation directions, thus reducing the dimension. The FPCA searches for a set of mutually orthogonal and normalized weight functions to summarize subject-specific features. This idea was generalized to multilevel FPCA, which captures both the intra- and inter-subject variation [33]. In addition, Greven et al. [60] proposed longitudinal FPCA to include dynamic subject-specific variability and Shou et al. [164] extended the analysis to decompose the variability of any functional model with a particular linear structure via structured

FPCA. Due to the hierarchical structure of our data with repeated days clustered within subjects and visits, longitudinal FPCA is implemented in this paper in order to obtain information from the entire accelerometer signal inputs, while at the same time accounting for the nested structure of our data.

Much statistical research has focused on developing regression models to evaluate associations between these functional measures of activity and health outcomes. Crainiceanu et al. [31] proposed a framework for regression models where the functional predictor is repeatedly observed but the response is a scalar variable. Along these lines, our previous study [193] implemented a multilevel FPCA to characterize subject- and visit-level variation, and used the corresponding principal component scores as predictors to examine associations between PA patterns and health outcomes. Similarly, several studies have utilized functional data methods to investigate accelerometer-measured physical activity and health [161, 6, 13], but these studies have been primarily cross-sectional, and/or the methods do not apply to longitudinally collected exposures (i.e., physical activity) and health outcomes, which is a focus of prospective epidemiologic studies. There have been methodological advances in the statistical realm. To model the longitudinal structure, Goldsmith et al. [59] extended the spline-based estimation strategy on functional predictors and added subject-specific random effects to the standard cross-sectional setting. Furthermore, combined with longitudinal FPCA [60], Gertheiss et al. [57] were able to incorporate the longitudinal structure of the functional predictors in the regression model. Thus, these models include subject-specific effects and functional predictors in the regression model, and can be summarized as functional mixed effect models.

In this paper, extending our earlier cross-sectional investigation [193], we implemented longitudinal FPCA and functional mixed effects models to investigate associations between diurnal PA patterns and longitudinal health outcomes. To this end, we leveraged data from a dietary intervention weight-loss trial of 245 overweight women (the MENU Study [154, 97]) with accelerommtry and a wide array of glucoregulatory and inflammatory biomarkers collected at three visits over 12 months. We used a two-step approach. In the first step, a longitudinal FPCA was applied to incorporate subject- and visit-specific variability when decomposing functional inputs. In the second step, mixed effect models were fitted with functional predictors from the first step to inform the association between PA and health outcomes. By applying this procedure, we not only addressed the subject-to-subject and visit-to-visit variation in activity patterns, but also made more nuanced inferences about how diurnal patterns of physical activity could longitudinally affect weight-loss and biomarkers related to obesity.

4.2 Study Overview

The MENU weight-loss study (2011-2017) [154, 97], a project in the UC San Diego NIH-funded Transdisciplinary Research on Energetics and Cancer (TREC) Center, comprised of 245 non-diabetic and overweight/obese women. Participants were randomized to one of three diet arms for investigating how variation in macronutrient diet composition impacted weight loss and cardiometabolic biomarkers. All participants across the diet arms also received a physical activity intervention. Eligibility criteria for study participation were age ≥ 21 years, body mass index (BMI; kg/m^2) between 27 and 40, and willingness and ability to participate in clinic visits, group sessions, and telephone and internet communications during the 12-month study.

Clinic visits, measurements and data collection occurred at three time points: baseline and 6 and 12 months. Demographic data, including age, ethnicity and smoking status, were collected only at the baseline visit. Fasting levels of C-reactive protein (CRP), insulin and body mass index (BMI) were measured at each visit and these constituted the longitudinal (scalar) health outcomes in our analysis. In general, larger values of each outcome indicate worse health status. Insulin and CRP were log-transformed so that the distribution of the transformed data is close to Gaussian. Details of the study protocol and main results have been previously published [154, 97].

Physical activity (PA), measured with accelerometer devices GT3X Actigraph (ActiGraph, LLC; Pensacola, FL), was recorded daily at each visit. The devices collect acceleration data at 30 Hz on the x, y, z axes and then the ActiLife program applies a band-pass filter to remove non-human acceleration signals from the data. The triaxial activity counts vector (AC_x, AC_y, AC_z) are summarized as magnitudes $\sqrt{AC_x^2 + AC_y^2 + AC_z^2}$, which are referred to as activity magnitude in the manuscript and related to intensity of the activity [11]. These activity counts can be categorized into minutes spent in sedentary, light, moderate, and vigorous activity using calibration thresholds. Participants were instructed to wear the devices for 7 days during waking hours, except when in contact with water. Non-wear time was identified via pre-defined algorithms of consecutive zero counts using standard protocols [26] and labeled as missing data. Records with at least 10 hours of device wear (per standard protocols) were retained. The final dataset includes accelerometer data for 4259 days for 245 participants; 4 records with fewer than 10 hours of wear were removed. All participants received the same physical activity intervention regardless of diet group. Thus for the current investigation the three diet arms are combined and the study is analyzed as a longitudinal cohort.

4.3 Statistical Model

4.3.1 Accelerometer Data Processing

We proposed analysis models based on the PA time-series inputs. Figure 4.1 presents an example of activity records from 6:00 am to 11:29 pm for one participant on the first day of each visit. Each data point (y-axis) represents minute-wise PA activity magnitude. Based on calibration studies on energy expenditure, sedentary time is defined as minutes with activity magnitude < 200, and moderate to vigorous physical activity (MVPA) time is defined as minutes with activity magnitude > 2690 [158].

As shown in Figure 4.1(a), the starting time and duration time are not constant for a given participant across days, and furthermore, these measures also vary among participants. To be specific, the mean time for participants to start wearing the devices was around 7 am (SD 127 min), indicating that participants, generally started daily activities in the morning. Therefore, to ensure a more consistent and balanced data structure, we re-aligned daily records, so that all participants had a “common” starting time of device wear denoted as ”0” on the x-axis in subsequent plots, so that 10-hours of device wear are recorded as 0 to 600 minutes (10 hours) on the x-axis. This realignment ensures that the start and end times across all days and participant activity profiles are on the same grid of points.

Lastly, the daily activity data for each participant were averaged over days within each visit to obtain an averaged PA profile for each visit. Of note, the mean (SD) number of days of device data per participant was 3.9 (SD 2.1). Sensitivity analysis was performed to assess the impact of the averaging on our findings, and results are included in the Appendix B.3.

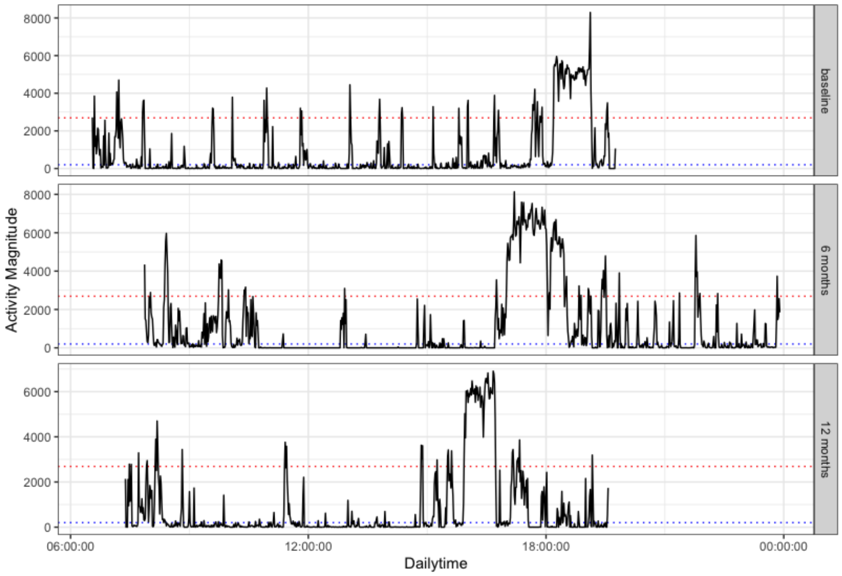
By smoothing the averaged daily activity, Figure 4.1(b) shows the overall population mean PA intensity curve over 600 consecutive minutes, as well as the mean at each visit. As noted above, time “0” on the x-axis (Figure 4.1(b)) indicates the common start time for all participants (after realignment), and 10-hours of device wear are recorded as 0 to 600 minutes.

To account for the hierarchical structure of the data (visits within subjects) and its longitudinal nature in both predictors (PA) and health outcomes, we applied a longitudinal FPCA model to decompose densely sampled PA data, and a (functional) mixed effects regression model to explore the association between predictors and outcomes.

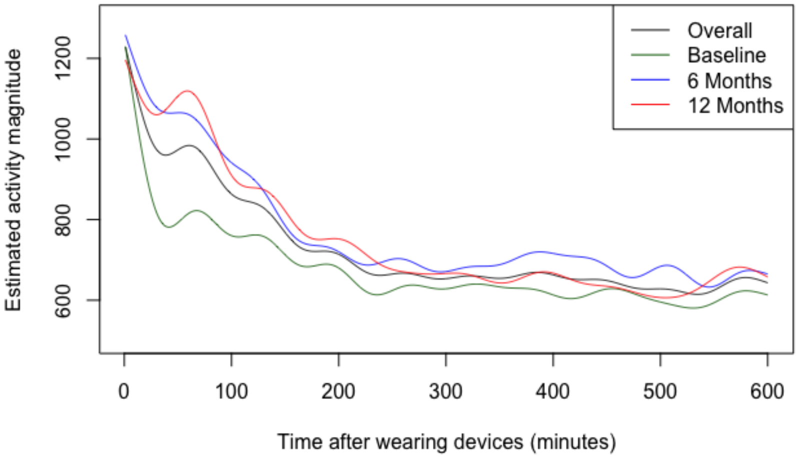
4.3.2 Longitudinal FPCA

Assuming no measurement error, a multilevel FPCA [33] can decompose an activity record $X_{ij}(t)$ for each subject i ($i = 1,2,...,N$) at time $t \in \mathscr{D}$ (measured at the minute-level in the current analysis and $\mathscr{D}$ can be treated as a set of grid points with length D) at each visit j ($j = 1,2,...,n_i$) in the form of

$$X_{ij}(t) = \mu(t) + U_i(t) + V_{ij}(t), \tag{4.1}$$



(a)



(b)

Figure 4.1. The plots provide (a) an example of activity patterns from minute-level accelerometer count data for one subject across three visits: the raw activity curve (black solid line), the sedentary count threshold(blue dotted line) and the MVPA count threshold (red dotted line); (b) the smoothed overall and visit-level mean activity magnitude curves at baseline, 6 months and 12 months. The y-axis denotes estimated activity magnitude and the x-axis depicts a time sequence from the start of devices wear (0) up to 600 minutes.

where $\mu(t)$ represents the overall population mean function at t. $U_i(t)$ is the subject-specific deviation from the overall mean function. $V_{ij}(t)$ is the subject- and visit- specific deviation from the subject-mean function. The subject-specific variation can be further decomposed into the sum of a static part and a longitudinal part, which forms the basis of the longitudinal FPCA structure [60]. The detailed derivation of the model was given in Greven et al. (2011) and we will briefly describe it under our study setup. Specifically, for a two-level model, the functional input can be rewritten as,

$$X_{ij}(t) = \mu(t) + U_{i0}(t) + U_{i1}(t)T_{ij} + V_{ij}(t), \tag{4.2}$$

where $U_{i0}(t)$ is the random functional intercept for subject i, $U_{i1}(t)$ is the random functional slope for subject i and T_{ij} is the time at visit j for subject i, and in our application T_{ij} has the form $T_{ij} = j$. To ensure the identifiability of the model, $\mathbf{U}_i(t) = (U_{i0}(t), U_{i1}(t))$ and $V_{ij}(t)$ are assumed to have mean zero and be mutually uncorrelated. $K_U(s,t) = cov\{U_i(s), U_i(t)\}$ and $K_V(s,t) = cov\{V_{ij}(s), V_{ij}(t)\}$ are covariance operators for the above random processes and K_U and K_V represent the corresponding covariance matrices for all $s,t \in \mathscr{D}$. Furthermore, for the subject-specific variation $K_U(s,t)$, the covariance operator between the bivariate process $\boldsymbol{U}_i(t)$ has two parts: the auto-covariance $K_{U_0}(s,t)$, $K_{U_1}(s,t)$ and the cross-covariance $K_{U_{01}}(s,t)$, which is represented as:

$$K_U(s,t) = \begin{pmatrix} K_{U_0}(s,t) & K_{U_{01}}(s,t) \\ K_{U_{01}}(t,s) & K_{U_1}(s,t) \end{pmatrix}. \tag{4.3}$$

Therefore, by Karhunen–Loéve expansion [89] on $\mathbf{U_i}(t)$ and $V_{ij}(t)$, we obtain

$$X_{ij}(t) = \mu(t) + \sum_{l=1}^{\infty}(1, T_{ij})\xi_{il}\phi_l^{(1)}(t) + \sum_{m=1}^{\infty}\zeta_{ijm}\phi_m^{(2)}(t), \tag{4.4}$$

where $\phi_l^{(1)}(t) = (\phi_l^{U_0}(t), \phi_l^{U_1}(t))'$ are the ordered eigenfunctions of $K_U(s,t)$ with corresponding eigenvalues λ_l^U and $\phi_m^{(2)}(t)$ are the ordered eigenfunctions of $K_V(s,t)$ with corresponding eigenvalues λ_m. Specifically, eigenfunctions $\phi_l^{(1)}(t), l \in \mathbb{N}$, are elements of $L^2[0,1] \times L^2[0,1]$

and satisfy the additive scalar product $< (f_0, g_0), (f_1, g_1) >= \int_0^1 f_0(t)g_0(t)dt + \int_0^1 f_1(t)g_1(t)dt$. Details of the derivation can be found in Appendix B.1. The corresponding principal component scores have the forms,

$$\xi_{il} = \int U_{i0}(s)\phi_l^{U_0}(s)ds + \int U_{i1}(s)\phi_l^{U_1}(s)ds \quad \text{and} \quad \zeta_{ijm} = \int V_{ij}(s)\phi_m^{(2)}(s)ds, \tag{4.5}$$

and are uncorrelated with mean zero and variances λ_l and λ_m, respectively. In this way, the covariance operator of the longitudinal functional model becomes

$$\begin{aligned} Cov\{X_{ij}(s), X_{ij'}(t)\} = K_{U_0}(s,t) + (T_{ij} + T_{ij'})K_{U_{01}}(s,t) + T_{ij}T_{ij'}K_{U_1}(s,t) + K_V(s,t)\delta_{jj'}, \\ \delta_{jj'} = \begin{cases} 1, & \text{if } j = j' \\ 0, & \text{otherwise} \end{cases}. \end{aligned} \tag{4.6}$$

Here, $\{K_{U_0}(s,t), K_{U_1}(s,t), K_{U_{01}}(s,t), K_V(s,t), s,t \in \mathscr{D}\}$ are estimated by linearly regressing $X_{ij}(s)X_{ij'}(t)$ on (1, T_{ij}, $T_{ij'}$, $T_{ij}T_{ij'}$,$\delta_{jj'}$) after mean-centering $X_{ij}(t)$. Eigenfunctions and eigenvalues of the estimated covariance matrices $\{\hat{K}_{U_0}, \hat{K}_{U_1}, \hat{K}_{U_{01}}, \hat{K}_V\}$ can be obtained via spectral decomposition, i.e. $\hat{K}_U = \sum_{l=1}^{2D} \hat{\lambda}_l^U \hat{\phi}_l^{(1)} \{\hat{\phi}_l^{(1)}\}'$ and $\hat{K}_V = \sum_{m=1}^{D} \hat{\lambda}_m^V \hat{\phi}_m^{(2)} \{\hat{\phi}_m^{(2)}\}'$. It is proved in Greven et al. (2011)[60] that if the time variable T_{ij} is standardized to have zero mean and unit variance, i,e, $E(T_{ij}) = 0$ and $Var(T_{ij}) = 1$, the variation in $X_{ij}(t)$ can be decomposed additively and expressed with respect to the estimated eigenvalues, $\int_{\mathscr{D}} var(X_{ij}(t))dt = \sum_{l=1}^{\infty} \lambda_l^U + \sum_{m=1}^{\infty} \lambda_m^V$. Usually a few most informative eigenfunctions are retained for further analysis. Criteria for selecting a finite number, N_U and N_V, of subject- and visit level eigenfunctions is discussed in Section 4.3.3. This finite sum then replaces the infinite sum in equation 4.4.

For fixed N_U and N_V, equation 4.4 is a linear mixed model and we use the best linear unbiased prediction (BLUP) to obtain the predicted principal component scores ξ_{il} and ζ_{ijm}. Let $\hat{\boldsymbol{\beta}} = (\hat{\xi}_{11}, \ldots, \hat{\xi}_{1N_U}, \ldots, \hat{\xi}_{N1}, \ldots, \hat{\xi}_{NN_U}, \hat{\zeta}_{111}, \ldots, \hat{\zeta}_{11N_V}, \ldots, \hat{\zeta}_{Nn_N1}, \ldots, \hat{\zeta}_{Nn_NN_V})$, then estimated

BLUP of $\hat{\boldsymbol{\beta}}$ is given by,

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X}. \tag{4.7}$$

where $\boldsymbol{Z} = [\boldsymbol{E_I} \otimes \boldsymbol{\Phi}^{U_0} + \boldsymbol{T} \otimes \boldsymbol{\Phi}^{U_1} | \boldsymbol{I} \otimes \boldsymbol{\Phi}^{V}]$, $\boldsymbol{E_I} = (\delta_{ijh})_{ij=11,\ldots,Nn_N;h=1,\ldots,N}$, $\boldsymbol{T} = (T_{ij}\delta_{ijh})_{ij=11,\ldots,Nn_N;h=1,\ldots,N}$, $\boldsymbol{\Phi}^{U_0} = \{\phi_l^{U_0}(t)\}_{t\in\mathscr{D},l=1,\ldots,N_U}$, $\boldsymbol{\Phi}^{U_1} = \{\phi_l^{U_1}(t)\}_{t\in\mathscr{D},l=1,\ldots,N_U}$, $\boldsymbol{\Phi}^V$ $= \{\phi_l^V(t)\}_{t\in\mathscr{D},l=1,\ldots,N_V}$, $\boldsymbol{I}$ is the $\sum_i N_i$ dimensional diagonal matrix with element 1, $\boldsymbol{X} = [\{X_{11}(t)\}_{t\in\mathscr{D}}, \ldots, \{X_{1N_1}(t)\}_{t\in\mathscr{D}}, \ldots, \{X_{N1}(t)\}_{t\in\mathscr{D}}, \ldots, \{X_{Nn_N}(t)\}_{t\in\mathscr{D}}]$, and $\otimes$ denotes the Kronecker product of matrices. $(\delta_{ijh})_{ij=11,\ldots,Nn_N}; h = 1, \ldots, N$ denotes the indicator matrix with entries δ_{ijh} at row $ij, i = 1, \ldots, N, j = 1, \ldots, n_i$ and column $h, h = 1, \ldots, N$, with $\delta_{ijh} = 1$ if $i = h$ and $\delta_{ijh} = 0$ otherwise.

Although the methods were described in detail in Greven et al. (2011) [60], our no measurement error setting differs slightly from the model specified in the original paper. Therefore, for completeness we provide the proof of the BLUP derivation (see Appendix B.1). In addition, we implemented simulation studies, in order to illustrate the applicability of the proposed methods, and to evaluate how higher values for subject-level versus visit-level variation (and vice-versa) influenced goodness of fit of the various model components. The simulation assumptions and results can be found in Appendix B.2. We discuss a few key results here. The boxplots of the estimated normalized errors of principal component scores show all parameters are unbiasedly estimated, demonstrating agreement with the simulation results in Greven et al. (2010) [60].

In addition to results from parameter estimation, we include residual mean square error (MSE) results in Appendix B.2 from each of the two simulation scenarios with three ways of computing residuals $R_{ij}(t)$, the residuals from subject level $X_{ij}(t) - U_i(t)$, the residuals from visit level $X_{ij}(t) - V_{ij}(t)$ and the overall residuals $X_{ij}(t) - U_i(t) - V_{ij}(t)$. Let M be the total number of observations, the residual MSE for one simulation replicate is defined as $\frac{1}{M}\sum_{i,j}(\sum_t |R_{ij}(t)|)^2$, which in fact reflects the total mean squared count difference per observation between the predicted and observed activity curves, when using only level-1 predictions, only level-2 predictions or both. Thus this mean-squared error represents the goodness-of-fit of the model when using different fitted components. Since larger eigenvalues indicate more explained variability, the goodness of fit of the subject- versus visit-level predictions depends on which component has the largest eigenvalue, as seen from the two simulation scenarios.

The simulations confirm that the estimated principal component scores ξ_{il}, ζ_{ijm} and hence the decomposed random processes $\mathbf{U_i}(t)$, $V_{ij}(t)$ obtained from the longitudinal FPCA model are reasonably accurate at recapitulating the observed temporal patterns of subject- and visit-level PA. We will use the PA patterns as predictors of outcomes in regression models, as detailed in the next section.

For data observed with white noise, denoted as $\tilde{X}_{ij}(t) = X_{ij}(t) + \varepsilon_{ij}(t)$, as suggested in Shou et al. 2015 [164], smoothing the raw data $\tilde{X}_{ij}(t)$ can be implemented before performing the longitudinal FPCA. Since the main purpose of this study is to explore the associations between general activity patterns and health outcomes, smoothing the raw inputs is preferable for these densely sampled accelerometer inputs. .

4.3.3 Regression model

With results from the longitudinal FPCA, the associations between physical activity and health outcomes are explored via regression modeling. Two regression models, regression modeling with principal component scores (PCR) and functional regression model with decomposed random processes (fPCR), are implemented in our analysis, and briefly discussed in this section. The first regression model directly incorporates subject- and visit-level principal component scores as predictors. To account for the repeated measures pattern in outcomes Y_{ij}, we use linear mixed models. Thus, the PCR is given as,

$$E(Y_{ij}) = \alpha_0 + \alpha_1 I(j > 1) + b_i + \sum_{l}^{N_U} \beta_l^U \xi_{il} + \sum_{m}^{N_V} \beta_m^V \zeta_{ijm} + \texttt{other covariates}, \qquad (4.8)$$

where the α and β parameters are fixed effects, namely, α_0 is the intercept at baseline visit and α_1 is the mean change at follow-up visits, and βs quantify associations between diurnal activity pattern (captured via subject- and subject-visit principal components and scores); b_i is a subject-specific random effect and the assumptions $b_i \sim N(0,\varepsilon^2)$ and b_i is conditionally independent of Y_{ij} hold. 'other covariates' refers to covariates which one might adjust for, which will depend on the particular study. In our application to the MENU study, we adjusted for age, ethnicity, smoking status, and follow-up visit. The number of components N_U and N_V are chosen to explain a pre-specified proportion of variance and in our application, we will choose enough components to explain over 85% variance. The fixed effects β^U, β^V and random effects b_i are estimated with R package `lme4` [12].

Another regression model we consider in this paper is the fPCR, which replaces principal component scores with functional curves as predictors. The functional predictors include between-subject variation $U_i(t)$ and between-visit variation $V_{ij}(t)$,which can be reconstructed in the form of $U_i(t) = \sum_l^{N_U} \xi_{il}\phi_l^{(1)}(t)$ and $V_{ij}(t) = \sum_m^{N_V} \zeta_{ijm}\phi_m^{(2)}(t)$. Here $U_i(t)$ is interpreted as the overall trend for subject i while subject-visit variation is captured by $V_{ij}(t)$. The fPCR model then has the form,

$$E(Y_{ij}) = \alpha_0 + \alpha_1 I(j>1) + b_i + \int \beta_U(t)U_i(t)dt + \int \beta_V(t)V_{ij}(t)dt + \texttt{other covariates}. \tag{4.9}$$

The αs and b_i have similar interpretation as the PCR model. The β parameters are now represented as smooth coefficient functions $\beta_U(t)$ and $\beta_V(t)$, and are estimated using penalized spline methods in our application via the R package `mgcv` [191, 190].

The estimated principal component scores quantify the extent to which a subject or subject-visit subscribe to the corresponding temporal patterns delineated by the principal components. Thus, as noted in Gertheiss et al. (2013) [57], by incorporating principal component scores as covariates, the PCR assesses associations between activity patterns and outcomes, and thus may have intuitive appeal. However, PCR is subject to overfitting, due to the need to a priori choose the number of principal components (N_U and N_V). The fPCR, on the other hand, is more flexible and can yield a more nuanced interpretation, especially when the coefficient functions are significant for some time domains. We will demonstrate the comparison in later sections.

4.4 Results

4.4.1 Sample characteristics

The study population had average age of 50.8 years (SD 9.9), with range 22-72 years; 81.6% were non-Hispanic and 69% had no history of smoking. In addition, summary information of insulin, C reactive protein (CRP) and body mass index (BMI) across the three visits are listed in Table 4.1. All three outcomes present a decreasing trend after the baseline visit, indicating improved health status at follow-up.

Summary statistics of physical activity by visit are included in Table 4.1 as well. Total magnitude computes the averaged sum of activity counts for a participant at each visit and is a measure of total activity. We also present standard metrics for PA study, including daily sedentary time and MVPA time. The increasing average movement magnitudes, shorter sedentary time and longer MVPA at follow-up visits imply that on average, participants increased physical activity after enrolling in this study. Boxplots of daily-average activity magnitudes at individual level are provided in Appendix B.4, which further establish an increase in PA magnitudes after baseline visits. Meanwhile, no notable seasonal variability was detected for this one-year longitudinal study, which is unsurprising for a study conducted in southern California.

4.4.2 Functional Physical Activity Patterns

For functional PA inputs, we fitted the longitudinal FPCA on averaged daily activity magnitudes, given the longitudinal design of our study. The number of principal components for subject (level 1) and visit (level 2) level patterns, i.e. N_U and N_V, were chosen based on the percentage of explained variation, and an attempt to achieve balance between under-fitting of the covariance matrix and over-fitting the regression model. In this study, we retained sufficient components to ensure that 95% overall variation in activity patterns could be explained.

Five level 1 principal components and nine level 2 principal components explained 95% of activity variation and were retained. The detailed results are included in Appendix B.4, which gives cumulative variation explained for the first five components at each level. For the level 1 principal components, the first component for subject-specific process **U** explains 25% of the variation. Also, within **U**, most of the variation is explained by the random functional intercept U_0 (38.55%) while the random functional slope only explains $< 5\%$ of the variation, suggesting that variation between subjects is largely captured by overall PA amount rather than by longitudinal trends. Another 25% of the variation is explained by the first principal component of the level 2 visit-specific process V. Overall, the first five components of the subject-level process and visit-level process each explain around 43% variation, indicating that they capture equal amount of variation in the data.

Figure 4.2 illustrates the first three estimated principal components for the random intercept, random slope and visit-specific process by columns. The plots in Figure 4.2(a) depict the overall mean curve $\mu(t)$ (black curve) with adding (red) or subtracting (blue) the value of 2 square root of eigen values multiplying first (or second level) principal component curves (i.e., $\pm 2\sqrt{\lambda_l}\phi_l^U$ or $2\sqrt{\lambda_m}\phi_m^V$), respectively. The plots in Figure 4.2(b) represent the eigenfunctions themselves and together these sets of plots can be used to interpret the PA patterns associated with each principal component. For instance, the first level 1 intercept principal component (top left in Figure 4.2(b)) is above the horizontal line at 0 throughout the 600 minutes, and represents an overall vertical shift of the mean activity curve. As seen in the corresponding top left plot in 4.2(a), the red curve, which represents adding (a multiple of) this principal component to the mean, is always higher than the mean curve. Thus a high score on this component indicates that a participant is on average more physically active throughout the time interval compared to one with a lower value. It is also observed that the peak of this curve appears at around an hour after wearing, showing that early activity is more notable for capturing between subject variability. The first level 1 random slope process curve (top middle plot of Figures 4.2(a) and 4.2(b)) show a similar pattern but with smaller variance. We also note that in the Karhunen–Loéve expansion, the level 1 intercept and slope eigenfunctions share the same level 1 score. This implies that a subject with a higher score in the first level 1 component will not only be more active overall, but also show a higher increase across visits.

The other level 1 components illustrate variation in timing of activity and identify periods of higher versus lower activity. For instance, the second level 1 intercept component (middle left plot in Figure 4.2(b)) is negative (i.e., below the horizontal line at 0) for the first 100 minutes and then becomes positive for the remaining 500 minutes, which indicates a contrast between earlier versus later activity. This is further evident in the middle left panel in Figure 4.2(a), where the red curve is below the mean for the first 100 minutes and then switches to being above the mean. A high positive score on this component would signify less activity in the early period (i.e., first 100 minutes) with increased activity later on.

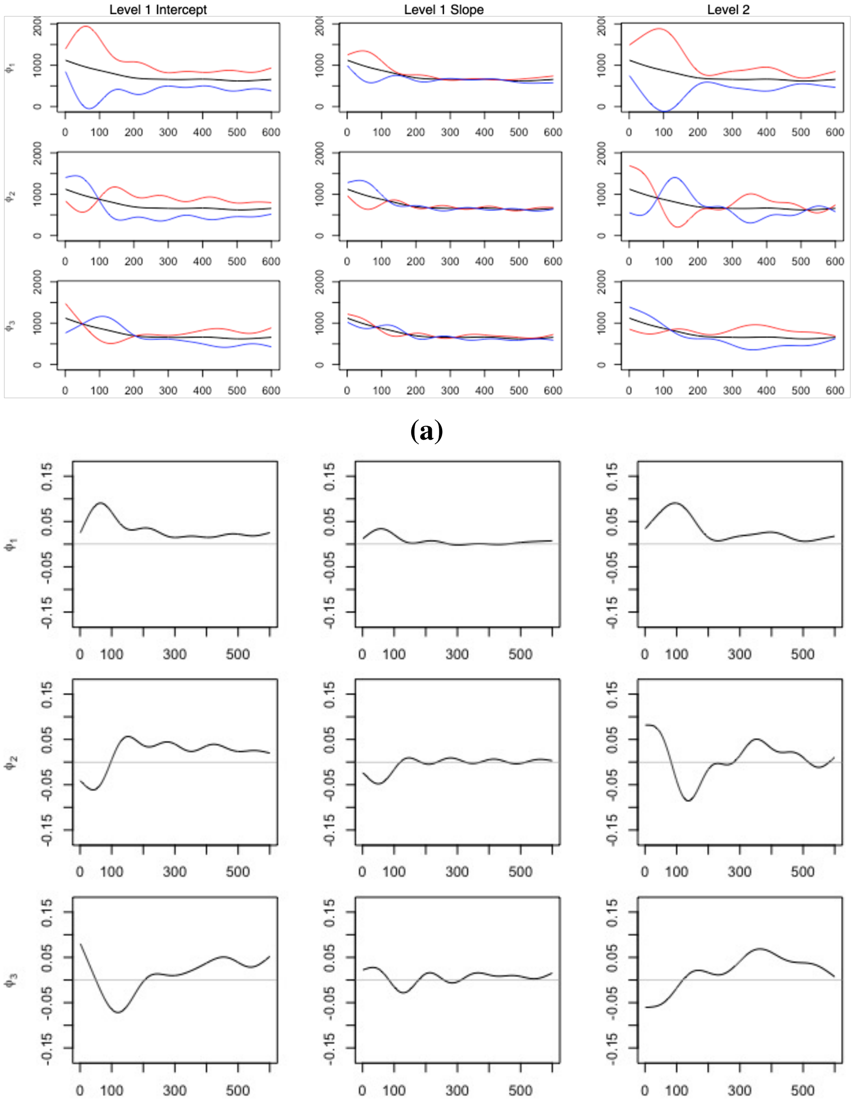
The first level 2 visit-specific curve, on the other hand, captures visit-to-visit shift from the subject-level curve. A participant with a higher score on this component would be more physically active longitudinally, based on the red curve in the top right panel of Figure 4.2(a) being always above the mean, or equivalently the curve in Figure 4.2(b) being always positive (i.e., above the horizontal line at 0). The peak for this curve appears at around 100 minutes and shows a delayed pattern compared with the first level 1 process, suggesting that visit-to-visit variation is more pronounced at the later morning time.

For each principal component, the corresponding principal component score quantifies the magnitude of the temporal pattern associated with that component. Thus, the principal component score itself can be used as a quantified indicator of the variation in PA records. To demonstrate this, two examples are given in Figure 4.3. In Figure 4.3(a), an individual example with a large first level 1 principal component score but a small first level 2 principal component score is given, showing a significant early-time bounce at both visits, with little variation between visits. Figure 4.3(b) presents an individual example with a small first level 1 principal component score but a large first level 2 principal component score and in this case, the large variation between visits is apparent. Detailed decomposition figures are included in Appendix B.4, illustrating a step-wise reconstruction after decomposition. These two examples to some extent also reflect our simulation results of residual MSEs (Section 4.3.2, Appendix B.2), since larger eigenvalues are more likely to have higher scores. Both examples further illustrate Figure 4.2, and demonstrate how level 1 versus level 2 principal component scores are useful for evaluating between- and within-subject activity patterns. It is also evident that the fitted (smoothed) FPCA curves track the original activity counts reasonably well, indicating that our fitted model provides a good fit to the data.

4.4.3 Regression Patterns: Associations between Physical Activity and Health Outcomes

In regression analysis, we first implemented the PCR models to explore the association between PA and health outcomes. In these models, physical activity patterns are modeled with estimated principal component scores as predictors, similar to the model in equation 4.8. Table 4.2 gives the results of the regression model, adjusting for baseline age, ethnicity, smoking status, and a logical variable indicating whether the participant is at a follow-up visit. The model accounts for individual variation by adding a random intercept b_i. The regression coefficients of the visit indicator and the first two principal component scores for both levels, which explained over 70% of variance jointly, are given in Table 4.2. For the purpose of comparing and interpreting model coefficients, all level 1 and level 2 principal component scores are also scaled to be in the range of 0 and 1.

All three health outcomes were negatively associated with the visit indicator, reflecting decreasing levels at follow-up, i.e., after the intervention. The first level 1 principal component scores were negatively associated with insulin, CRP and BMI, suggesting that more PA was associated with lower levels of these health outcomes, i.e., higher PA is associated with better



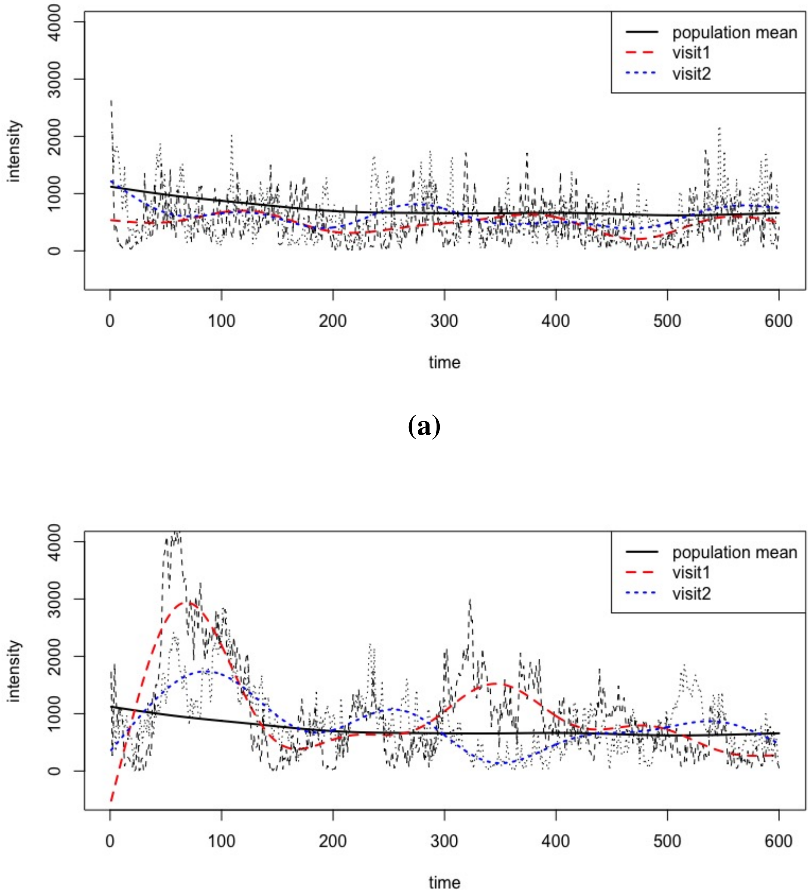
(a)

(b)

Figure 4.2. The first three estimated principal components for the random intercept (left column), random slope (middle column) and visit-specific process (right column). The plots give the (a) overall mean value curve $\mu(t)$ (black) with addition (red) or subtraction (blue) of 2 square root of eigen values multiplying first or second level principal component curves ($\pm 2\sqrt{\lambda_l}\phi_l^U$ or $2\sqrt{\lambda_m}\phi_m^V$) respectively; (b) estimated eigenfunctions of the first three principal components. The horizontal gray line represents 0.

metabolic health. In addition, the first level 2 principal component scores were negatively associated with insulin and CRP, suggesting that increased PA between visits (within an individual) was associated with greater decline in biomarkers.

To compare PCR to standard methods which use physical activity summaries, we also



(a)

(b)

Figure 4.3. Two examples of PA records with raw count inputs (black) and estimated curves at each visit (red, blue): (a) is an example with a large first level 1 principal component score but a small first level 2 principal component score; (b) is an example with a small first level 1 principal component score but a large first level 2 principal component score.

fitted a mixed effect regression model by including total (averaged) activity counts and MVPA as predictors respectively (Appendix B.4), whose values were also scaled to be in the range of 0 and 1. It shows that both total activity counts and MVPA also exhibit a negative association with health outcomes, which supports findings from PCR models. However, the analysis based on daily summary PA estimates such as total activity counts or MVPA fails to capture the temporal aspect of PA accumulation, e.g., the level 1 first principal component of the intercept process suggests that peak activity occurred at around an hour from the start time. We further elucidate on these and other differences between standard and functional regressions methods in the Discussion section.

Along with the PCR, fPCR models were also implemented to better exemplify the diurnal association between PA and health outcomes. Figure 4.4 presents the estimated coefficient functions with 95% pointwise confidence intervals. The coefficients at a given time-point (on the x-axis) are considered significant if the 95% confidence limits at that time do not cross the reference horizontal line at $y = 0$. As shown in the figure, the coefficient functions for the level 1 and 2 processes for log(insulin) were negative and significant for most time-points of the day, suggesting that participants with more PA (irrespective of time of accumulation) than the ”average participant” or the ”previous visit” tended to have lower insulin. These effects were stronger (and significant) if PA occurred during earlier times (of day) for the level 2 coefficients for log(insulin), indicating stronger effects for visit-to-visit change in PA earlier in the day. Similar results were observed for BMI, although the level 1 coefficients were minute-wise significant up to approximately the first 300 minutes (see x-axis), whereas the level 2 coefficients were significant throughout the day. Interestingly, both level 1 and level 2 coefficients for BMI showed an initial increasing pattern with leveling off later, suggesting that PA earlier (rather than later) in the day was more beneficial for reducing weight. The effect of PA on log(CRP) was not significant in the first level but showed a pointwise negative association in the second level coefficients only during the first 100 minutes of wear.

4.5 Discussion

In this work we have demonstrated the use of functional principal component analysis to extract patterns of physical activity from accelerometer data, and use these patterns to evaluate associations between PA and health outcomes. Functional data analysis provides a rich statistical framework for modeling the variation of physical activity curves. While summary statistics such as weekly total activity counts or MVPA provide aggregated metrics, functional data analysis can unravel temporal patterns, and presents varying activity patterns of individuals throughout the day.

Conventional approaches usually summarize statistical characteristics from accelerator data (e.g., mean weekly MVPA), and then use these summaries to examine longitudinal associations between PA and health outcomes. These methods ignore the full spectrum of activity magnitude trajectories. On the other hand, functional modeling allows a more robust decomposition of the original accelerometer inputs, and thus could provide a richer framework for examining PA-health associations. From mixed effect regression models by including conventional summary measures of PA as predictors, such as total activity counts and MVPA, the results (Appendix B.4) show high concordance with coefficients of the first level 1 and level 2 principal components scores from the PCR models (Table 4.2), which in fact can be interpreted as measuring the average amount and visit-visit change of activity for each individual. However, the total activity counts and MVPA in the model did not explicitly separate the subject-level (level 1) and between-visit variations (level 2). Importantly, summary PA measures such as weekly total activity or MVPA cannot identify associations between diurnal variation in PA accumulation and health, which is exemplified by fPCR to further extend the regression model, by using smooth coefficient functions to explain predictors' influence on health outcomes. From the coefficient functions for the health outcomes (Figure 4.4), the fPCR shows advantages by providing a trend of changing coefficients over time (of day). The level 1 and level 2 functional coefficients are negatively associated with the outcomes, which is in conformity with the findings for both, standard summary measures and PCR. However, from fPCR we are also able to discern that the level 1 and 2 coefficients for BMI, albeit negative, increase during the day, indicating that earlier activity is potentially more beneficial for weight management among overweight women. Interestingly, while also negative, the level 1 coefficient function for log(insulin) is relatively stable throughout the day, suggesting that PA, irrespective of time of accumulation, is equally beneficial for controlling insulin level. Thus, the timing of activity during the day may differentially impact biomarker outcomes, a fact that would be useful for designing personalized activity interventions.

In addition, the implementation of longitudinal models emphasizes the statistical analysis of cross-visit variation in both functional PA predictors and scalar health outcomes. On one hand, longitudinal FPCA reveals how different PA patterns within one participant reflect either a more active or sedentary style, as the examples shown in Figure 4.3. On the other hand, the application of the fPCR, extends the interpretation of the regression coefficients to minute-level at each visit. This is advantageous because the PA inputs and predicted coefficients are correspondingly matched in the same scale. In contrast with ordinary regression coefficients of principal component scores (Table 4.2), which provide the association between the full daily activity profile and health outcomes, the fPCR approach treated regression coefficients as smooth functions of time and computed an estimated coefficient at each time-point. As it was mentioned in Dziak et al. [40], a motivation of incorporating coefficient curves is to look for a period of time during which the predictors are more strongly associated with outcomes. However, we urge caution when interpreting time intervals, as our results are based on pointwise 95% intervals, and thus could be subject to increased Type 1 error when considering multiple time-points.

Another advantage of using fPCR compared with PCR is consideration of the number of functional principal components. In PCR, the number of principal components to retain is usually determined based on explaining sufficient variability in functional inputs, which might result in overfitting the regression model. On the contrary, when fitting a fPCR model, this is not an important concern, since a penalty is used to avoid overfitting. Construction of random process predictors often requires a large enough number of principal components to capture important features, which makes the fPCR more robust when the first few principal components do not explain enough variation in the predictor.

Further research is needed to address several limitations of this work. Firstly, we only implemented the model on visit-level data averaged across days. Additional methodological work, beyond the scope of the current investigation, is required to extend our current model to a three-level longitudinal FPCA model, which can be fitted with daily inputs. For our application, a sensitivity analysis (Appendix B.3) demonstrated that averaging the day-level PA inputs did not materially affect our analysis or results. Also, we realigned all PA profiles to a common start time. We believe that this realignment initializes each record at the participant's own starting time which seems more appropriate for capturing an individual participant's wake-time activity patterns, compared to using an arbitrary and fixed clock time for all individuals. Even so, more advanced analytic and registration approaches may need to be considered, especially when performing analysis directly on the day-level data, and particularly for applications where the variation could be larger within this level. Thirdly, we used pointwise 95% confidence intervals, which are specific to a given time-point. Estimating confidence bands for functional data that will account for all time-points simultaneously, is an area of active research; we will consider these extensions in future work. Also, although our findings could be used to optimally design timing of PA interventions, we recommend replication in independent cohorts, given the exploratory nature of principal components analysis. Finally, the functional approach does not specifically delineate levels of activity intensity, e.g., MVPA from light activity. Compositional data analysis, an emerging and highly relevant area of research[38], allows evaluation of different (correlated) activities (e.g.,sleep, sedentary time, MVPA) in the same model. While the focus of our functional approach is to elicit diurnal patterns of overall activity, it may be interesting to incorporate multiple behaviors into a functional model, thus leveraging the strengths of both functional and compositional data analysis methodologies. We leave this to future investigations.

4.6 Conclusion

In summary, our longitudinal FPCA model offers a new approach for analyzing the association of physical activity patterns with health outcomes. We have demonstrated that functional modeling can not only yield comparable results with traditional PA summary statistics with longitudinal outcomes, but also provide further information on the time domain of daily activities, including the association between PA effects at certain times of the day and health outcomes. These findings could be useful for providing individualized activity guidelines for overweight women and to promote health and weight control. Importantly, the use of wearable sensors for PA is becoming more and more common in public health research. Use of functional data methods to explore PA patterns could offer a useful complement to summary-based PA measures.

Acknowledgements

DDS, CLR and LN were partially supported by funding from the National Cancer Institute (U54 CA155435-01); JZ, CD, and LN were partially supported by funding from the National Institute of Aging (PO1AG052352) and the National Institute of Diabetes and Digestive and Kidney Disease (R01DK114945); CD was partially supported by funding from the National Heart, Lung and Blood Institute (R01HL130483). The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Table 4.1. Summary statistics of health outcomes and daily physical activity at each visit (mean (SD)).

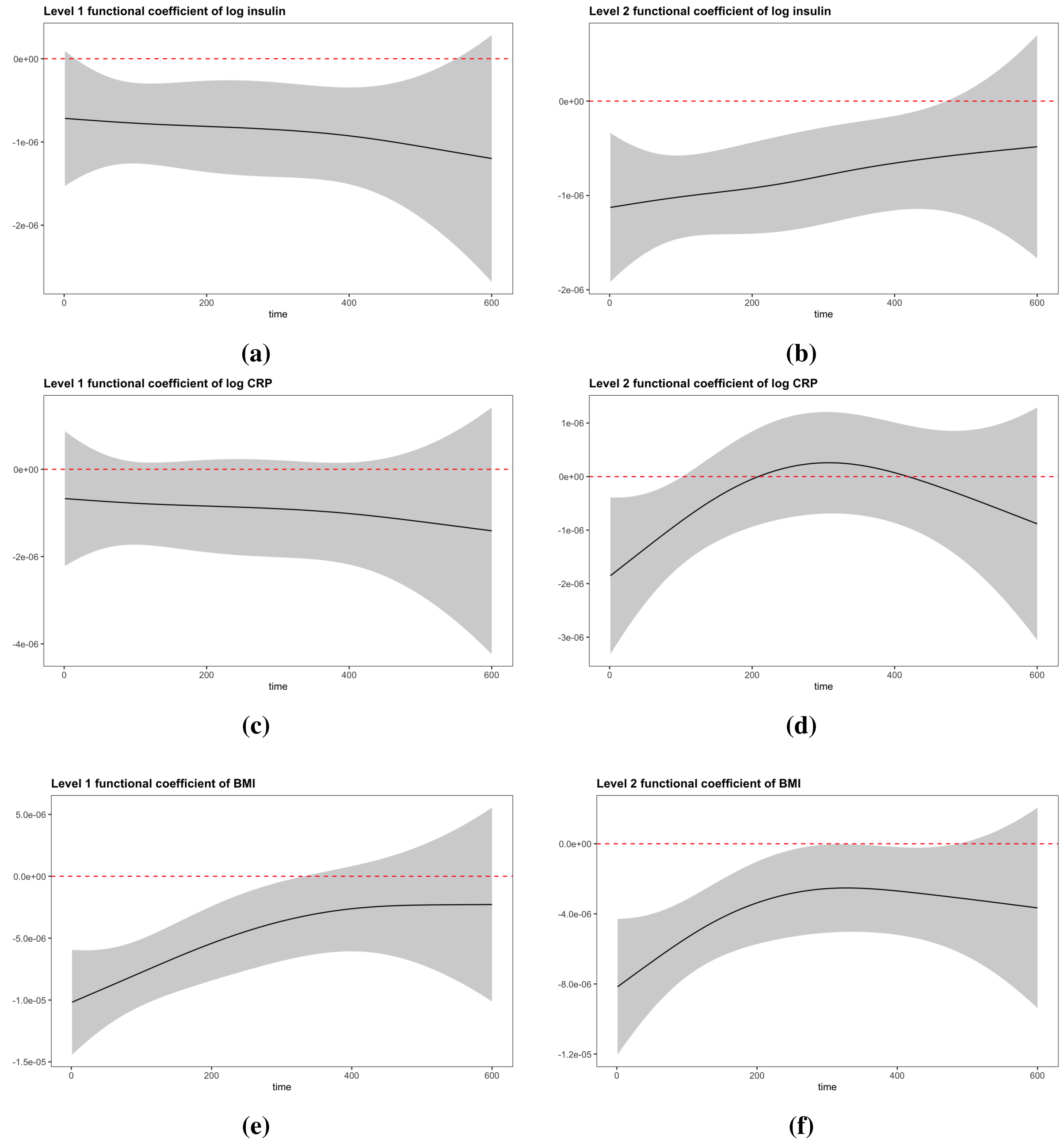
	Overall	Baseline	6 months	12 months
Health Outcomes				
Insulin (pg/mL)	13.44 (7.68)	14.49 (8.26)	12.63 (6.60)	12.98 (7.90)
CRP (μg/mL)	3.99 (4.70)	4.86 (5.28)	3.64 (3.88)	3.28 (4.60)
BMI	31.79 (4.12)	33.50 (3.34)	30.91 (3.98)	30.74 (4.43)
Activity magnitude				
Total magnitude counts	437,824.09 (201,831.34)	403,903.16 (174,556.80)	463,070.61 (217,989.27)	455,464.38 (210,793.04)
Sedentary time (mins)	312.33 (87.81)	318.22 (87.62)	308.93 (86.43)	308.14 (89.17)
MVPA time (mins)	40.62 (32.42)	34.43 (27.07)	45.42 (35.62)	43.61 (33.92)

Table 4.2. Linear mixed effect regression results of health outcomes on the first two level 1 and level 2 principal component scores of physical activity

Outcome	Predictor	Coefficient estimate	SE	Confidence interval
Log(insulin)	PC11	-0.15	0.04	(-0.23,-0.06)
	PC12	-0.08	0.03	(-0.14,-0.03)
	PC21	-0.16	0.04	(-0.24,-0.08)
	PC22	0.03	0.03	(-0.03,0.08)
	visit $>$ 1	-0.09	0.03	(-015,-0.04)
Log(CRP)	PC11	-0.23	0.09	(-0.41,-0.06)
	PC12	-0.11	0.06	(-0.24,0.01)
	PC21	-0.09	0.08	(-0.24,-0.06)
	PC22	-0.04	0.05	(-0.14,0.06)
	visit $>$ 1	-0.36	0.05	(-0.45,-0.27)
BMI	PC11	-0.74	0.29	(-1.31,-0.17)
	PC12	-0.31	0.23	(-0.75,0.14)
	PC21	-0.23	0.21	(-0.66,0.18)
	PC22	-0.02	0.14	(-0.3,0.25)
	visit $>$ 1	-2.56	0.13	(-2.81,-2.3)

*Adjusted for baseline age, ethnicity, smoking history and visit indicator.

**PC11 represents the first level 1, PC12 the second level 1, PC21 the first level 2, and PC22 the second level 2 principal component scores.



(a) **(b)**

(c) **(d)**

(e) **(f)**

Figure 4.4. Estimated functional coefficients curve (with 95% pointwise confidence intervals) when functional principal component regression (fPCR) models with functional predictors $U_i(t)$, $V_{ij}(t)$ and random intercept b_i are fitted (adjusted for baseline age, ethnicity, smoking history and visit indicator).

Chapter 5

Multilevel Longitudinal Functional Principal Component Analysis

5.1 Introduction

Currently, accelerometer-based devices are commonly used to characterize physical activity (PA) behavior in research and clinical trials [174]. These devices, such as GT3X Actigraph, are able to provide estimates at minute-level. Meanwhile, to acquire valid data, participants were required to wear the device at least 10 hours in 5-7 consecutive days at each visit [30, 109, 175, 97, 154]. However, other characteristics, such as related health outcomes, are normally measured once per subject or visit. Therefore, the flexibility of accelerometer-based devices allows a rich amount of data being objectively collected, but also brings challenges when more frequent high-dimensional PA data were acquired but relatively infrequent the outcomes of interest were measured. In this paper, we intend to propose a multilevel longitudinal functional principal component model to address the case when PA is measured more frequently than related health outcomes in a longitudinal study. Meanwhile, a comprehensive simulation study is applied to evaluate performance of scalar-on-function regression models when unbalanced data structure is observed between predictors and outcomes, in both cross-sectional and longitudinal scenarios.

The motivating dataset comes from the MENU trial, a 12-month behavioral intervention longitudinal study consisted of 245 overweight non-diabetic women [97, 154]. PA was recorded with GT3X Actigraph monitors, set to collect data at 30 Hz [11], for about a week per subject at each clinical visit. Each day's PA record for a subject and a visit is a densely and finely sampled function across the time interval. Therefore, these PA data are considered to have a three-level hierarchical structure under a longitudinal study design. Figure 5.1 displays an example of one subject's daily PA records on three random days obtained at baseline, 6 months and 12 months. Scalar health outcomes, such as body mass index (BMI), were acquired at subject and visit level, thus having a two-level structure. In fact, the study is one example of many biomedical studies, either having cross-sectional or longitudinal study designs, where the predictors are more frequently observed than the outcomes. The goal of this study is to explore the question regarding unbalanced study design, typically with accelerometer-measured functional predictors and scalar outcomes.

There are several existing statistical approaches for analyzing accelerometer data. The most common one is to derive algorithms that can translate the dense signal into many metrics, such as total/average time spent or counts of activity with varying intensities [48, 172, 176]. For instance, sedentary behaviors are defined as activity with less than 100 counts/minute [112] and moderate to vigorous physical activity (MVPA) time is defined as minutes with activity counts ¿ 2020 [172]. Bürgi et al. (2011) [22] investigated the cross-sectional and longitudinal relationship of PA with body fat and other health outcomes for preschool children, using total PA, moderate PA and vigorous PA summarized from at least 3 days of PA recording. Though these metrics provide useful summary of overall activity, they lose the ability to capture the correlation over time within a subject activity profile.

Functional data analysis (FDA) is applied recently to address this concern, since we are more interested in analyzing minute-by-minute temporal pattern of PA. As it was summarized in Ramsay and Silverman (2005) [149], FDA treats a sequence of observations, such as the daily activity profile curve in our case, as a single unit rather than disjoint minutes spent in varying types of activity. Specifically, our proposed model is based on Functional Principal Component Analysis (FPCA), which is performed on densely-sampled PA data to get the principal directions of variation and achieve dimension reduction. Current studies of FPCA have extended its application in modeling multilevel functional data [33, 164], longitudinal functional data [60] and longitudinal association with scalar outcomes [59, 57], based on corresponding study design or data structure. Our previous works ([193], Wenyi et al. to appear) have implemented these methods on the MENU study. For example, Selene et al. [193] applied a two-level FPCA model and explored the cross-sectional association between extracted principal component scores and health outcomes. Wenyi et al. (to appear) further explored the longitudinal association by means of longitudinal FPCA modeling. However, both of them were not optimal solutions for the PA data have repeated measures at visit-level. The issue was previously addressed by taking the average of the daily records or selecting one single day at each visit, which may cause a loss of information within the corresponding level. Therefore, we propose the multi-level longitudinal FPCA approach by extending the longitudinal FPCA model in Greven et al. (2010) [60] to accommodate the situation where additional levels of inputs were observed in a longitudinal study.

Another research question of interest in this paper is to assess the regression performance when functional predictors are measured more frequently than scalar outcomes. In real analysis, models can be easily misspecified due to lack of background information or the preference to simplified models. In particular, misspecification may happen at the stage of model construction or data preprocessing. As far as we know, there is no related work that has been done to explore the effects of misspecification on model performance, in both cross-sectional and longitudinal studies. In this paper, we provide a comprehensive investigation on the question using both mathematical derivation and simulation studies.

We organize this chapter as follows: Section 5.2 first describes existing methods in the field of FPCA modeling, then provides our proposed multi-level longitudinal FPCA. Section 5.3 illustrates the estimation procedure of our model and its comparison with misspecified models. Section 5.4 shows the performance of our model and existing methods with extensive simulation studies. Section 5.5 presents the application of the multi-level longitudinal FPCA methods to the

MENU study. And Section 5.6 concludes the article with a discussion.

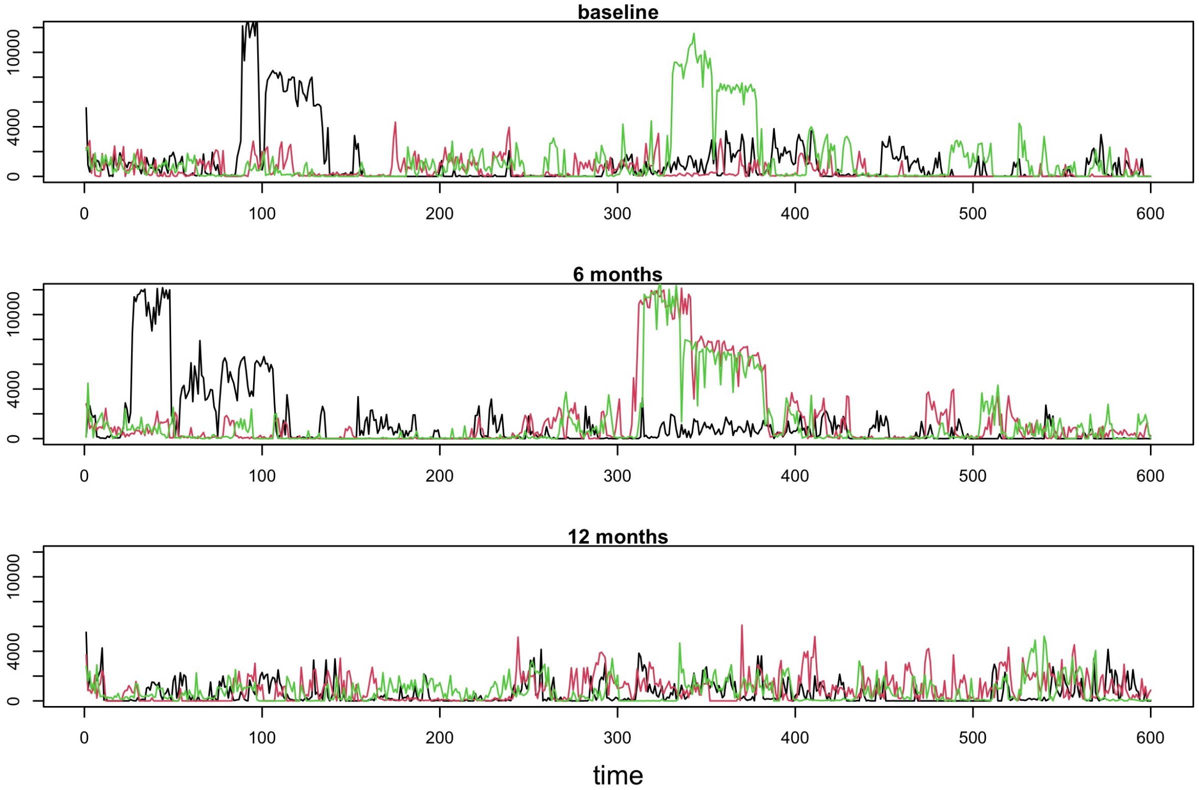


Figure 5.1. An example of daily activity patterns in three days from minute-level accelerometer count data for one subject across three visits.

5.2 Statistical Model

The observed physical activity records are functional data $\boldsymbol{X}_{ijk} = \{X_{ijk}(t), t \in \mathscr{D}\}$, which are random functions in $L^2[0,1]$ measured at minute-level time t on a set of grid points $\mathscr{D}$ with length D, for subject $i = 1,2,\ldots,n$ at visit $j = 1,2,\ldots,n_i$ and day $k = 1,2,\ldots,n_{ij}$. The total number of observations are denoted as $I = \sum_{i,j} n_{ij}$ and the number of observations for subject i is $I_i = \sum_j n_{ij}$. In the following sections, we first summarize previous works in Di et al. (2009)[33], Greven et al. (2010)[60], Shou et al. (2015)[164], and then present our proposed algorithm in the multi-level longitudinal setup.

5.2.1 Overview of FPCA models

We first review existing FPCA models for one-, two- and three-level data, which may also combine longitudinal setting. FPCA plays an important role in functional data analysis, whose basic purpose is to decompose the functional curves into principal directions of variation. For the sake of simplicity, we use $X_i(t)$, $X_{ij}(t)$ and $X_{ijk}(t)$, $t \in \mathscr{D}$ to denote the one-,two- and three-level functional inputs, respectively, where the hierarchical structure of the data can be analogous to subject i, visit j and day k. Assuming no measurement error, in the one-level setting, $X_i(t)$ can be represented as,

$$X_i(t) = \mu(t) + U_i(t). \tag{5.1}$$

$\mu(t)$ is the overall population mean function at t and $U_i(t)$ is the subject-specific deviation from the overall mean function. Specifically, $\mu(t)$ is a fixed function and $U_i(t)$ are assumed to be i.i.d. with mean zero and covariance operator $K_U(s,t) = cov\{U_i(s), U_i(t)\}$. By Mercer's theorem [121], the spectral decomposition is provided as $K_U(s,t) = \sum_{l=1}^{\infty} \lambda_l \phi_l^U(s)\phi_l^U(t)$, where $\lambda_1 \geq \lambda_2 \geq \ldots$ are ordered nonnegative eigenvalues and ϕ_l^U are corresponding orthogonal eigenvectors. Using the Karhunen-Loève (KL) expansion [89], model M1 becomes $X_i(t) = \mu(t) + \sum_{l=1}^{\infty} \xi_{il}\phi_l^U(t)$, where $\xi_{il} = \int U_i(s)\phi_l^U(t)dt$ are uncorrelated principal component scores with mean zero and variance λ_l.

Di et al. (2009) [33] expanded the one-level FPCA model to a two-level FPCA when the data $X_{ij}(t)$ are measured at both subject- and visit-level. The decomposition has the form of

$$X_{ij}(t) = \mu(t) + U_i(t) + V_{ij}(t), \tag{5.2}$$

where $U_i(t)$ is the subject-specific (level 1) deviation from the overall mean function and $V_{ij}(t)$ is the subject- and visit- specific (level 2) deviation from the subject-mean function. It assumed that $U_i(t)$ and $V_{ij}(t)$ are uncorrelated stochastic processes with zero mean and continuous covariance functions. $K_U(s,t) = cov\{U_i(s), U_i(t)\}$ and $K_V(s,t) = cov\{V_{ij}(s), V_{ij}(t)\}$ are covariance operators for the above random processes. Therefore, the variability of $X_{ij}(t)$ is determined by the sum of K_U and K_V, that is, $K_X = K_U + K_V$. With the KL expansion, level 1 and level 2 processes can be decomposed as $X_{ij}(t) = \mu(t) + \sum_{l=1}^{\infty} \xi_{il}\phi_l^U(t) + \sum_{m=1}^{\infty} \zeta_{ijm}\phi_m^V(t)$, where $\phi_l^U(t)$ and $\phi_m^V(t)$ are the eigenfunctions of covariance operators K_U and K_V, respectively. $\xi_{il} = \int U_i(s)\phi_l^U(s)ds$ and $\zeta_{ijm} = \int V_{ij}(s)\phi_m^V(s)ds$ are uncorrelated level 1 and level 2 principal component scores, with mean 0 and variance λ_l and λ_m, respectively. λ_l and λ_m are ordered eigenvalues for every l and m.

Greven et al. (2010) [60] further extended the two-level FPCA to longitudinal FPCA, analogous to a classical longitudinal model but in functional format, which has the form,

$$X_{ij}(t) = \mu(t) + U_{i0}(t) + U_{i1}(t)T_{ij} + V_{ij}(t), \tag{5.3}$$

where $U_{i0}(t)$ is the random functional intercept and $U_{i1}(t)$ is the random functional slope for subject i, respectively, and T_{ij} is the time at visit j for subject i, which can be either the visit indicator with $T_{ij} = j$ or a continuous time variable. One major difference between the longitudinal FPCA and multilevel FPCA is the construction of the subject-specific variation $K_U(s,t)$, which is the covariance operator between the bivariate process $\boldsymbol{U}_i(t) = (U_{i0}(t), U_{i1}(t))$ and has two parts: the auto-covariance $K_{U_0}(s,t)$, $K_{U_1}(s,t)$ and the cross-covariance $K_{U_{01}}(s,t)$. It is represented as:

$$K_U(s,t) = \begin{pmatrix} K_{U_0}(s,t) & K_{U_{01}}(s,t) \\ K_{U_{01}}(t,s) & K_{U_1}(s,t) \end{pmatrix}.$$

Corresponding KL expansion is given as, $X_{ij}(t) = \mu(t) + \sum_{l=1}^{\infty}(1, T_{ij})\xi_{il}\phi_l^U(t) + \sum_{m=1}^{\infty} \zeta_{ijm}\phi_m^v(t)$. Similarly, $\phi_l^U(t) = (\phi_l^{U_0}(t), \phi_l^{U_1}(t))'$ and $\phi_m^V(t)$ are the eigenfunctions of covariance operators K_U and K_V, respectively. $\xi_{il} = \int U_{i0}(s)\phi_l^{U_0}(s)ds + \int U_{i1}(s)\phi_l^{U_1}(s)ds$ and $\zeta_{ijm} = \int V_{ij}(s)\phi_m^V(s)ds$ are uncorrelated level 1 and level 2 principal component scores with mean 0 and variance λ_l and λ_m.

Additional levels of data structure were considered in Shou et al. (2015) [164], referred as structured FPCA. Specifically, with three-level data $\{X_{ijk}(t)\}$, the three-way FPCA model decomposes the data into three parts, subject-specific process $U_i(t)$, visit-specific process $V_{ij}(t)$ and day-specific process $W_{ijk}(t)$, which can be written as,

$$X_{ijk}(t) = \mu(t) + U_i(t) + V_{ij}(t) + W_{ijk}(t), \tag{5.4}$$

where $U_i(t)$ is the subject-specific process, $V_{ij}(t)$ is the subject-visit-specific deviation and $W_{ijk}(t)$ quantifies the daily (level 3) deviation from the the subject- and visit-mean function. $U_i(t)$, $V_{ij}(t)$ and $W_{ijk}(t)$ are mutually uncorrelated random processes with mean zero and covariance operators K_U, K_V and K_W. Using the KL expansion, model 5.3 becomes, $X_{ijk}(t) = \mu(t) + \sum_{l=1}^{\infty} \xi_{il}\phi_l^U(t) + \sum_{m=1}^{\infty} \zeta_{ijm}\phi_m^V(t) + \sum_{r=1}^{\infty} \eta_{ijkr}\phi_r^W(t)$, where $\phi_l^U(t)$, $\phi_m^V(t)$ and $\phi_r^W(t)$ are the eigenfunctions of covariance operators K_U, K_V and K_W, respectively. $\xi_{il} = \int U_i(s)\phi_l^U(s)ds$, $\zeta_{ijm} = \int V_{ij}(s)\phi_m^V(s)ds$ and $\eta_{ijkr} = \int W_{ijk}(s)\phi_r^W(s)ds$ are uncorrelated level 1, level 2 and level 3 principal component scores, with mean zero and variance λ_l, λ_m and λ_r, respectively. Therefore, the variability of $X_{ijk}(t)$ is fully determined by processes $U_i(t)$, $V_{ij}(t)$ and $W_{ijk}(t)$, i.e. $K_X = K_U + K_V + K_W$.

5.2.2 Multi-level Longitudinal FPCA Model

We proposed our multi-level longitudinal model by extending the previously introduced FPCA models, for better accommodating the data structure from our study, that is, a longitudinal study with repeated day-level records, i.e. day k at visit j, for each subject i. Let $\boldsymbol{U}_i(t) = (U_{i0}(t), U_{i1}(t))$, $V_{ij}(t)$ and $W_{ijk}(t)$ be mutually uncorrelated random processes with mean zero as described in section 5.2.1. In our particular implementation with the MENU study, the method can also be referred as a three-level longitudinal FPCA model.

We assume that $U_{i0}(t)$ and $U_{i1}(t)$ have covariance functions $K_{U_0}(s,t)$ and $K_{U_1}(s,t)$, respectively, and cross-covariance function $K_{U_{01}}(s,t)$; $V_{ij}(t)$ has covariance function $K_V(s,t)$ and

$W_{ijk}(t)$ has covariance function $K_W(s,t)$. The model becomes,

$$X_{ijk}(t) = \mu(t) + U_{i0}(t) + U_{i1}(t)T_{ij} + V_{ij}(t) + W_{ijk}(t). \tag{5.5}$$

The model in fact is a natural generalization of the longitudinal FPCA and structured FPCA. Similarly, we provide the KL expansion as,

$$X_{ijk}(t) = \mu(t) + \sum_{l=1}^{\infty}(1, T_{ij})\xi_{il}\phi_l^U(t) + \sum_{m=1}^{\infty}\zeta_{ijm}\phi_m^V(t) + \sum_{r=1}^{\infty}\eta_{ijkr}\phi_r^W(t), \tag{5.6}$$

where $\phi_l^U(t) = (\phi_l^{U_0}(t), \phi_l^{U_1}(t))'$, $\phi_m^V(t)$ and $\phi_r^W(t)$ are the eigenfunctions of covariance operators K_U, K_V and K_W, respectively. $\xi_{il} = \int U_{i0}(s)\phi_l^{U_0}(s)ds + \int U_{i1}(s)\phi_l^{U_1}(s)ds$, $\zeta_{ijm} = \int V_{ij}(s)\phi_m^V(s)ds$ and $\eta_{ijkr} = \int W_{ijk}(s)\varphi_r^W(s)ds$ are uncorrelated random variables with mean zero and variance λ_l^U, λ_m^V and λ_r^W, respectively. Since the infinite expansions is impractical, we consider the case the finite-dimensional approximations of processes $\boldsymbol{U}$, V and W, when most variability of each process is captured by the first N_U, N_V, and N_W principal components,

$$X_{ijk}(t) = \mu(t) + \sum_{l}^{N_U}(1, T_{ij})\xi_{il}\phi_l^U(t) + \sum_{m}^{N_V}\zeta_{ijm}\phi_m^v(t) + \sum_{r}^{N_W}\eta_{ijkr}\phi_r^W(t). \tag{5.7}$$

5.3 Estimation

We assume $X_{ijk}(t)$ are measured on a set of grid points $\mathscr{D}$ with finite length D. Missing data, either in terms of visits or days, can be easily handled with our method. Estimation can be done in the following steps, and more details are provided in the next few sections.

Step 1. Estimating the mean function $\mu(t)$ by the sample average $\hat{\mu}(t) = \frac{1}{I}\sum_{i,j,k} X_{ijk}(t)$. Denote the centered data as $\tilde{X}_{ijk}(t) = X_{ijk}(t) - \hat{\mu}(t)$

Step 2. Estimating the covariance function $\hat{K}_W$ from $\tilde{X}_{ijk}(t)$ via method of moment (MoM) estimators (Koch 1968, Shou et al. 2015).

Step 3. Estimating the covariance functions $\hat{K}_U$ for $\boldsymbol{U}_i = (U_{i0}, U_{i1})$ and $\hat{K}_V$ for V_{ij} via mixed linear regression model.

Step 4. Performing eigen decompositions of the estimated covariance functions to provide bases for representing $\boldsymbol{U}_i = (U_{i0}, U_{i1})$, V_{ij} and W_{ijk}.

Step 5. Estimating the best linear unbiased prediction (BLUP) to provide subject-, visit- and day-specific principal component scores.

5.3.1 Estimation of the mean and covariance operators

The fixed population mean function $\mu(t)$ is estimated by taking the sample mean in our implementation. When the observations across visits and subjects are relatively dense, a bivariate smoother in s and T may be considered to form the mean surface $\mu(s,T)$, such as penalized splines smoothers (Greven et al. 2010). Similarly, as for sparser collection of T_{ij}, $\mu(t)$ can be approximated via the univariate smoother $\mu_j(t)$. With the estimated mean function $\hat{\mu}(t)$ from any of the optional methods, data are centered via $X_{ijk}(t) - \hat{\mu}(t)$ and without loss of generality, we assume that $X_{ijk}(t)$ has mean zero.

The main challenge of our proposed method is to estimate the covariance operators $K_U = \begin{pmatrix} K_{U_0} & K_{U_{01}} \\ K_{U_{01}} & K_{U_1} \end{pmatrix}$, K_V and K_W. Under the setup and assumptions of model 5.5, for all i, j, j', k, k', s, t, we have

$$
\begin{aligned}
Cov(X_{ijk}(s), X_{ij'k'}(t)) = E(X_{ijk}(s)X_{ij'k'}(t))& \\
= Cov(U_{i0}(s), U_{i0}(t)) + T_{ij}Cov(U_{i0}(s), U_{i1}(t)) &+ T_{ij'}Cov(U_{i0}(t), U_{i1}(s)) \\
+ T_{ij}T_{ij'}Cov(U_{i1}(s), U_{i1}(t)) + Cov(V_{ij}(s), V_{ij'}(t)) &+ Cov(W_{ijk}(s), W_{ij'k'}(t)).
\end{aligned}
\tag{5.8}
$$

The estimation of these covariance operators is not straightforward, and we cannot simply apply the method of Greven et al.(2010) [60], that is, linearly regressing the left side ”outcome”

$X_{ijk}(s)X_{ij'k'}(t)$ on the right side "covariates". This is because the total number of the observations $I = \sum_{i,j} n_{ij}$ in the three-level model can be much larger than that in a two-level setup, and hence it is impractical and computationally inefficient to fit a regression model at once. Therefore, we proposed a two-step procedure to estimate these covariance estimators, combining the method of moment (MoM) estimators and regression strategy. Let δ denote the Kronecker's delta defined as $\delta_{ii'} = \begin{cases} 1, & \text{if } i = i' \\ 0, & \text{otherwise} \end{cases}$, Equation 5.8 can be rewritten as,

$$E(X_{ijk}(s)X_{ij'k'}(t)) = \begin{cases} K_{U_0}(s,t) + 2T_{ij}K_{U_{01}}(s,t) + T_{ij}^2 K_{U_1}(s,t) + K_V(s,t) + \delta_{kk'}K_W(s,t), \\ \qquad \text{if} j = j' \\ K_{U_0}(s,t) + T_{ij}K_{U_{01}}(s,t) + T_{ij'}K_{U_{01}}(t,s) + T_{ij}T_{ij'}K_{U_1}(s,t), \text{otherwise.} \end{cases} \tag{5.9}$$

We first estimate the day-level covariance operator K_W using a MoM estimator,

$$\hat{K}_W = \frac{1}{\sum_{i,j} n_{ij}(n_{ij}-1)} \sum_{i,j} \sum_{k,k'} (\boldsymbol{X}_{ijk} - \boldsymbol{X}_{ijk'})(\boldsymbol{X}_{ijk} - \boldsymbol{X}_{ijk'})^T. \tag{5.10}$$

Substituting K_W with the empirical estimator $\hat{K}_W$ in both sides of the first line of Equation 5.9, we eliminate the day-to-day variation from the total variation. The remaining proportion of variation therefore only involves variations at subject and visit levels. Denote $\widetilde{\boldsymbol{X}_{ij.}\boldsymbol{X}_{ij'.}^T}$ as the resulting residual variance, the estimators of the covariance operators of the U and V processes can be expressed as,

$$E(\widetilde{\boldsymbol{X}_{ij.}\boldsymbol{X}_{ij'.}^T}) = K_{U_0} + T_{ij}K_{U_{01}} + T_{ij'}K_{U_{01}} + T_{ij}T_{ij'}K_{U_1} + \delta_{jj'}K_V. \tag{5.11}$$

By implementing the two-step procedure, we are able to reduce the dimension of the 'outcome' variable in the regression modeling. The computational feasibility for the 2-level case has been proved in Greven et al.(2010), and we then regress the product $\widetilde{\boldsymbol{X}_{ij.}\boldsymbol{X}_{ij'.}^T}$ on "predictors" $(1, T_{ij}, T_{ij'}, T_{ij}T_{ij'}, \delta_{jj'})$ and get the estimated $(\hat{K}_{U_0}, \hat{K}_{U_{01}}, \hat{K}_{U_1}, \hat{K}_V)$.

5.3.2 Estimation of eigenfunctions and scores

With the estimated covariance operators from the two-step procedure, $\hat{K}_U = \begin{pmatrix} \hat{K}_{U_0} & \hat{K}_{U_{01}} \\ \hat{K}_{U_{01}} & \hat{K}_{U_1} \end{pmatrix}$, $\hat{K}_V$ and $\hat{K}_W$ in the previous section, using the spectral decomposition, we can easily estimate the eigenvalues λ_l^U, λ_m^V, and λ_r^W, and eigenfunctions $\phi_l^U(t)$, $\phi_m^V(t)$, and $\phi_l^W(t)$, $t \in \mathscr{D}$, at grid points D, that is, $\hat{K}_U = \sum_{l=1}^{2D} \hat{\lambda}_l \phi_l^U (\phi_l^U)^T$, $\hat{K}_V = \sum_{m=1}^{D} \hat{\lambda}_m \hat{\phi}_r^V (\hat{\phi}_r^V)^T$ and $\hat{K}_W = \sum_{r=1}^{D} \hat{\lambda}_r \hat{\phi}_r^W (\hat{\phi}_r^W)^T$. The eigenfunctions, $\phi_l^U = \{\phi_l^{U_0}(t), \phi_l^{U_1}(t), t \in \mathscr{D}\}$ are orthogonal vectors in IR^{2D}, and ϕ_m^V and ϕ_r^W are orthogonal vectors in IR^D. If the time variable T_{ij} is standardized to have zero mean and unit variance, i,e, $E(T_{ij}) = 0$ and $Var(T_{ij}) = 1$, the variation in $X_{ij}(t)$ can be decomposed additively and expressed with respect to the estimated eigenvalue,

$$\int_{\mathscr{D}} var(X_{ijk}(t))dt = \sum_l \lambda_l^U + \sum_m \lambda_m^V + \sum_r \lambda_r^W. \tag{5.12}$$

It has been proved in a two-level longitudinal FPCA setting in Greven et al.(2010) and we extend its validity to three-level scenario (Appendix B.1). We usually retain finite numbers of eigenfunctions of subject (N_U), visit (N_V) and day (N_W) levels for further analysis, which is based on a pre-specified percentage of explained variation.

For fixed N_U, N_V and N_W, it is evident that model 5.7 is a three-level linear mixed model. Therefore, the principal component scores $\hat{\xi}_{il}$, $\hat{\zeta}_{ijm}$ and $\hat{\eta}_{ijkr}$ can be obtained via the best linear unbiased prediction (BLUP). Let $\boldsymbol{X}_i = vec\{\boldsymbol{X}_{i11}, \ldots, \boldsymbol{X}_{i1n_{i1}}, \ldots, \boldsymbol{X}_{ij1}, \ldots \boldsymbol{X}_{ijn_{ij}}\}$ be a vector with stacked functional inputs for subject i and has length of $D \times I_i$ and $\boldsymbol{\beta_i} = (\xi_{i1}, \ldots, \xi_{iN_U}, \ldots, \zeta_{i11}, \ldots, \zeta_{i1N_V}, \ldots, \zeta_{in_i1}, \ldots, \hat{\zeta}_{in_iN_V}, \eta_{i111}, \ldots, \eta_{i11N_W}, \ldots, \eta_{i1n_in_{ij}1}, \ldots, \eta_{in_in_{ij}N_W})$ be the vector of scores to be estimated. The BLUP for $\boldsymbol{\beta_i}$ is given as,

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{i}} = (\mathbf{Z}_i'\mathbf{Z}_i)^{-1}\mathbf{Z}_i'\boldsymbol{X}_{\boldsymbol{i}}, \tag{5.13}$$

where $\boldsymbol{Z}_i = [\mathbf{1}_{I_i} \otimes \mathbf{\Phi}^{U_0} + \boldsymbol{T}_i \otimes \mathbf{\Phi}^{U_1} | \boldsymbol{I}_{n_i} \otimes (\mathbf{1}_{n_{ij}} \otimes \mathbf{\Phi}^V) | \boldsymbol{I}_{I_i} \otimes \mathbf{\Phi}^W]$, $\boldsymbol{T}_i = (T_{ij}\delta_{jh})_{j=1,\ldots,n_i}; h = 1, \ldots, n_i$, $\mathbf{\Phi}^{U_0} = \{\phi_l^{U_0}(t)\}_{t \in \mathscr{D}, l=1,\ldots,N_U}$, $\mathbf{\Phi}^{U_1} = \{\phi_l^{U_1}(t)\}_{t \in \mathscr{D}, l=1,\ldots,N_U}$, $\mathbf{\Phi}^V = \{\phi_l^V(t)\}_{t \in \mathscr{D}, l=1,\ldots,N_V}$,

$\mathbf{\Phi}^W = \{\phi_l^W(t)\}_{t\in\mathscr{D},l=1,\ldots,N_W}$, $\boldsymbol{I}$ is the diagonal matrix with element 1, and $\otimes$ denotes the Kronecker product of matrices. $(\delta_{jh})_{j=1,\ldots,n_i}; h = 1,\ldots,n_i$ denotes the indicator matrix with entries δ_{jh} at row $j, i = 1,\ldots,N, j = 1,\ldots,n_i$ and column $h, h = 1,\ldots,n_i$, with $\delta_{jh} = 1$ if $j = h$ and $\delta_{jh} = 0$ otherwise.

5.3.3 Comparing different models

The proposed three-level longitudinal FPCA model was motivated by the MENU study [97, 154], which is a longitudinal study in which the functional measurements were collected daily for each visit for each participant, thus encompassing a three-level nested structure. To further illustrate the need for deriving our method, we explored the relationship between our three-level method and previous approaches with the aim of gaining deeper insights. For instance, with the three-level longitudinal data, one may consider ignoring the random functional slope process and applying three-level (structured) FPCA [164]. Alternatively, assuming that data are well aligned, the two-level longitudinal FPCA provided in Greven et al.(2010) [60], can be fitted on the mean values of daily measurement curves at each visit. While it is intuitive that either of these simplifications can cause a loss of information, we wanted to demonstrate their concrete effects.

Suppose that we fit the data with three-level FPCA model, and decompose the total variance with the idea of symmetric sum MoM estimators from Shou et al. (2015) [164]. Rewriting the covariance operators as $E\{X_{ijk}(s) - X_{i'j'k'}(s)\}\{X_{ijk}(t) - X_{i'j'k'}(t)\}'$, gives the following decomposed form,

$$\begin{cases} 2K_W(s,t), & \text{if } i = i', j = j', k \neq k' \\ 2(T_{ij}K_{U_{01}}(s,t) + T_{ij'}K_{U_{01}}(t,s) + T_{ij}T_{ij'}K_{U_1}(s,t) + K_V(s,t) + K_W(s,t)), & \text{if } i = i', j \neq j' \\ 2(K_{U_0}(s,t) + T_{ij}K_{U_{01}}(s,t) + T_{ij'}K_{U_{01}}(t,s) + T_{ij}T_{ij'}K_{U_1}(s,t) + K_V(s,t) + K_W(s,t)), & \text{if } i \neq i' \end{cases} \tag{5.14}$$

Let $K_{V_U}(s,t) = T_{ij}K_{U_{01}}(s,t) + T_{ij'}K_{U_{01}}(t,s) + T_{ij}T_{ij'}K_{U_1}(s,t) + K_V(s,t)$, which combines variation from random slope auto-covariance K_{U_1}, cross-covariance $K_{U_{01}}$ and subject-visit specific covariance K_V. As a result, the total covariance will be only decomposed into three parts, the K_{U_0}, K_{V_U} and K_W. Therefore, if the model is misspecified as a three-level FPCA model which ignores the slope process, it is expected to witness an inflation of variation at visit level, while the rest of the variation at subject and day level will not be changed. Given that the total amount of variation is a fixed number, the proportions explained by subject will be underestimated.

On the other hand, if we take the mean of the observed curves at visit level, that is, let $\bar{X}_{ij} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} X_{ijk}$, we then have,

$$\begin{aligned} \bar{X}_{ij}(t) &= \mu(t) + U_{i0}(t) + U_{i1}(t)T_{ij} + V_{ij}(t) + \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} W_{ijk}(t) \\ Var(\bar{\boldsymbol{X}}_{ij}) &= K_{U_0} + T_{ij}K_{U_{01}} + T_{ij}K_{U_{01}} + T_{ij}^2 K_{U_1} + K_V + \frac{1}{n_{ij}^2} \sum_{k,k'} Cov(W_{ijk}, W_{ijk'}). \end{aligned} \tag{5.15}$$

The total variation is increased with an additional part $\frac{1}{n_{ij}^2} \sum_{k,k'} Cov(W_{ijk}, W_{ijk'})$, compared with Equation 5.11. As a result, the estimated explained variance within both K_U and K_V are increased by an approximately same amount, and correspondingly, the proportions of the respective explained variations. However, the relative variation, i.e. the ratios between eigenvalues of $\hat{K}_U$ and $\hat{K}_V$ are fixed. In fact, taking the average can be considered as a form of smoothing, especially when the processes $\hat{W}_{ijk}$ are random errors. Therefore, a two-level model may be applicable if the day-to-day variation in the data is ignorable or the major objective is more focused on subject and visit levels.

Both comparisons with three-level FPCA and longitudinal FPCA are further demonstrated with our simulation studies in Section 5.4.

5.3.4 Regression model

Traditional multivariate linear models are extended to scalar-on-function regression models to explore the associations between scalar outcomes and functional predictors. With normally-distributed outcomes, a functional principal components regression (FPCR) model [150] with functional predictor $U_i(t)$ has the form,

$$E(Y_i) = \alpha_0 + \int_{t \in \mathscr{D}} U_i(t)\beta_U(t)dt, \tag{5.16}$$

where $Y_i \in \mathscr{R}$ is a scalar outcome and α_0 is the regression intercept. The subject-level $U_i(t)$ can be reconstructed from any FPCA models introduced in Section 5.2 and $\beta_U(t)$ is the corresponding functional regression coefficient. Gertheiss et al. (2013) [57] extended the method to a longitudinal setup, where the outcome Y_{ij} was recorded for each subject i at visit j and includes both subject-level functional predictors $U_i(t)$ and visit-level functional predictors $V_{ij}(t)$. The longitudinal FPCR model then has the form,

$$E(Y_{ij}) = \alpha_0 + b_i + \int \beta_U(t)U_i(t)dt + \int \beta_V(t)V_{ij}(t)dt, \tag{5.17}$$

where b_i is a subject-specific random effect. We assume $b_i \sim N(0, \tau^2)$ and it is conditionally independent Y_{ij}. The $\beta_U(t)$ and $\beta_V(t)$ are smooth coefficient functions for processes $U_i(t)$ and $V_{ij}(t)$, respectively. Both standard FPCR and longitudinal FPCR yield smooth coefficient functions, which have nice interpretation over time and don't depend on the number of principal components selected [57].

These functional coefficients in Equation 5.16 and 5.17 can be estimated using penalized spline methods via the R package `mgcv` [189, 190].

5.4 Simulation Study

In this section, simulation studies were implemented to explore the properties of the methods provided in Section 5.2. In addition to testing the robustness of our proposed methods, another goal is to explore how these methods perform when the model is misspecified under varying simulation settings. Specifically, we performed simulation studies in both unbalanced cross-sectional (one-level outcome and two-level predictor) and longitudinal (two-level outcome and three-level predictor) setups. Our motivation for considering these two setups is to explicitly evaluate the impact of ignoring the multilevel structure of the data versus ignoring the longitudinal structure. For instance, taking the average at visit level of a two-level predictor could result in simultaneous loss of information in both multilevel and longitudinal structure. However, for a three-level predictor, the averaging process performed on day-level measures still retains the longitudinal structure of the input data and is expected to lose the three-level structure. We perform a series of simulations studies to validate these assumptions in cross-sectional and longitudinal setups.

For both simulation studies, we compared the performance in both functional models and regression models. The normalized errors between the estimated and true eigenvalues and principal component scores were used as the evaluation criteria for functional modeling. As for regression results, we computed the observed mean squared errors (MSE). For each simulation setting, we generated 100 replicates with $N = 100$ subjects. The corresponding R code for our proposed method and other models used in simulation studies is available at https://github.com/wendylin23/MixedFPCA.

5.4.1 Study 1: Two-level functional inputs and one-level scalar outcomes

We assumed a balanced design with $n_i = 3$ visits for each subject and the time variable T_{ij} is generated by standardizing the visits, i.e. $T_{ij} = \frac{j-\frac{1}{n_i}\sum j}{var(j)}$, to have unit variance. The functional curves $X_{ij}(t)$ with length of $D = 600$ were generated according to the two-level longitudinal FPCA model 5.2 and the true model was set as,

$$\begin{cases} y_i & = \int \beta_U(t)U_i(t)dt + \varepsilon_\sigma, \\ X_{ij}(t) & = \sum_{l=1}^{N_U} \xi_{il}\phi_l^{(U_0)}(t) + \sum_{l=1}^{N_U} T_{ij}\xi_{il}\phi_l^{(U_1)}(t) + \sum_{m=1}^{N_V} \zeta_{ijm}\phi_m^{(V)}(t), , \\ \xi_{il} & \overset{\text{i.i.d.}}{\sim} \mathscr{N}(0.\lambda_l^U), \quad \zeta_{ijm} \overset{\text{i.i.d.}}{\sim} \mathscr{N}(0.\lambda_m^V), \quad t \in \mathscr{D} \end{cases} \tag{5.18}$$

where the number of eigenfunctions was set as $N_U = N_V = 4$. The eigenfunctions bases can be orthonormal sine/cosine basis (F-basis) and Legendre polynomials basis (L-basis). It is noted that the sine/cosine basis is orthogonal with each other but it is correlated with the Legendre polynomials basis. In addition, we also considered the cases where only smoothed random error curves were added to the subject-level slope and visit-level curves, to mimic the situations when the between-visit variability is small. The random error curves (E-basis) were generated via $e_{ij} \sim N(0, \sigma_e^2)$, $\sigma_e = 0.3$, which was used to replace the random slope or visit-specific term in Equation 5.18. In the following sections, we used abbreviations to represent the combination of different basis. For instance, ”FFF” refers to the combination of all three orthogonal Fourier basis in U_0, U_1 and V processes. In this part, we simulated data from five combinations, (a) FFF (b) FFL (c) FLF (d) FFE (e) FEF. Meanwhile, we set the eigenvalues to be $\lambda_l^U = \lambda_l^V = 0.5^{l-1}, l = 1,2,3,4$.

The scalar outcomes y_i in the regression models were assumed to be normally distributed with variance $\sigma^2 = 2$. Following the regression simulation settings in Gertheiss et al. (2013) [57], we considered a nonlinear function having the shape of a Gamma-density for the true coefficient function $\beta_U(t)$.

For each of the 100 simulated datasets, we implemented the two-level longitudinal FPCA and two-level FPCA on the two-level simulated functional data $X_{ij}(t)$, and a simple one-level FPCA model on the visit-average functional inputs $\overline{X}_{i.}(t) = \frac{1}{n_i}\sum_j X_{ij}(t)$. We estimated the eigenfunctions, eigenvalues, scores and predicted functional trajectories $\hat{U}_i(t)$. To assess the performance of functional modeling, we computed the normalized errors between the estimated and true eigenvalues $(\hat{\lambda}_l^U - \lambda_l^U)/\lambda_l^U$, and scores $(\hat{\xi}_{il} - \xi_{il})/\sqrt{\lambda_l^U}$ at subject-level. With the simulated scalar outcomes y_i, we then computed the MSE from the regression fitting $\frac{1}{M}\sum_i (y_i - \hat{y}_i)^2$ as the assessment for regression modeling.

Figure 5.2 presents results of normalized errors between the estimated and true eigenvalues for subject-level process $(\hat{\lambda}_l^U - \lambda_l^U)/\lambda_l^U$ based on 100 replicates from the five simulation scenarios. Eigenvalues estimates are generally unbiased if the model is correctly fitted with a two-level longitudinal FPCA, although scenarios with correlated basis, i.e., FFL and FLF show slight bias, which is not evident for FFF. If the model was misspecified, compared with results from one-level FPCA modeling, the two-level FPCA models yield less biased results. Meanwhile, in the last two scenarios, FFE and FEF, estimates from both two-level FPCA models are comparable, which is perhaps unsurprising since in these scenarios, the random processes at slope or visit-level were replaced with error terms.

Figure 5.3 displays results of the normalized errors between the estimated and true scores for subject-level process $(\hat{\xi}_{il} - \xi_{il})/\sqrt{\lambda_l^U}$. All three models provide unbiased estimates of the level 1 principal component scores, while the correctly-specified two-level longitudinal model always has the least variance.

Figure 5.4 shows the MSE results $\frac{1}{M}\sum_i (y_i - \hat{y}_i)^2$ from the regression fitting. Two-level longitudinal FPCA models consistently have the least MSE compared to the other two methods in all scenarios. It is noted when we replace the visit-specific processes with random error terms, i.e., the FFE scenario, the misspecified one-level FPCA model works similarly well as the true model, which is not surprising since in this case, one-level FPCA is simply averaging out the visit-level ”noise”. However, one-level FPCA provides biased estimations in all other scenarios. Fitting a two-level FPCA model was generally more robust to the misspecification, and in general had better performance than one-level FPCA models, except in scenario (c) with FLF.

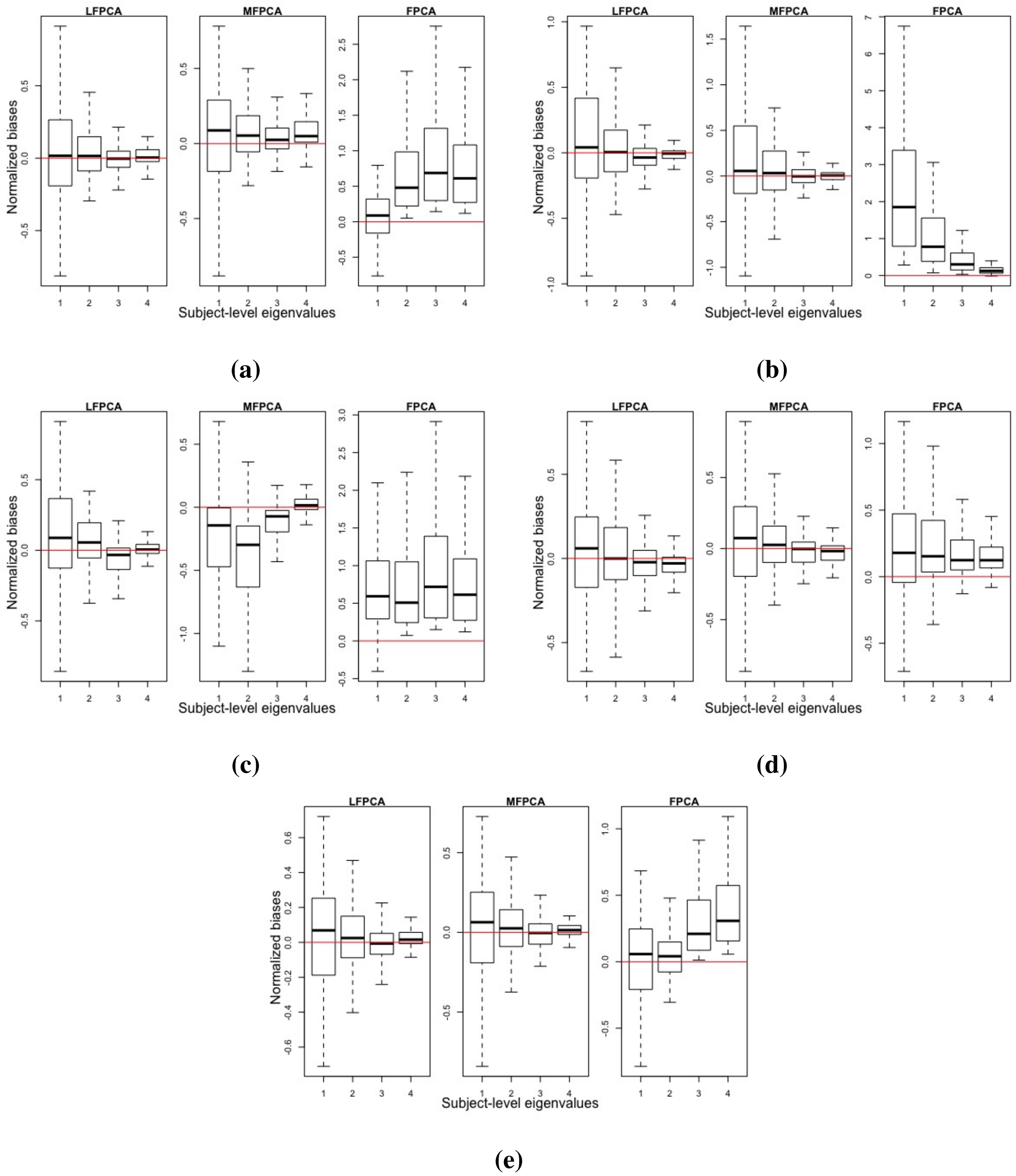
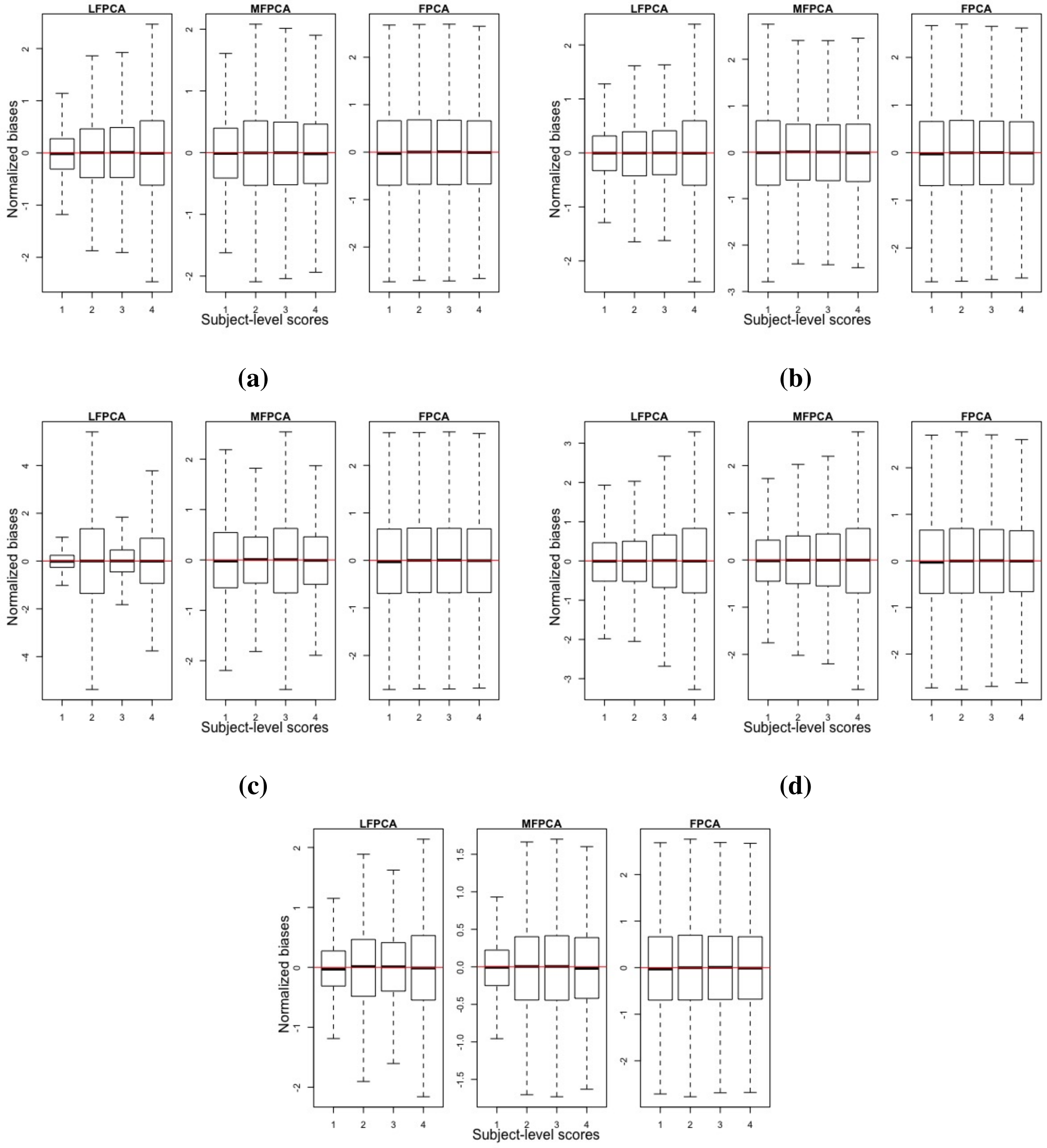


Figure 5.2. Boxplots of the normalized errors between the estimated and true eigenvalues for subject-level process $(\hat{\lambda}_l^U - \lambda_l^U)/\lambda_l^U$ based on 100 replicates, comparing the two-level models with the one-level model. Red line represents the zero. The simulation scenario includes (a) FFF (b) FFL (c) FLF (d) FFE (e) FEF.



(a)

(b)

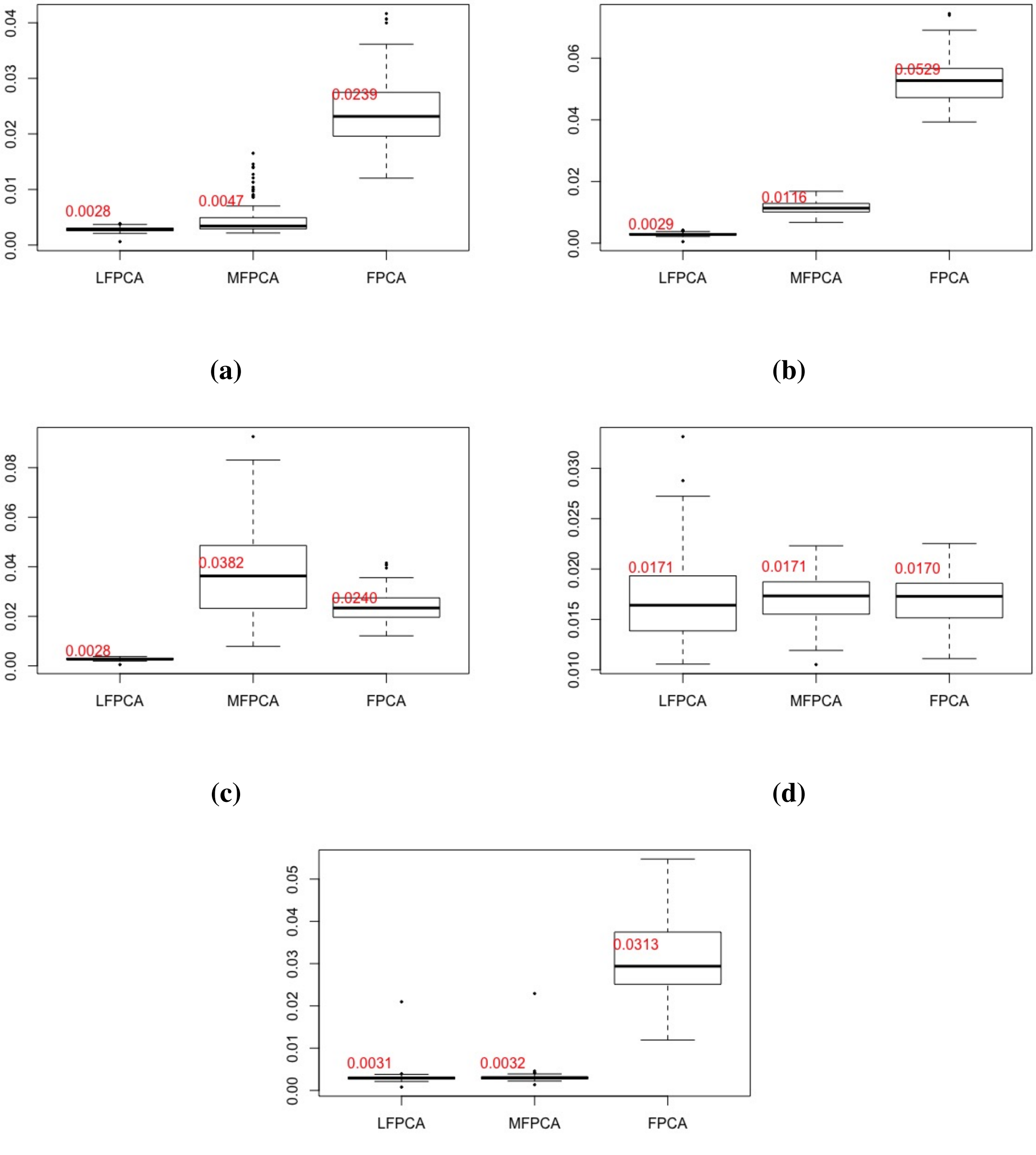
(c)

(d)

(e)

Figure 5.3. Boxplots of the normalized errors between the estimated and true scores for subject-level process $(\hat{\xi}_{il} - \xi_{il})/\sqrt{\lambda_l^U}$ based on 100 replicates, comparing the two-level models with the one-level model. Red line represents the zero. The simulation scenario includes (a) FFF (b) FFL (c) FLF (d) FFE (e) FEF.

5.4.2 Study 2: Three-level functional inputs and two-level scalar outcomes



(a) **(b)**

(c) **(d)**

(e)

Figure 5.4. Boxplots of the MSE $\frac{1}{M}\sum_i (y_i - \hat{y}_i)^2$ from the regression fitting based on 100 replicates, comparing the two-level models with the one-level model. The simulation scenario includes (a) FFF (b) FFL (c) FLF (d) FFE (e) FEF.

In this section, the simulation studies were extended to three-level settings and we assumed a fixed numbers of visits $n_i = 3$ and days $n_{ij} = 3$ for each subject. The functional curves $X_{ijk}(t)$ with length of $D = 600$ were generated according to the three-level longitudinal FPCA model 5.4 and the true model was set as,

$$\begin{cases} y_{ij} & = b_i + \int \beta_U(t)U_i(t)dt + \int \beta_V(t)V_{ij}(t)dt + \varepsilon_\sigma, \\ X_{ijk}(t) & = \sum_{l=1}^{N_U} \xi_{il}\phi_l^{(U_0)}(t) + \sum_{l=1}^{N_U} T_{ij}\xi_{il}\phi_l^{(U_1)}(t) + \sum_{m=1}^{N_V} \zeta_{ijm}\phi_m^{(V)}(t) + \sum_{r=1}^{N_W} \eta_{ijkr}\phi_r^W(t), , \\ \xi_{il} & \overset{\text{i.i.d.}}{\sim} \mathscr{N}(0.\lambda_l^U), \quad \zeta_{ijm} \overset{\text{i.i.d.}}{\sim} \mathscr{N}(0.\lambda_m^V), \quad \eta_{ijkr} \overset{\text{i.i.d.}}{\sim} \mathscr{N}(0.\lambda_r^W), \quad t \in \mathscr{D} \end{cases} \tag{5.19}$$

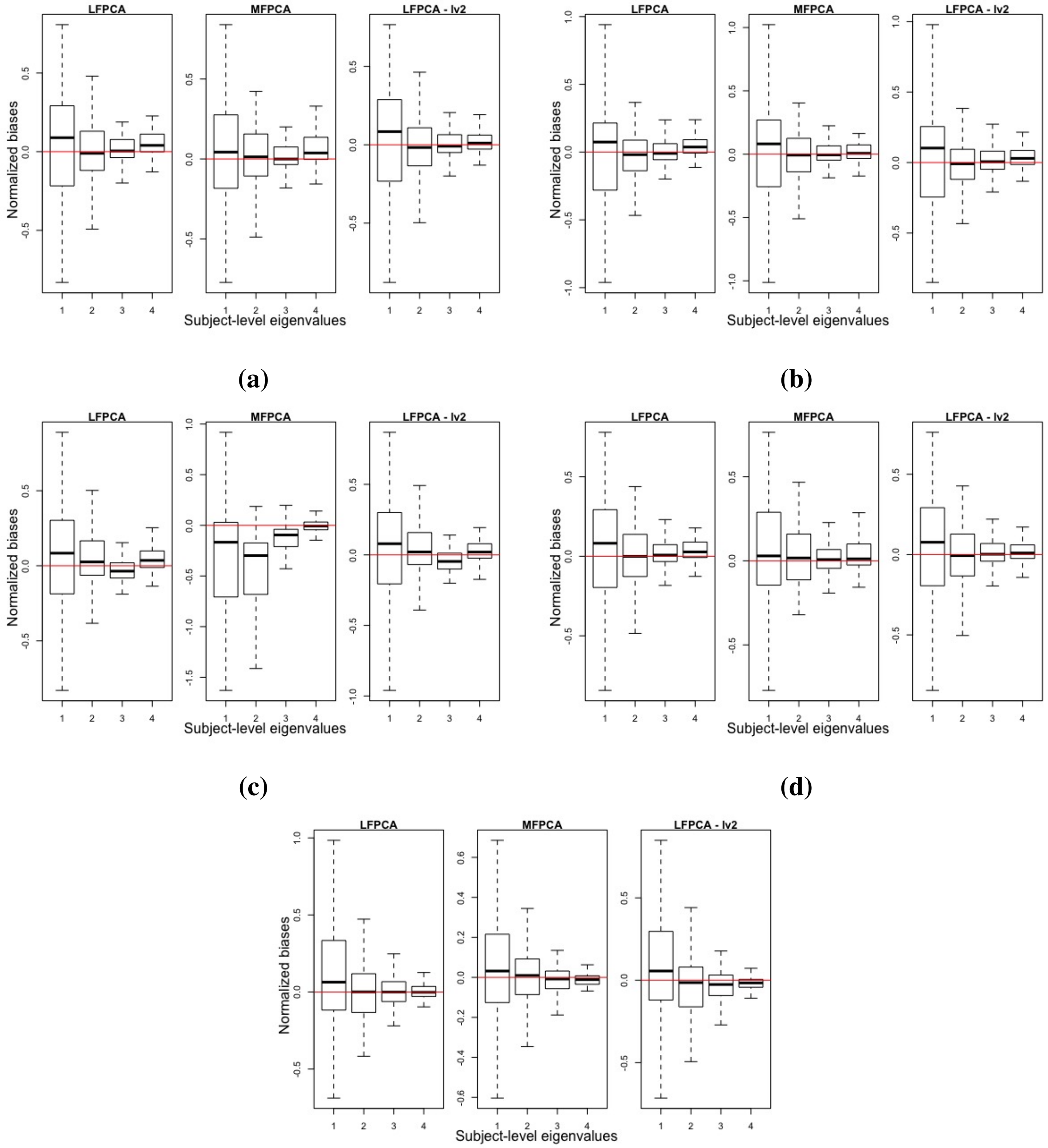
where the number of eigenfunctions is set as $N_U = N_V = N_W = 4$. Similar to the previous simulations, the eigenfunctions bases were from orthonormal sine/cosine basis, Legendre polynomials basis or the smoothed random error curves. These error curves were generated via $e_{ijk} \sim N(0, \sigma_e^2)$, $\sigma_e = 0.3$. In this study, we simulated data based on five types of basis combinations, including (a) FFFF (b) FFFL (c) FLFF (d) FFFE (e) FEFF. Similarly, we set the eigenvalues to be $\lambda_l^U = \lambda_l^V = \lambda_l^W = 0.5^{l-1}, l = 1,2,3,4$.

The two-level scalar outcomes y_{ij} in the regression models were assumed to be normal with variance $\sigma^2 = 2$. b_i is a random intercept process and follows $b_i \sim N(0,1)$. We also used Gamma-density to simulate the true coefficient functions $\beta_U(t)$ and $\beta_V(t)$. For each of the 100 simulated datasets, we implemented the three-level longitudinal FPCA and three-level FPCA on the three-level simulated functional data $X_{ijk}(t)$, and the two-level longitudinal FPCA model on the day-average functional inputs $\overline{X}_{ij.}(t) = \frac{1}{n_{ij}} \sum_j X_{ijk}(t)$. The eigenfunctions, eigenvalues, scores and predicted functional trajectories $\hat{U}_i(t)$ and $\hat{V}_{ij}(t)$ were estimated from each model. Furthermore, we considered the case where varying amounts of variation are explained in the visit- and day-level and assumed that the true eigenvalues can vary among levels. Additional tests, such as unbalanced design (missing visits), were also implemented.

We include the results using equal eigenvalues at each level, assuming no missing data. In Figure 5.5 and Figure 5.6, we show the results of normalized errors between the estimated and true eigenvalues for subject-level process $(\hat{\lambda}_l^U - \lambda_l^U)/\lambda_l^U$ and visit-level process $(\hat{\lambda}_m^V - \lambda_m^V)/\lambda_m^V$, respectively, based on 100 replicates from five simulation scenarios. At subject level, all three methods provide similarly unbiased estimates of the eigenvalues. But at visit level, only the proposed three-level longitudinal model can unbiasedly estimate the eigenvalues in all five scenarios. Meanwhile, the two-level longitudinal model performs consistently better than the three-level FPCA model, except for the last scenario, where only random error terms are added to the day-level. This finding conforms with our theoretical derivation in Section 5.3.3, where we proved that when the model fitting ignores the random slope process and is misspecified as a three-level FPCA model, estimated variation at visit level is inflated and the consequences are reflected by these overestimated visit-level eigenvalues. On the contrary, when the two-level longitudinal FPCA is fitted to the day-averaged data, because the relative explained variation at the subject and visit levels are not as affected, the estimation biases are hence much less.

Finally, as an interesting parenthetical remark, the last two scenarios can be considered as special cases in model misspecification, and can also guide decisions on when our proposed three-level model is most needed. Specifically, if the proportion of explained variation is small for the random slope processes or day-specific processes, the three-level longitudinal model essentially reduces to a three-level FPCA model (i.e., ignoring the slope process) or a two-level longitudinal model (i.e., averaging over days). Table 5.1 provides the percentages of explained variance by different levels of the first two principal components, comparing results from three fitted model with the true setting of the first simulation scenario. It further validates our derivation in Equation 5.14 and 5.15, showing the impact of misspecification at different levels.

Figure 5.7 and Figure 5.8 display results of the normalized errors between the estimated and true scores for subject-level process $(\hat{\xi}_{il} - \xi_{il})/\sqrt{\lambda_l^U}$ and visit-level process $(\hat{\zeta}_{ijm} - \zeta_{ijm})/\sqrt{\lambda_m^V}$, respectively. The scores are unbiasedly estimated and the three-level longitudinal models have the least variance, which is similar as we seen in Study 1. Figure 5.9 presents MSE results $\frac{1}{M}\sum_{i,j}(y_{ij} - \hat{y}_{ij})^2$ from the regression fitting and the three-level longitudinal models always have the best prediction performance. Compared with three-level FPCA models, the two-level longitudinal FPCA models still perform better in the first four scenarios. Combined with the similar findings in eigenvalues, we conclude that in real analysis, misspecifying a model with the form of a two-level longitudinal structure may be a more acceptable than misspecifying it as three-level FPCA models. However, it is important to note that all simulated data in our study are well-aligned, which may also increase the relative robustness of the averaging procedure.



(a)

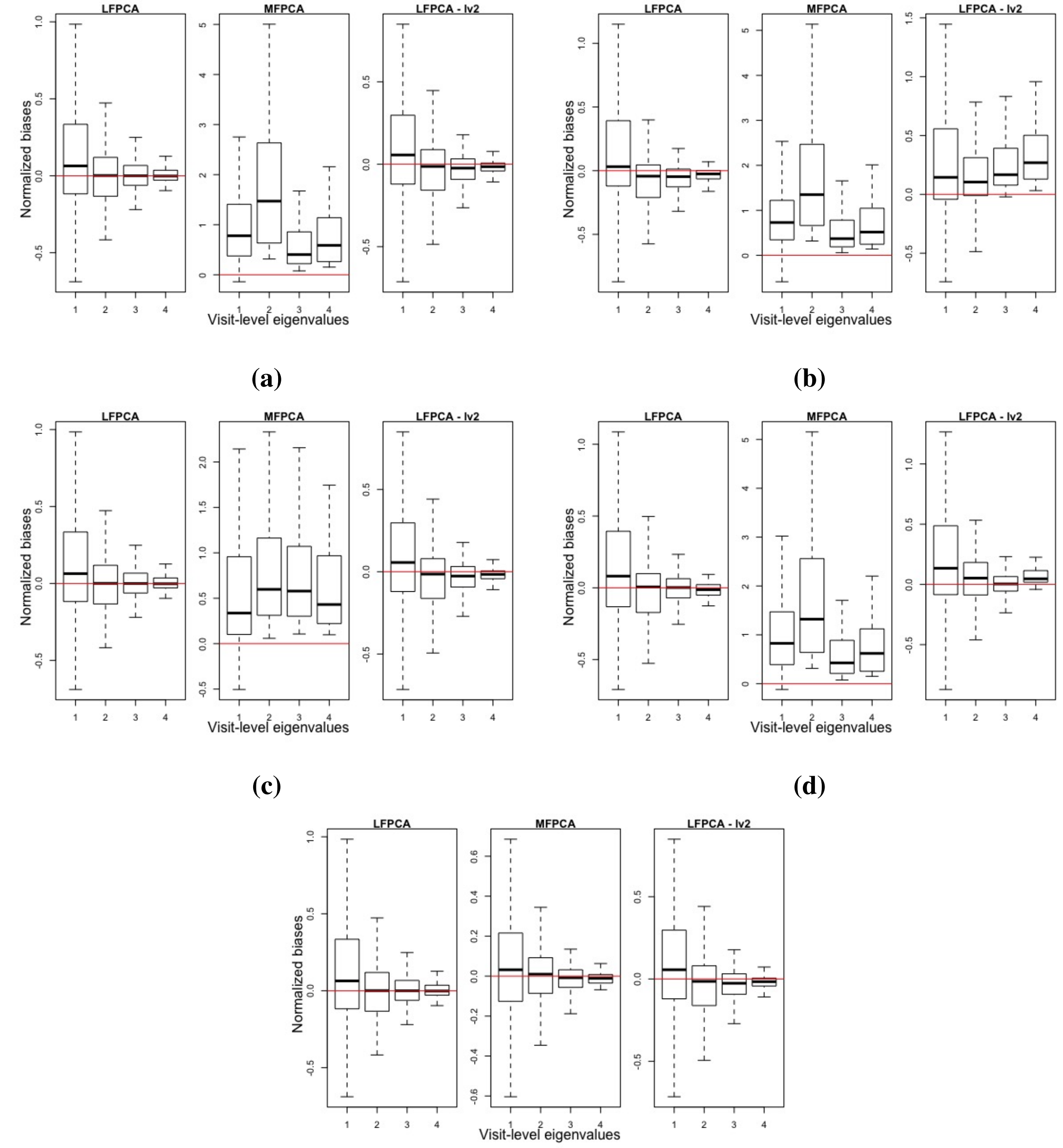
(b)

(c)

(d)

(e)

Figure 5.5. Boxplots of the normalized errors between the estimated and true eigenvalues for subject-level process $(\hat{\lambda}_l^U - \lambda_l^U)/\lambda_l^U$ based on 100 replicates, comparing the three-level models with the two-level model. Red line represents the zero. The simulation scenario includes (a) FFFF (b) FFFL (c) FLFF (d) FFFE (e) FEFF.



(a)

(b)

(c)

(d)

(e)

Figure 5.6. Boxplots of the normalized errors between the estimated and true eigenvalues for visit-level process $(\hat{\lambda}_m^V - \lambda_m^V)/\lambda_m^V$ based on 100 replicates, comparing the three-level models with the two-level model. Red line represents the zero. The simulation scenario includes (a) FFFF (b) FFFL (c) FLFF (d) FFFE (e) FEFF.

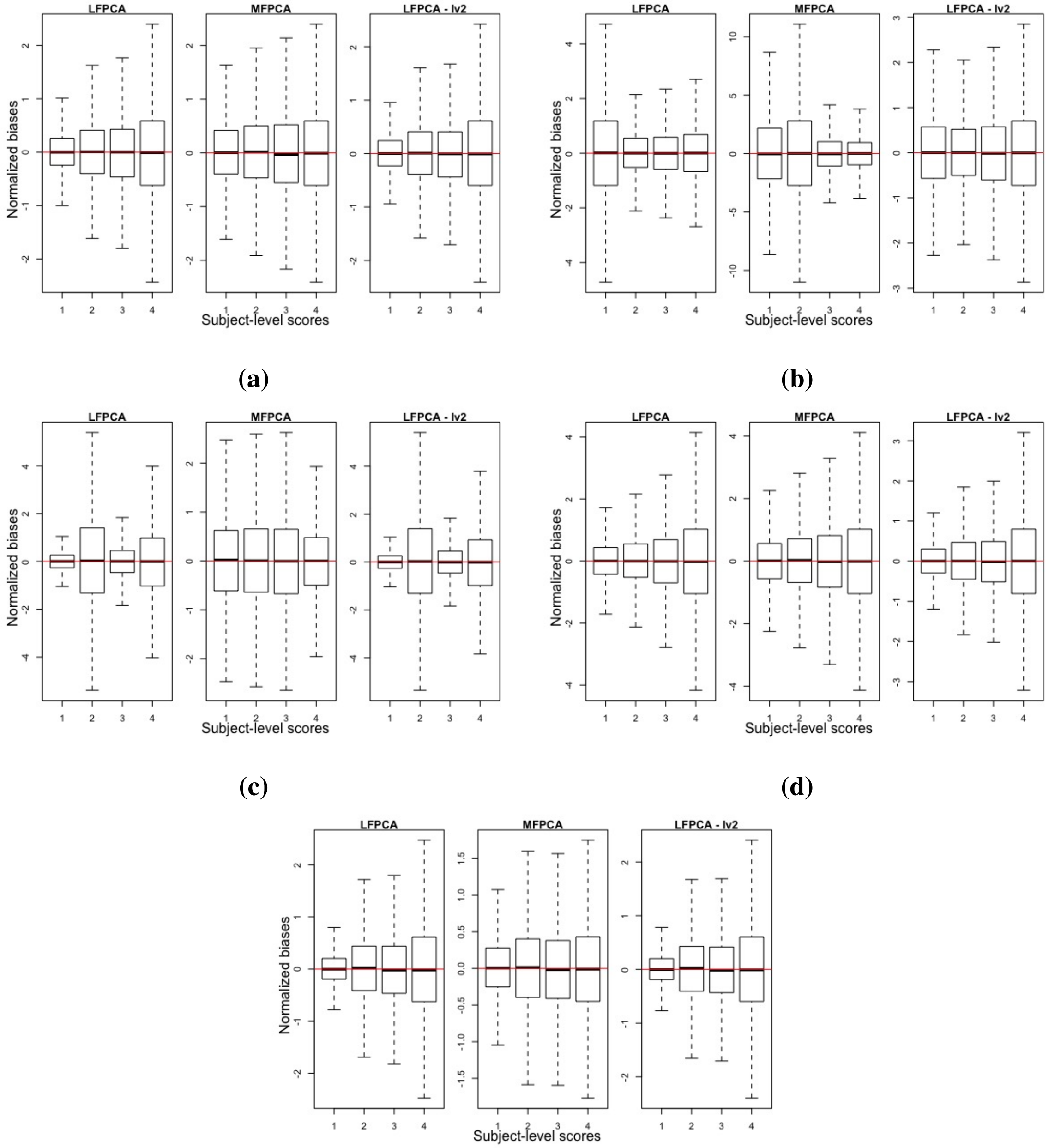
Table 5.1. Percentages of average variance explained by different levels of the first two principal components in (a) Simulated model (b) Three-level longitudinal FPCA (c) Three-level FPCA (d) Two-level longitudinal FPCA.

	# Component	$\phi_l^{U_0}$	$\phi_l^{U_1}$	ϕ_m^{V}	ϕ_r^{W}
(a) Simulated model					
	1	0.09	0.09	0.18	0.18
	2	0.04	0.04	0.09	0.09
(b) Three-level longitudinal FPCA					
	1	0.09	0.09	0.18	0.17
	2	0.04	0.04	0.09	0.09
(c) Three-level FPCA					
	1	0.09	0	0.22	0.17
	2	0.04	0	0.17	0.08
(d) Two-level longitudinal FPCA					
	1	0.13	0.14	0.27	0
	2	0.07	0.07	0.13	0

5.5 Application in MENU Study

The prevalence of obesity in the US has been steadily increasing over the last 20 years with recent age-adjusted estimates indicating that 42.4% of US adults are obese [61, 54]. Obesity can be associated with serious health risks [120]. For instance, compared with persons with normal weight, overweight or obese persons are more vulnerable to dyslipidemia, which is a major risk factor for cardiovascular disease and other comorbidities [134, 41, 46]. In addition, overweight status and obesity increase the risk of end-stage renal disease and many types of cancer [74, 93]. Since weight gain occurs when energy expenditure (EE) remains low while dietary consumption levels are high, certain amounts of physical activity (PA) for increasing EE are commonly considered as part of treatment plans for achieving weight-loss in obese individuals [65].

The MENU trial, conducted under the auspices of the NIH-funded Transdisciplinary Research on Energetics and Cancer (TREC) Study at UCSD from 2011–2017, recruited $n = 245$ overweight women to a 12-month three-arm dietary weight-loss intervention. All participants



(a)

(b)

(c)

(d)

(e)

Figure 5.7. Boxplots of the normalized errors between the estimated and true scores for subject-level process $(\hat{\xi}_{il} - \xi_{il})/\sqrt{\lambda_l^U}$ based on 100 replicates, comparing the three-level models with the two-level model. Red line represents the zero. The simulation scenario includes (a) FFFF (b) FFFL (c) FLFF (d) FFFE (e) FEFF.

received the same PA intervention. There were three study-related clinic visits at baseline, 6 and 12 months. [97, 154]. PA was measured using a triaxial accelerometer device, the GT3X

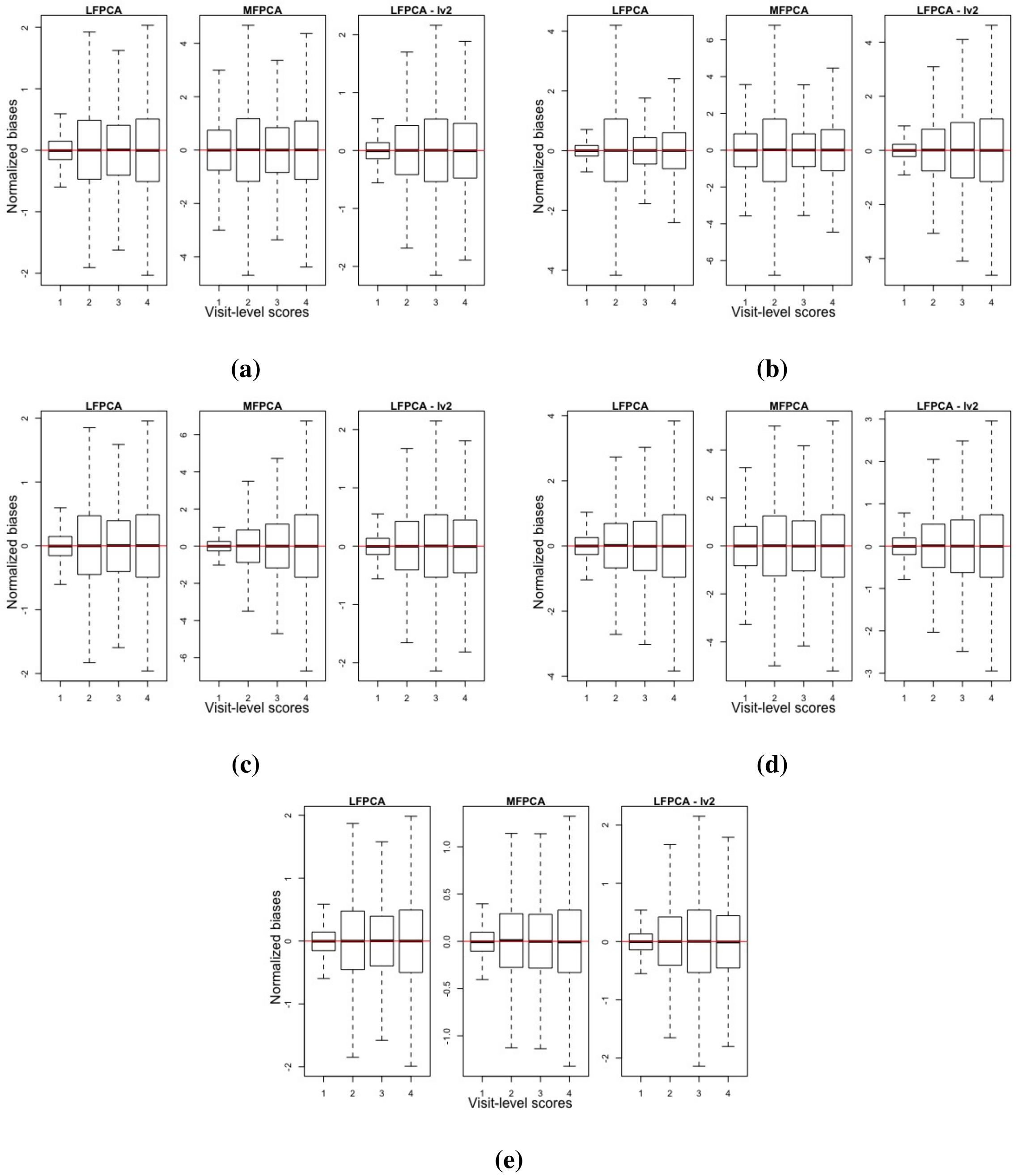
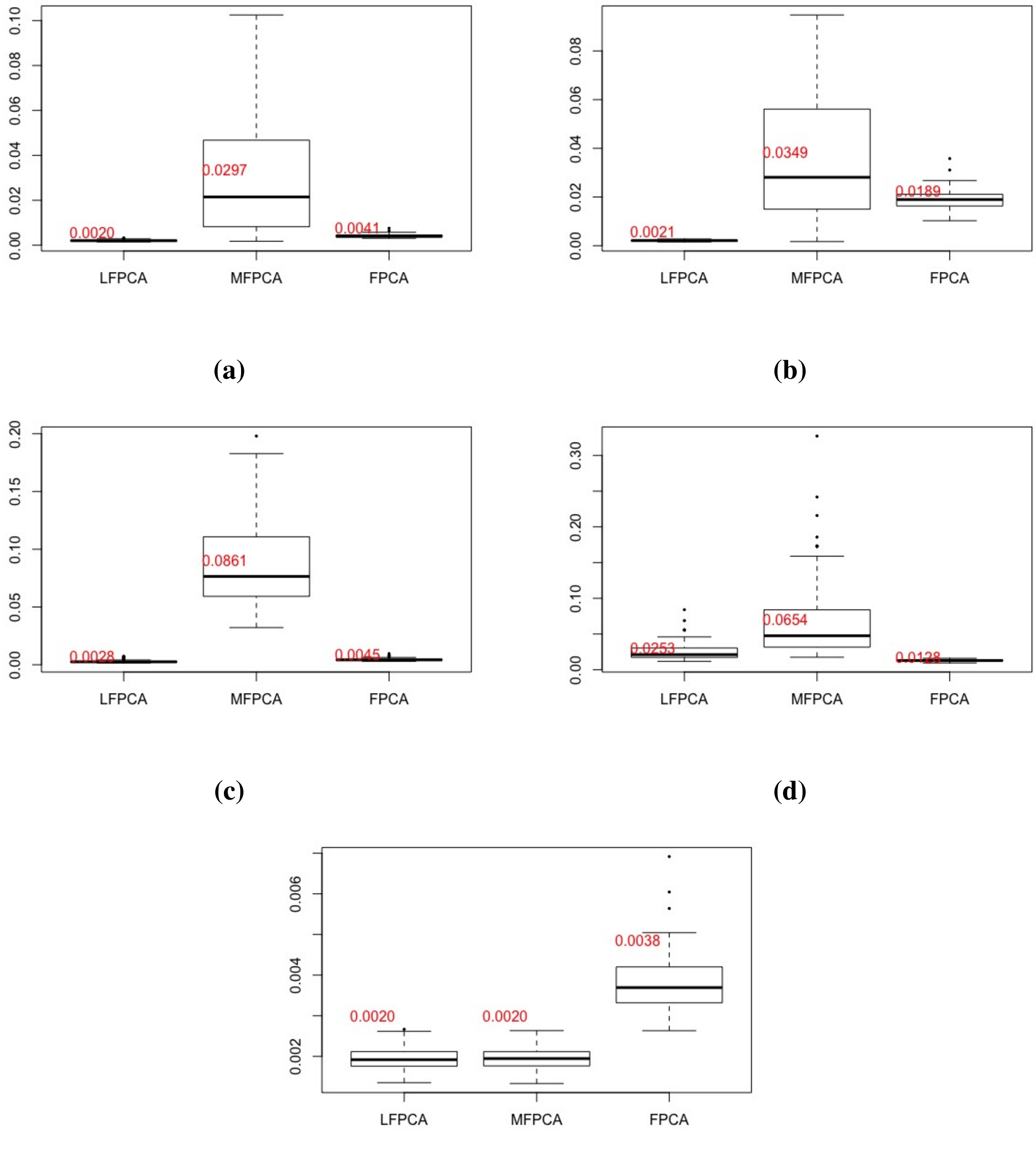


Figure 5.8. Boxplots of the normalized errors between the estimated and true scores for visit-level process $(\hat{\zeta}_{ijm} - \zeta_{ijm})/\sqrt{\lambda_m^V}$ based on 100 replicates, comparing the three-level models with the two-level model. Red line represents the zero. The simulation scenario includes (a) FFFF (b) FFFL (c) FLFF (d) FFFE (e) FEFF.

Actigraph (ActiGraph LLC, Pensacola FL). Participants were instructed to wear the devices for 7 days during waking hours and measurements of health outcomes were collected at each visit.



(e)

Figure 5.9. Boxplots of the MSE from the regression fitting $\frac{1}{M}\sum_{i,j}(y_{ij}-\hat{y}_{ij})^2$ based on 100 replicates, comparing the three-level models with the two-level model.The simulation scenario includes (a) FFFF (b) FFFL (c) FLFF (d) FFFE (e) FEFF.

The goal of the current work is to utilize the longitudinal accelerometer-based PA to implement three-level functional data methods, and evaluate associations with longitudinal health outcomes. For this purpose, to ensure consistent data availability across participants, we extracted daily

PA counts on three random days, including weekdays and weekends, for each participant at each visit. Sensitivity analysis were performed to show that the selected three-day data were representative of whole-week measures, explaining similar amount of variation at day-level.

For exploring the association between PA and overweight/obese status, we considered several related health outcomes, including body mass index (BMI), insulin levels and homeostatic model assessment (HOMA). The BMI, computed as weight in kilograms divided by height in meters squared (kg/m2), is commonly used to identify overweight/obese status if BMI ¿ 25.0 [54]. In addition, high-level of insulin was proven to be associated with lifestyle-dependent obesity risk factors [90]. The HOMA index is computed using fasting plasma glucose (mg/dL) and plasma insulin concentration (μU/mL), which further quantifies insulin resistance status. Therefore, lower values of each outcome indicate better metabolic health.

We first fitted the proposed three-level longitudinal FPCA model on the daily PA counts data for all subjects at each visit. Figure 5.10 presents the first estimated principal components for the random intercept, random slope, visit-specific and day-specific process by columns. The top row of the figure provides the first principal component at each level. It shows that the red curve, which represents adding (a multiple of) the principal component to the mean, is always higher than the mean (black) curve in each figure. Specifically, a high score on this component at the subject-level (level 1) indicates that a participant is on average more physically active, as well as a higher increase across visits, compared to a participant with a low score. Similarly, a high score at the visit-level (level 2) indicates that the participant has higher activity on that visit. Interestingly, the first principal component at the day-level emphasizes higher (or lower activity) during the first 200 minutes. Figure 5.11 provides an example of the daily raw, smoothed and model-recovered PA curves at each visit of one participant. As is evident, the model recovered curves mirror closely the (smoothed) observed PA data, which further illustrates the robustness and applicability of our proposed model. Finally, to have 95% explained variance, we acquired $N_U = 10$, $N_V = 5$ and $N_W = 13$ principal components at three levels, respectively.

Next, we aimed to fit a regression model to evaluate associations between functional PA inputs and longitudinal health outcomes. For this, as described in Section 5.3.4, the estimated scores and principal components are used to reconstruct the subject-level process U_i and visit-level process V_{ij} for each subject i and visit j. We then fitted the longitudinal FPCR model on each health outcome, respectively, with the reconstructed curves $\hat{U}_i$ and $\hat{V}_{ij}$, as denoted in equation 5.16. The model also includes a random intercept and additional covariates including age, ethnicity, smoking status and $\mathbf{1}(\text{visit} > 1)$. Figure 5.12 gives the estimated coefficient functions for log(Insulin), BMI and HOMA levels. From the coefficient functions for the subject-level (level 1) process of BMI, it shows that women with higher levels of PA than the 'average participant' are inclined to have lower BMI levels, especially in the first 300 minutes. The subject-level effects are not significant for either log(Insulin) and HOMA, as is evdenced by the 95% confidence bands including the null (zero) value. However, the estimated coefficient curves at visit-level (level 2) of these two health outcomes were negative and significant for the first 300 minutes, indicating that women with more PA than the "previous visit" tended to have lower insulin and HOMA. Interestingly, though the regression patterns at subject and visit level slightly varied among three health outcomes, we found that having PA earlier in the day was more beneficial for mitigating the overweight/obese status.

As a comparison, we also applied a two-level longitudinal FPCA model with the day-averaged data, similar as we did in the simulation study. Figure 5.13 provides the corresponding coefficient functions for three outcomes at the subject and visit level, which presents that none of these coefficient curves has significant pattern. Since the day-to-day PA patterns can be quite different within a week for one participant, taking the average of daily measures can eliminate or dampen trends in the data.

5.6 Discussion

In this work, we proposed a multi-level longitudinal functional principal component analysis approach and compared its performance with different functional principal component

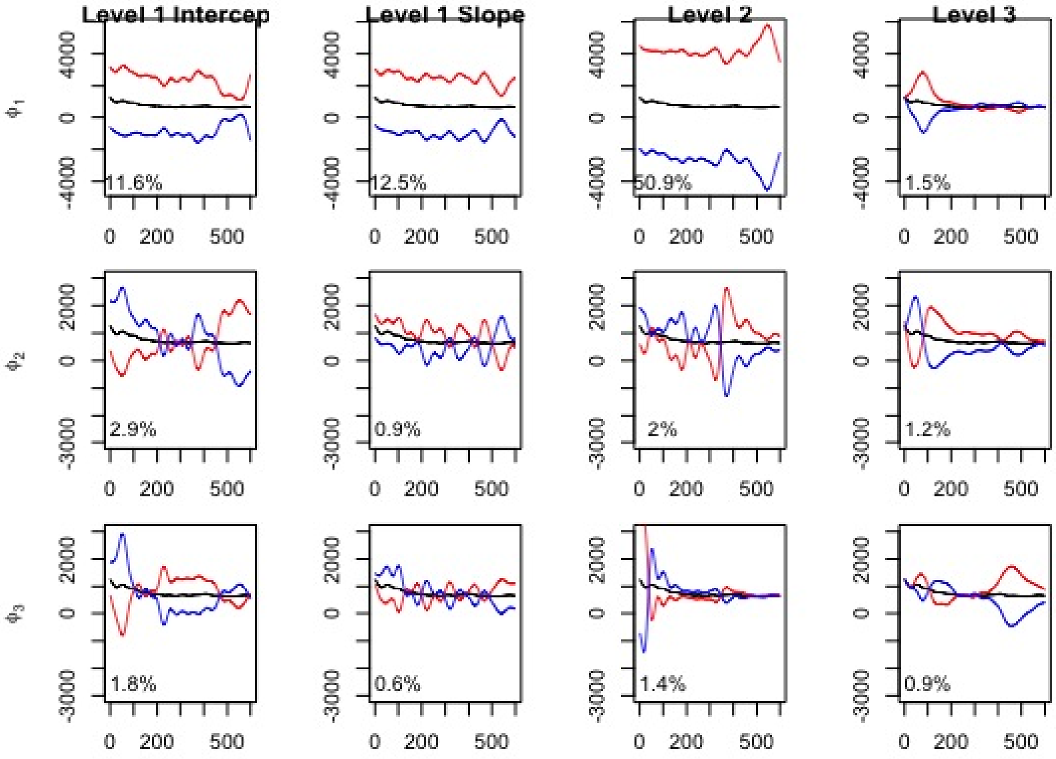


Figure 5.10. The first three estimated principal components for the random intercept (1st column), random slope (2nd column), visit-specific process (3rd column) and day-specific process (4th column). The plots give the overall mean value curve $\mu(t)$ (black) with addition (red) or subtraction (blue) of 2 square root of eigenvalues multiplying first, second or third level principal component curves.

models that have been previously applied on multilevel data, by means of both simulation study and real data application. Specifically, the proposed model was designed to fit data from a longitudinal study but has three-level functional inputs. It includes a two-step estimation procedure and eigen-expansion based methods to capture and decompose the covariance structures of the observed PA curves. The association between PA and overweight/obesity related health outcomes was then examined via functional regression approaches. In addition, a wide range of simulation studies were performed to validate and compare model performances.

Our proposed model can be considered as a natural extension of previous methodology on multilevel and longitudinal FPCA [33, 164, 60]. To demonstrate the necessity of such an extension, we provided both theoretical illustration in Section 5.3.3 and simulations in

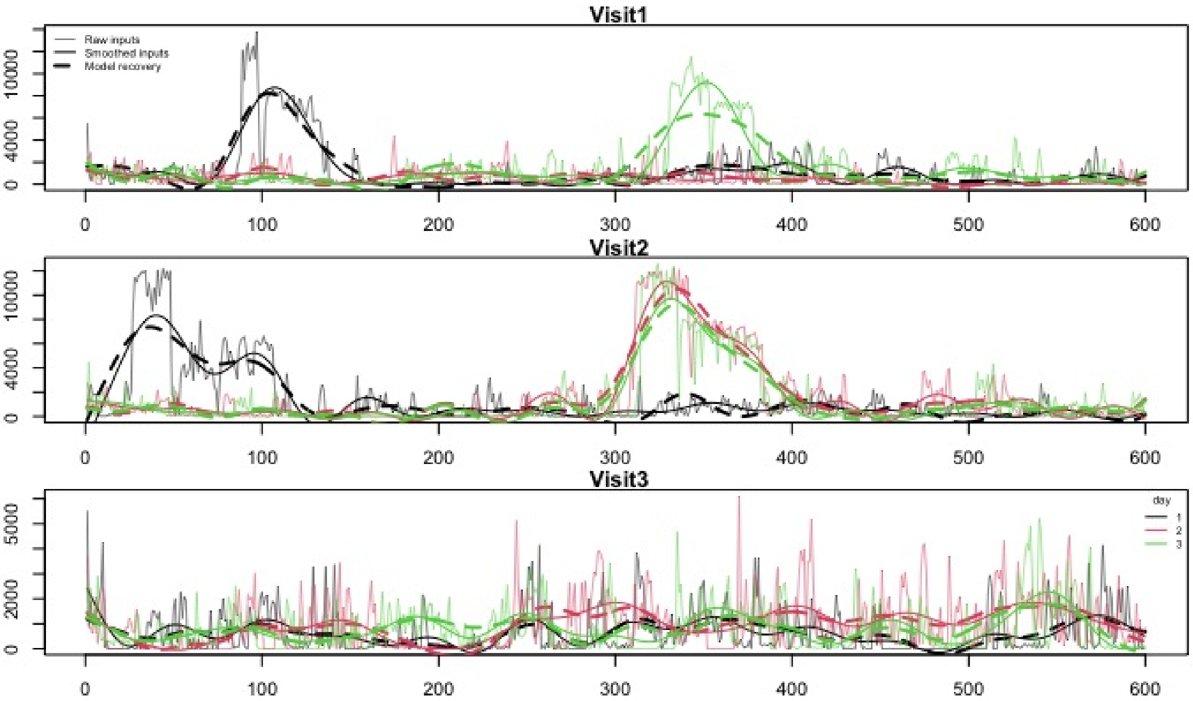
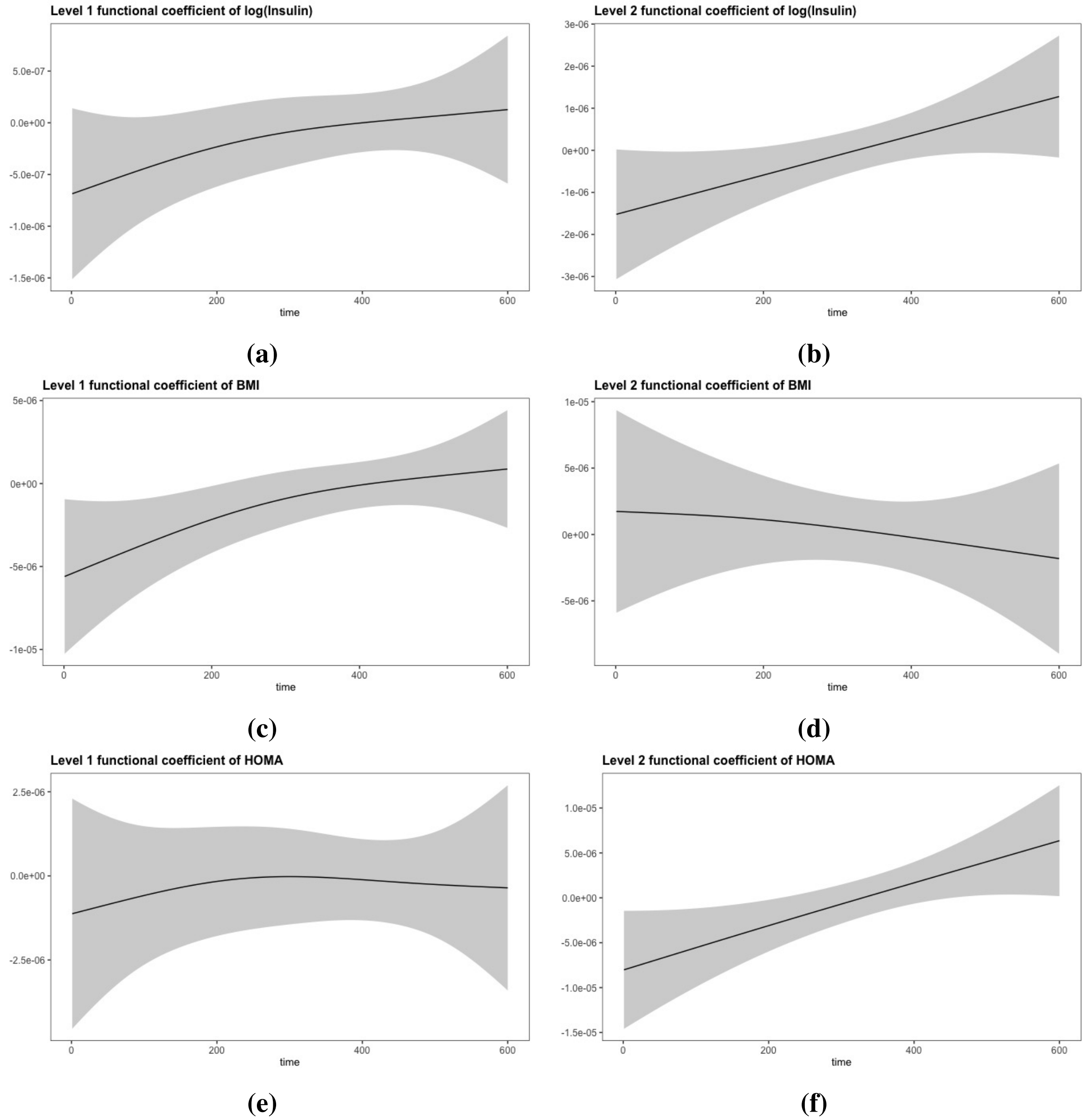


Figure 5.11. An example of PA records with raw count inputs (thin solid), smoothed curves (thick solid) and model-recovered curves (thick dashed) at baseline (top), 6 months (middle) and 12 months (bottom). Different colors of the line represent the day of the measurement.

Section 5.4. On the one hand, compared with previous implementations which used averaging to reduce the number of nested levels (e.g., averaging over days at each visit), our method retained the three-level longitudinal design structure, and thus can fully extract the variation information included in all nested levels and provide solid inference. On the other hand, though our proposed method consistently yields better performance, under certain scenarios, simpler models may be acceptable depending on the study aims, and when they accurately reflect the pertinent information contained in the data. For instance, if the day-to-day variation in a three-level dataset explains a relatively small proportion of total variability, a two-level longitudinal FPCA model may be applicable for inputs averaged over days. Importantly, our simulation studies illustrate different misspecification effects in cross-sectional and longitudinal setups. We found that in the setting with cross-sectional outcomes, even with longitudinal functional inputs (Section 5.4.1 Study 1), the simpler misspecified multilevel FPCA models (which ignore the slope term but retain the multilevel structure in the functional inputs) had



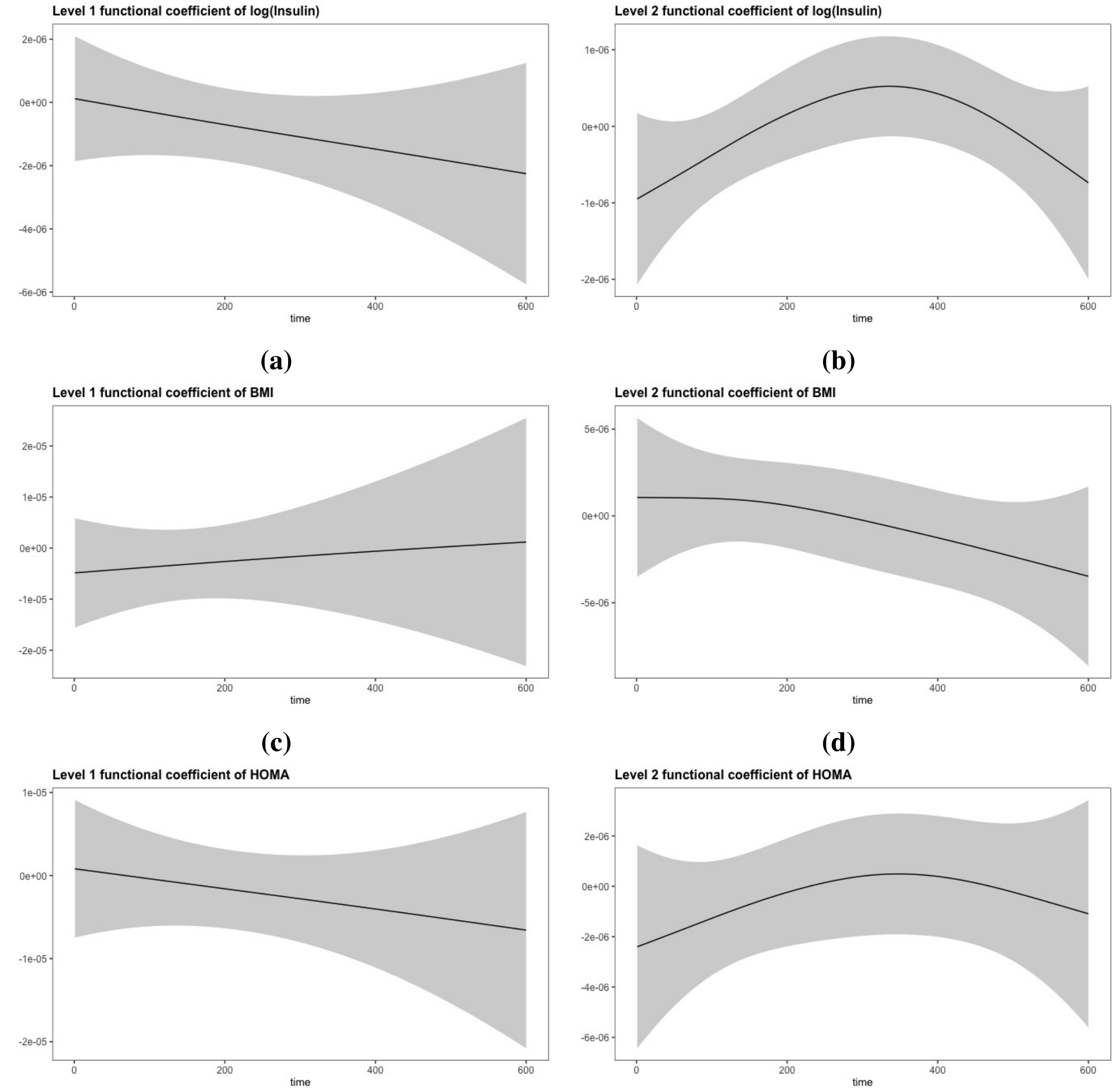
(a) **(b)**

(c) **(d)**

(e) **(f)**

Figure 5.12. Estimated coefficient functions when implementing the longitudinal FPCR model on log(Insulin) (top), BMI (middle) and HOMA (bottom), with U and V processes reconstructed from a two-level longitudinal FPCA model as functional predictors, after adjusting for age, ethnicity, smoking status, and visit$>$1.

superior performance in terms of estimation and prediction, compared to single-level FPCA which reduced the levels of the functional inputs by averaging (Figures 5.2, 5.4). However, in the setting with longitudinal outcomes (Section 5.4.2 Study 2), averaging the functional inputs (over the third-level) had superior performance compared to multilevel FPCA (which ignored the



(a) **(b)**

(c) **(d)**

(e) **(f)**

Figure 5.13. Estimated coefficient functions when implementing the longitudinal FPCR model on log(Insulin) (top), BMI (middle) and HOMA (bottom), with U and V processes reconstructed from a two-level longitudinal FPCA model as functional predictors, after adjusting for age, ethnicity, smoking status, and visit>1.

longitudinal component)[Figures 5.6, 5.9]. Thus, depending on the structure of the outcome data, misspecifying the longitudinal functional component appears to strongly influence results. We believe that these results can guide researchers in how to choose simpler approaches, should they wish to do so. Of course, preserving the full data structure, i.e., all levels of the functional data, performs best, and although this method appears to be more complex, the simulation studies and application indicate that the computation is in fact fairly efficient.

We also implemented the three-level longitudinal FPCA model in data application. Our analysis of the MENU study, revealed a negative association between physical activity and overweight/obese related health outcomes, that is, more diurnal physical activity is related with healthier status. Understanding how timing of physical activity most impacts health could be useful when designing intervention trials and informing public health recommendations.

Future studies can be further extended to functional data in longitudinal studies with more than three levels, such as studies in which the variation between morning versus evening physical activity are of interest. In fact, the structural FPCA proposed by Shou et al. (2015) [164] provided a general estimation procedure for data with any number of levels, but they do not explicitly consider longitudinal designs. By combining their approach and ours, we expect that these methods could be extended to multilevel longitudinal data, and we aim to pursue a similar approach in future work, in particular if relevant clinical questions are posed. In addition, with functional regression models where predictors are measured more frequently than outcomes, we could consider other summary metrics that can incorporate information from the higher levels in the predictors. For instance, Steele et al. (2017) [168] proposed multilevel structural equation models for longitudinal data, but the implementation involved with functional data needs further exploration, which we aim to address in future studies.

In summary, in this work, we propose an efficient two-step estimator for three-level longitudinal functional data, as are common in physical activity studies. Through simulations and theory we examine and compare this estimator to potentially simpler but missspecified model structures, and provide guidance when the simpler models may be appropriate. We applied our method to data from a longitudinal study on obesity measures and physical activity assessed via accelerometry and obtained meaningful results. Importantly, our approach can be applied to other applications with densely sampled data e.g., continuous glucose monitoring, heart rate monitoring etc. We believe that this work could add to the body of methods for analyzing data from wearable sensors, which are becoming more and more common in public health and biomedical applications.

Chapter 6

Spatial-Temporal Modeling and Spatial Inference Using NA-CORDEX Climate Data

6.1 Introduction

The historical and future behavior of regional climate change is of great interest because of its potential effects on policymakers in managing and mitigating the impacts of global climate change on local communities [1]. For climate conditions at any given location, the previous stationarity assumption states that the future statistics of climate conditions (e.g., temperature, precipitation) will be similar to the recent past when averaged over a sufficiently long time. However, this assumption of stationarity is becoming unreliable today, due to increasing emissions of carbon dioxide, methane, and other heat-trapping greenhouse gases from human activities [91]. It has been reported that annual and seasonal temperatures have increased by 1.3 to 1.9 °F (0.7 to 1.1 °C) since records began in 1895 in the United States [180]. Meanwhile, as the temperature increases, more water evaporates from aquatic systems, i.e. oceans, lakes, etc., which have already caused more heavy rainfall and precipitation events over the past 50 years. In climate research, a trend is defined as the gradual change of climate variables, such as temperature and precipitation, over time and the assumption of stationarity intrinsically implies no significant trend exists. Therefore, statistical modeling and testing the trend of current climate change, especially at regional or local scale, is beneficial and crucial for monitoring currently non-stationary environment.

The global climate models (GCMs) were initially developed to model the earth system, seeking to understand the relationship between multiple aspects of the environment and modeling the dynamics in the future. The GCMs operate at fairly coarse resolutions of approximately 100–500 km [70], which may not be sufficient for capturing the local-scale climate features. Regional Climate Models (RCMs) were used to address this limitation, with both increased resolution and improved representation of physical processes in the model [91, 70]. In fact, the GCMs provide the multiple boundary conditions for the RCMs, then the RCMs can dynamically downscale and model climate behavior within a limited area with finer scale. With higher-resolution data from RCMs, we were able to statistically model the climate variables, such as temperature and precipitation, and perform significance tests on parameters of interest at regional and local scales.

The estimation and prediction of the climate change trend and its associated significance are of high priority in climate research. Simple linear regression is the simplest model for estimating the linear trend (slope), and the corresponding statistical significance can be tested via a Z-test or Student-t test. However, certain limitations exist in this commonly used model. Firstly, since the climate data were observed over a period of time, these observations are not necessarily independent and could potentially have temporal correlations with each other. In addition, Gaussian assumptions in a linear model may not be appropriate with non-negative climate measures, such as precipitation. Ye et al. proposed a time-series model consisted with both deterministic and stochastic processes for monthly absolute temperature data [196]. Specifically, deterministic processes contained the trend term and stochastic processes were explained by seasonal autoregressive integrated moving average models. The model was shown to have a promising performance in predicting future temperature, while the parameter of the linear trend can not be directly interpreted as the change over time. Other non-parametric tests, such as the Mann-Kendall (M-K) test, were widely used to detect trends in precipitation modeling, due to less sensitive to outliers and skewed distributions [197, 170, 5]. However, these tests might not be appropriate for climate data with seasonal effects and don't explicitly inform the magnitude of change. To address these limitations, we proposed multivariate time series regression modeling strategies for temperature and precipitation under different distribution assumptions, along with a series of diagnostic tools for validating the model selections. Data from each season were treated as a sequence of time series observations, and they were combined to form a multivariate outcome matrix with correlations between seasons.

In addition to modeling the temporal trend, because climate data derived from RCMs reside on a finer grid, the spatial correlation between time series is a nonnegligible factor. The geographically weighted regression (GWR) was proposed to investigate heterogeneity in data relationships across geographic space and constructed local spatial relationships via spatial weight matrix [21]. Huang et al. (2010) [76] extended the GWR for modeling local spatial dependency across time using a spatiotemporal weight matrix. However, calculating distance in three dimensions can be challenging since the scales of geographical distance and time are usually different. Fotheringham et al. (2015) [45] further proposed a geographical and temporal weighted regression (GTWR) model by constructing spatiotemporal kernel functions to ensure the data points are both spatially and temporally weighted. These implementations motivated us to propose geographically weighted multivariate spatial-temporal models for climate data, which not only incorporated geographical correlation between locations but also kept the interpretability of parameters in the trend analysis. Under different model assumptions of temperature and precipitation, we validated that these geographically weighted models provide unbiased estimates and improve the model performance.

With estimated climate change parameters, appropriate significance testing procedure over a space domain is necessary for identifying statements about the climate change effects. Pointwise tests, such as the Z-test on the point-by-point basis, are basic methods to determine where an effect is significant and easily to perform in climate research [32]. However, this pointwise inference can not control familywise error rate (FWER), the probability of making at least one Type I error when a series of hypothesis tests are simultaneously performed. In normal situations, it can be fixed with the traditional Bonferroni correction, but may be too conservative for climate models, where data are sampled over a dense grid and many tests are required. French et al. (2017) [50] summarized approaches for simultaneous inference by controlling the false discovery rate (FDR) or FWER, while they normally require the field to be Gaussian. In contrast, Sommerfeld et al. (2015) [165] proposed a approach for addressing this issue by constructing Coverage Probability Excursion (CoPE) sets. For data observed on a fine grid of fixed locations, the CoPE method is able to account for the simultaneous inference problem and computes statistically significant spatial region based on proposed hypothesis. Meanwhile, it is also fast to apply and only requires mild assumptions about the data structure. Using the CoPE method, we intended to perform simultaneous tests of the trend parameters over the space and determine the region where a climate change effect is significant.

Spatial-temporal random field using Bayesian hierarchical model (BHM) paradigm was also commonly implemented to provide estimation and inference framework for climate evolution, where temporal data were considered as a realization of random field [15, 186]. In a Bayesian framework, a common and useful approach is to use Markov chain Monte Carlo (MCMC) techniques [132] for spatial modeling, given their power and simultaneously accounting for parameter uncertainty. Samantaray et al. [157], for instance, investigated changes in regionalization and regional hydroclimatic patterns over India using Markov random field model and provided spatial inference of estimated variable. However, Bayesian approaches usually come at a cost of greater computational complexity, especially in large spatial datasets [10]. It mainly arises from the difficulty when modeling large covariance matrices, which also needs prior information to specify them with correct formats. Thus, our implementation is more applicable from both efficiency and flexibility points of view. Firstly, because the GWR provides unbiased estimators, the algorithm can still provide robust parameter estimates even when covariance structure between locations is misspecified. Secondly, the CoPE-based spatial inference is performed on the obtained parameter space rather than the original data, thus it only requires minimal diagnostic checking.

For modeling and testing the climate change effects for temperature and precipitation in both spatial and temporal domains, we proposed a two-step analysis procedure in this study. In the first step, we selected the optimal spatial-temporal models using strategies from geographically weighted multivariate time series regression models, to extract interested climate change parameters based on series of diagnostic analysis and model comparisons. Secondly, we applied spatial inference, i.e. the CoPE method, on these estimated parameters and identified spatial confidence regions where a significant climate change pattern exists. With a case study using data from North American CORDEX (NA-CORDEX) program, we aimed to provide a pipeline and a reference for constructing spatial-temporal models with data generated from RCMs, as well as evaluating the effect of regional climate change based on novel statistical inference method.

6.2 Data Overview

The data we used for this study was acquired from the North American CORDEX Program (NA-CORDEX) [117], which contains output from RCMs run over a domain covering most of North America using boundary conditions from GCM simulations in the Coupled Model Intercomparison Project Phase 5 (CMIP5) archive. Because we mainly focused on providing a pipeline of statistical modeling and inference in regional climate data in this paper, we extracted raw monthly near-surface temperature ($^\circ C$) and precipitation (mm/day) data of 0.44° (50-km projected grid) in Kansas (KS), Colorado (CO) and California (CA), with both history data (1950 - 2005) and future data (2006 - 2100) from the a typical combination of CanRCM4 (RCM 8.5) and CanESM2 (GCM) [159]. A detailed description and characteristics of NA-CORDEX models can be found on https://na-cordex.org/rcm-characteristics.

CA, CO and KS were selected as regional examples for representing varying geographical and topological conditions in the US. Figure 6.1 presents the elevation maps of the three states. KS has a square shape and is located on the great central plain of the US, with a generally flat or undulating surface among two-thirds of the state and a mild elevation increase from east to west. CO is inside the Mountain States region and is famous for its diverse geography, which includes alpine mountains, high plains, desert lands, and deep canyons, which could affect local climate. CA is considered to be the one with the most complex diversity in both geography and topology among all three states. It has a vast land, irregular border shape and is beside the ocean coast, where ocean may play an important role in moderating the climate. Also, CA is also home to both the highest (Mount Whitney) and lowest (Death Valley) points of the continental US. Considering the geographical variability, we seek to derive statistical models which can capture a systematical change of climate and meanwhile, can reflect distinct characteristics across the three states. In addition, we also picked several typical locations in these states for further illustrating local features at specific locations. For instance, San Francisco, San Diego, Death Valley and Yosemite were selected to represent Northern/Southern and mountain/dessert areas in CA. In addition, for constructing a robust statistical model and inference, ocean areas are not considered while partial areas adjacent to CA in Nevada are included in our analysis, which establish a squared shape at the east of CA map.

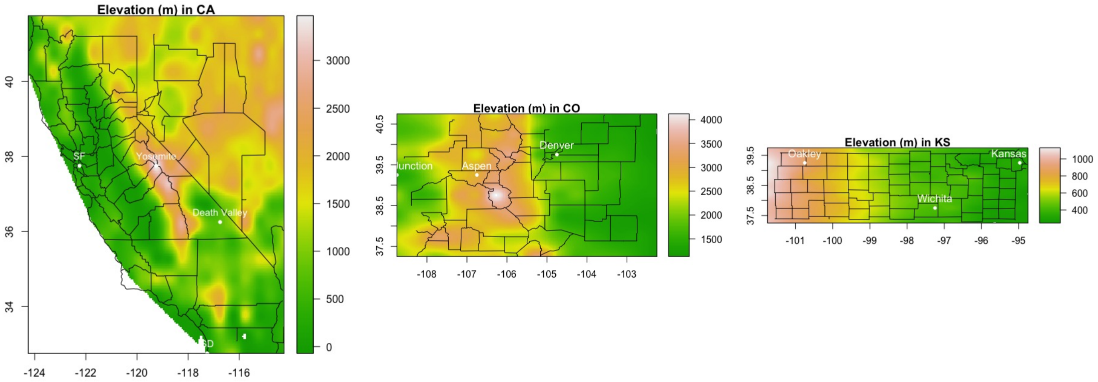


Figure 6.1. Elevation (m) map in CA, CO and KS.

In this study, we targeted to model climate change of each season in a year, therefore, the seasonal means of monthly temperature and precipitation were computed as the outcome variables. In particular, since all monthly records start in January, we define seasons as Winter

(January, February, March), Spring (April, May, June), Summer (July, August, September) and Fall (October, November, December). Figure 6.2 illustrates examples of seasonal climate data in CA. Two heat maps on the top show the spatial distributions of temperature (left) and precipitation (right), respectively, in 1992. In consistent with the common knowledge, here we use darker blue to represent either lower temperature or more precipitation. It can be seen from the maps that, at the given grid resolution, measurements of temperature and precipitation over space look smooth and can be treated as spatially continuous data. Meanwhile, we provide two figures on the bottom to show time series of temperature and precipitation at these locations in historical scenario from 1950 to 2005 and RCM-based future scenario from 2006 to 2100. It is apparent that the temperature has an increasing trend over years, with similar patterns for all four seasons. Compared with historical series, the simulated future series tend to have a faster increase. It is also noticeable that the patterns and progression of both temperature and precipitation differ significantly at these locations, indicating that the climate variability can be common in CA. Therefore, we also intend to derive models to reflect the continuity, similarity and discrepancy of climate data within a state, in both time and spatial domains.

6.3 Analysis Overview

The statistical strategy for analyzing the climate data has three major steps, including temporal modeling, spatial-temporal modeling and statistical inference, to form a comprehensive and solid analysis pipeline. We used the historical temperature and precipitation data as training data for constructing our final models, while the future climate data will be used as the validating set. Meanwhile, we only presented modeling results in CA as an illustrating example, while for the purpose of comparison, statistical inference results of all three states are included. Additional results can be found in the Supplementary Materials.

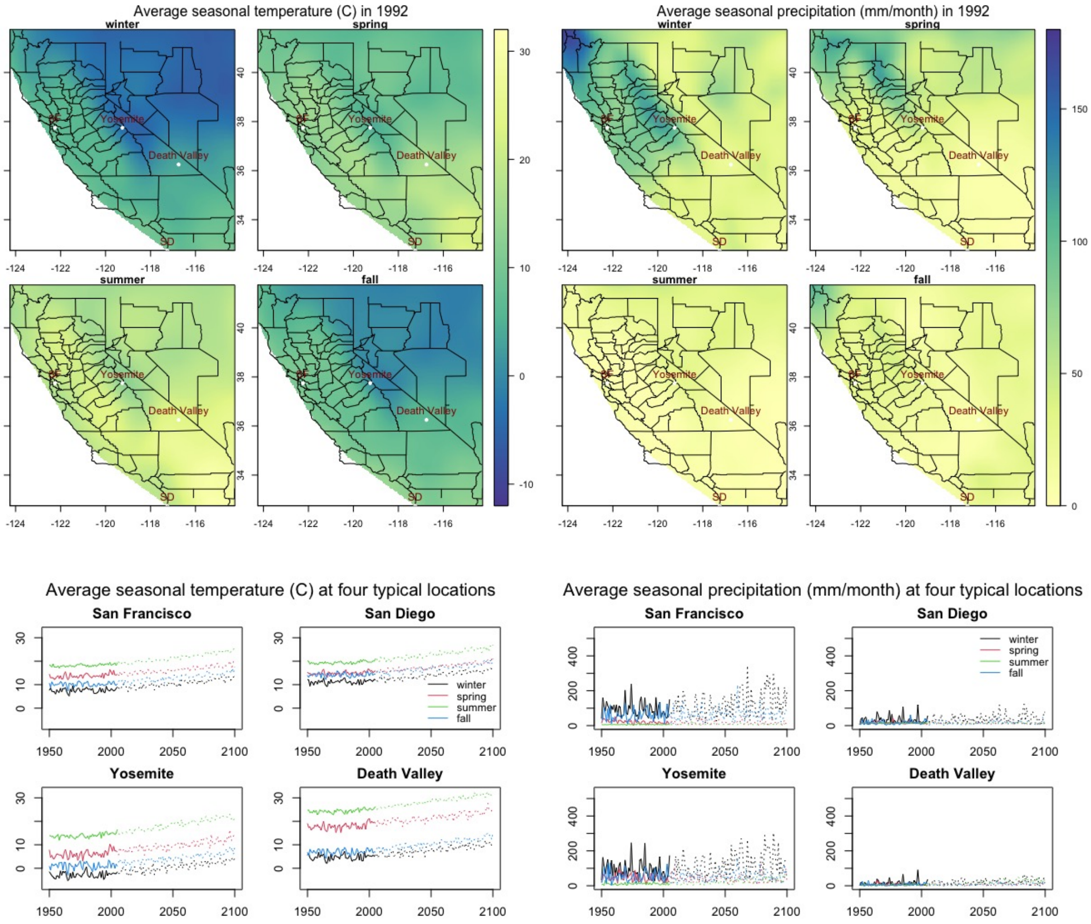


Figure 6.2. Example of temperature (left) and precipitation (right) data in CA. Figures on the top present the climate data in 1992 on map. Lower temperature and higher precipitation are represented with darker blue, respectively. Four typical locations are marked on the map and their corresponding seasonal temperature and precipitation time series from 1950 to 2100.

6.3.1 Temporal modeling

We started constructing models for temperature and precipitation with only temporal modeling and fitted models at each location individually. Regarding distinct assumptions of data distributions, linear and gamma regression models were considered for temperature and precipitation, respectively. The temporal models are supposed to have a form of multivariate linear or gamma time series regression models for temperature and precipitation separately, while they may be simplified based on further model checking. Details of temporal modeling are introduced in Section 6.4 and the main procedure includes:

1. Checking the validity of data assumptions. Specifically, linear models are considered for fitting temperature data and the Shapiro-Wilk test [163] was performed to check the normality assumption. As for positively skewed distributed precipitation data, it is assumed that they have gamma distributions and gamma regression models were implemented [188].
2. Investigating within-season correlations, such as the autoregressive (AR) effect of regression residuals. Temporal dependence of temperature and precipitation data in each season could be assessed using autocorrelation function (ACF), partial autocorrelation function (PACF) or the Durbin Watson test [39].
3. Assessing the between-season correlations using a simple linear regression model.
4. Constructing final models based on selection results, with a form of multivariate linear or gamma time series regression model for temperature or precipitation, respectively.

6.3.2 Spatial-temporal modeling

The temporal model only considered modeling the data in time domain and fitted independent models at each location, which may neglect potential spatial correlations between neighboring locations. Meanwhile, the geographical conditions at the location, such as local elevation, may also affect the climate change. Therefore, we extended individual temporal models to geographically weighted models, incorporating spatial effects using GWR strategy [21]. For temperature and precipitation data, the following steps are performed and modeling details are illustrated in Section 6.5,

1. Computing the pairwise Pearson correlation coefficients of seasonal residual sequences from the temporal modeling between locations and constructing weight functions in by inspecting the association between the pairwise correlation and distance.

2. Incorporating the weight information to build the geographical weighted multivariate time series regression (GWMTSR) model and gamma regression (GWGR) model for temperature and precipitation, respectively.

6.3.3 Statistical inference

We implemented inferential procedure to identify locations with significant climate change effects in the spatial domain, using the coverage probability excursion (CoPE) sets approach proposed by Sommerfeld et al. (2015) [165]. Details of the CoPE methodology are provided in Section 6.6, which in brief have two steps,

1. Assessing whether the assumptions, including continuity and Gaussianity of estimated parameters and the independence of errors at each location, apply for results obtained from temperature and precipitation modeling.

2. Constructing CoPE sets on historical and future temperature and precipitation results, with some prespefied sets of threshold levels.

6.4 Temporal modeling

6.4.1 Multivariate time series regression model

In this section, we seek to generalize models which can explain the overall seasonal temporal changing of temperature and precipitation, independently fitted at each location. For a time series vector $\boldsymbol{y}_{is} = (y_{i1s}, \ldots, y_{iTs})'$ at location $i, i = 1, \ldots, n$ and in season $s, s = 1, \ldots, 4$, on a fine grid time interval $t \in \{t_1, \ldots, t_T\}$, a univariate time series regression model is defined with the form,

$$\boldsymbol{y}_{is} = \boldsymbol{X}_i' \boldsymbol{\beta}_{i(s)} + \boldsymbol{\phi}_{is} \boldsymbol{b}_{i(s)} + \boldsymbol{\varepsilon}_{is}, \tag{6.1}$$

where $\boldsymbol{X}_i$ is a $(k+1)\times T$ predictor matrix and $\boldsymbol{\beta}_{i(s)}$ is the corresponding $(k+1)\times 1$ coefficient vector. $\boldsymbol{\phi}_{is}$ is the lagged response vector variable of $\boldsymbol{y}_{is}$, such as AR(p), and $\boldsymbol{b}_{i(s)}$ is a $p\times 1$ parameter matrix. $\boldsymbol{\varepsilon}_{is}$ is a sequence of independent and identically distributed white noise vector, $N(0,\sigma_{is})$. Specifically, if the X_{ij} is specified as $(1,t_j)$ and $\boldsymbol{\phi}_{is}=0$, the model decays to a simple linear regression for trend estimation.

In addition to modeling each season sequence independently and considering the potential cross-season correlations, we stacked all four seasonal time series together to construct a $T\times 4$ outcome matrix $\boldsymbol{Y}_i=(\boldsymbol{y}_{i1},\boldsymbol{y}_{i2},\boldsymbol{y}_{i3},\boldsymbol{y}_{i4})$ and proposed the multivariate time series regression (MTSR) model,

$$\boldsymbol{Y}_i=\boldsymbol{X}_i'\boldsymbol{\beta}_i+\boldsymbol{\Phi}_i\boldsymbol{b}_i+\boldsymbol{\varepsilon}_i, \tag{6.2}$$

where $\boldsymbol{\Phi}_i$ is the lagged response matrix variable for $\boldsymbol{Y}_i$. $\boldsymbol{\beta}_i=[\boldsymbol{\beta}_{i(1)},\boldsymbol{\beta}_{i(2)},\boldsymbol{\beta}_{i(3)},\boldsymbol{\beta}_{i(4)}]$ and $\boldsymbol{b}_i=[\boldsymbol{b}_{i(1)},\boldsymbol{b}_{i(2)},\boldsymbol{b}_{i(3)},\boldsymbol{b}_{i(4)}]$ are the parameter matrices. Rows in $\boldsymbol{\varepsilon}_i$ are white noise processes and have distributions $N_4(0,\Sigma_i)$. Estimation of coefficients $\boldsymbol{\beta}_i$ and $\boldsymbol{b}_i$ is based on the assumptions of outcome variable distributions. For instance, least square estimates are generally provided for temperature data with Gaussian distribution. We implemented the ideas of MTSR models for temperature and precipitation, respectively. In particular, to separate the notations of two types of climate measures, we use y to denote temperature data and z to denote precipitation data in following sections.

6.4.2 Temperature model

We first provided an illustration example for showing the procedure of selecting a model for modeling temperature data, by inspecting the normality of the seasonal temperature sequences, as well as within-seasonal and cross-seasonal correlations. For assessing the validity of the first two assumptions, a simple linear regression model was fitted at each location i and season s

individually, that is, fitting Equation 6.1 with covariates specified as $\boldsymbol{\phi}_{is} = 0$ and $\boldsymbol{X}_i$ with column vectors $(1, t_j)'$. The acquired residual sequences $\boldsymbol{\varepsilon}_{is}$ can be considered as a detrending process in analyzing time series data. The normality of the residuals was evaluated using Shapiro-Wilk tests and the autocorrelation test was performed with the `ACF` function in R. We further examined the cross-seasonal correlations by fitting simple linear models between two consecutive seasons, where a significant slope indicates nonnegligible association. Figure 6.3 shows diagnostic results for selecting temporal models for temperature data in CA.The normality assumption is confirmed with Figure 3a, since the empirical cumulative distribution function (CDF) of the p values from Shapiro-Wilk tests approximately follow the CDF of a standard uniform distribution for all four seasons. Figure 3b presents lag-1 autocorrelation values of each season and their corresponding 95% confidence intervals, which shows no significant within-seasonal autocorrelations. Figure 3c displays the p values of coefficients from the one-on-one between-season linear regression at each location, indicating that certain numbers of locations have significant correlation between consecutive seasons, except the correlation between Fall and Summer. From these diagnostic results, selected models might vary among locations and seasons, it is recommended to fit one single model which fits most of cases for the purpose of simplicity and consistency. Meanwhile, we further demonstrated that the proposed MTSR models can be reformalized to approximate the limit of trend parameter in a simple linear regression, which is the parameter of interest in this study.

Based on preliminary diagnostic results, the MTSR model for modeling temperature was given as,

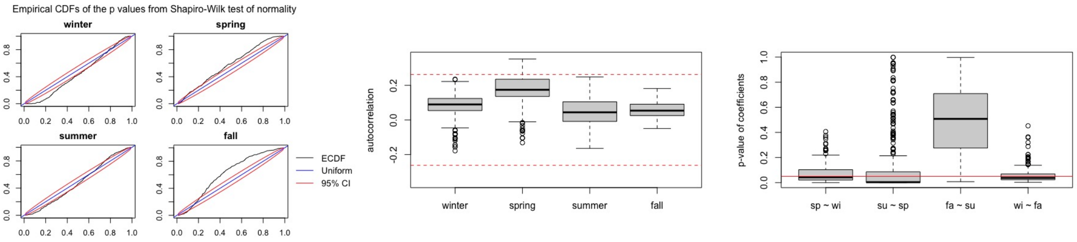
$$\boldsymbol{Y}_i = \boldsymbol{X}_i'\boldsymbol{\beta}_i + \boldsymbol{Y}_i^{(1)}\boldsymbol{b}_i + \boldsymbol{\varepsilon}_i, \quad \boldsymbol{\varepsilon}_i \sim N_4(0, \Sigma_i). \tag{6.3}$$

$\boldsymbol{X}_i$ is the design matrix with column vectors $(1, t_j)$, representing the deterministic trend in the regression model. t_j is the centralized time variable of year $j = 2, \ldots, T$ divided by 10, which leads to the corresponding trend coefficient $\boldsymbol{\beta}_{i1}$ be interpreted as the change per decade. Based on model selection, $\boldsymbol{\Phi}_i$ in Equation 6.2 is replaced with $\boldsymbol{Y}_i^{(1)}$, with typical row elements $\boldsymbol{y}_{ij}^{(1)} = (y_{i(j-1)4}, y_{ij1}, y_{ij2}, y_{ij3})$. It provides a stochastic process in the regression model, which also reflects the between-season effects. The superscript 1 represents the consecutive or lag-1 cross-seasonal effect and we can further denote the *k*-year lagged seasonal effects with $\boldsymbol{y}_{ij}^{(k)}$. It is easy to note that $\boldsymbol{y}_{ij}^{(4)} = \boldsymbol{y}_{ij-1}$. We denote the full set of covariate matrix as $\boldsymbol{S}_i = (\boldsymbol{X}_i, \boldsymbol{Y}_i^{(1)})$, including both deterministic and stochastic processes.

$\boldsymbol{\beta}_i = (\boldsymbol{\beta}_{i0}, \boldsymbol{\beta}_{i1})$ and $\boldsymbol{b}_i$ are the corresponding parameters matrix with dimension 2×4 and 4×4. In this temperature model, $\boldsymbol{b}_i$ is a diagonal matrix, represented as $\text{diag}(b_{i1}, b_{i2}, b_{i3}, b_{i4})$. We use $\boldsymbol{B}_i = \{\boldsymbol{\beta}_i, \boldsymbol{b}_i\}$ and $\Sigma_{\boldsymbol{B}_i}$ to denote the full parameter space at location *i* and its corresponding variance-covariance matrix, respectively. Therefore, $\hat{\boldsymbol{B}}_i$ can be estimated using a least square estimator $(\boldsymbol{S}_i'\boldsymbol{S}_i)^{-1}\boldsymbol{S}_i'\boldsymbol{Y}_i$, and covariance of the parameter estimates is computed as $\hat{\Sigma}_{\boldsymbol{B}_i} = (\boldsymbol{S}_i'\boldsymbol{S}_i)^{-1}\hat{\Sigma}_i$, where $\hat{\Sigma}_i$ is the empirical estimator of Σ_i and can be estimated from residuals. The process of estimation follows the least square theory for multivariate linear regression and details of the derivation can be found in Johnson and Wichern (2014, Chapter 7) [86].

Figure 6.4 presents the heat maps of estimated linear time trend parameters ($\hat{\boldsymbol{\beta}}_{i1}, i = 1, \ldots, n$) in four seasons and their corresponding Z-scores, defined as $\hat{\boldsymbol{\beta}}_{i1}/se(\hat{\boldsymbol{\beta}}_{i1})$. In general, the estimated parameters over the space are smooth and the field can be treated as continuous at the given grid resolution. Adjusting with the cross-seasonal effects in regression models, these estimates are positive mostly at all locations in Winter, Spring, Summer and at southern part in Fall, indicating an increasing trend of the temperature in historical series. Specifically, Spring was estimated to have the most drastic warming effects, reflected in both parameter and Z-Score levels. Meanwhile, the Z-score maps also indicate that temporal temperature changing patterns were not significant in Winter and Fall, with the values very close to 0.

It is noted that $\boldsymbol{\beta}_{i1}$'s are not equivalent with the slope estimates in trend analysis, since the model is adjusted with neighboring season variable. To compensate for the differing interpretation of trend parameters in our model and simple linear regression, we further derived a changing parameter $\boldsymbol{v}_{ij}$, which has the same interpretation as the trend slope when j is large. Details of the derivation are included in Appendix A.



(a) **(b)** **(c)**

Figure 6.3. Temporal model selection for CA temperature data. (a) Empirical CDFs of the p values from the Shapiro-Wilk test of gaussianity of residuals. The black line is the CDF of a standard uniform distribution and red lines provide the 95% confidence band of it. This figure demonstrates the validity of gaussianity assumption. (b) Boxplots of autocorrelation of the residuals at lag 1. The red lines are the 95% confidence band based on $\pm 1/\sqrt{T-1}$. The figure shows that within-seasonal autoregressive model are not needed. (c) Boxplots of coefficient p values from fitting linear regression between consecutive seasons. The red line provides the threshold p values of 0.05. This figure proves that a large amount of places have significant correlation between consecutive seasons.

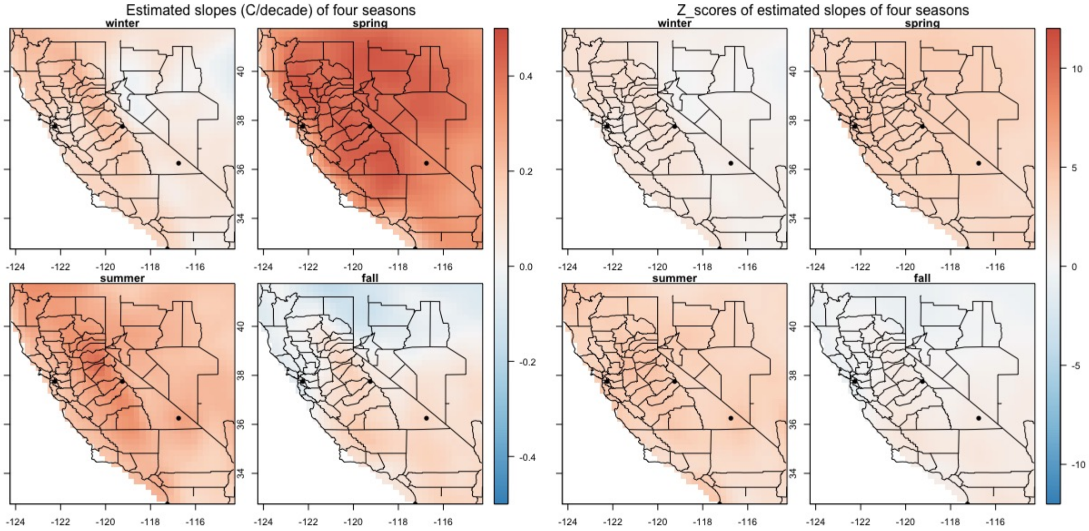


Figure 6.4. Heat maps of estimated linear time trend parameters (left) and corresponding Z-scores (right) of temperature change ($^{\circ}C$) in a decade of four seasons from MTSR models, fitted individually at each location.

6.4.3 Precipitation model

The gamma distribution is assumed to be suitable for modeling distributions of precipitation data in previous studies [188, 122, 129]. It assumes that the response variable z has a gamma distribution, with the form $z \sim Gamma(\alpha, \theta)$ and indicating $E(z) = \alpha\theta$ and $Var(z) = \alpha\theta^2$. The probability density function of gamma distribution with parameters α and θ is given as $f(z, \alpha, \theta) = \frac{1}{\theta^\alpha \Gamma(\alpha)} z^{\alpha-1} e^{-\frac{z}{\theta}}$, where $\Gamma(a)$ represents the gamma function evaluated at α. Similarly, we performed within-season and between-season tests (Figure 6.5) for precipitation data, and no significant effect was found from both tests, meaning that temporal correlation can be neglected. Therefore, a univariate gamma regression model is sufficient for modeling precipitation data at each season.

Given the precipitation data $\boldsymbol{z}_{is} = (z_{i1s}, \ldots, z_{iTs})'$ at location i and season s, the gamma regression model, using a log link function such that the expected value (μ_{ijs}) is always positive, is provided as,

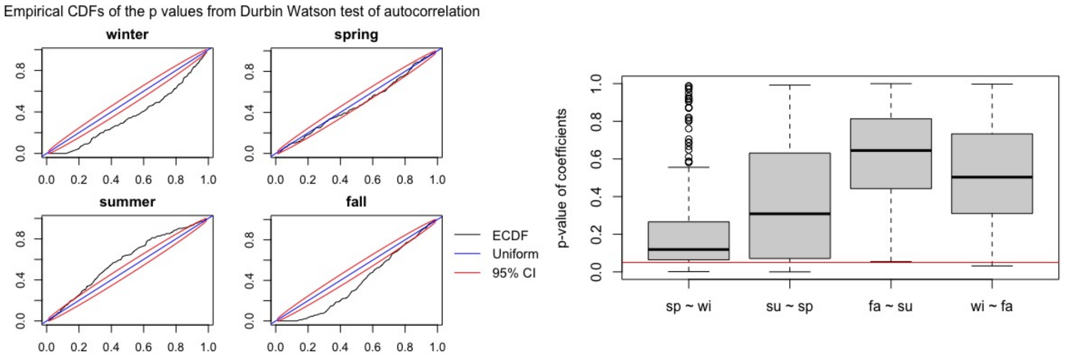
$$E(\boldsymbol{z}_{ijs}) = \exp(\boldsymbol{X}_i' \boldsymbol{\beta}_{is}) = \alpha_{is} \boldsymbol{\theta}_{is}, \tag{6.4}$$

such that $\boldsymbol{\theta}_{is} = \frac{\exp(\boldsymbol{X}_i' \boldsymbol{\beta}_{is})}{\alpha_{is}}$. $\boldsymbol{X}_i$ is the design matrix with j-th column input $(1, t_j)$. The log-likelihood is given as,

$$\begin{split} L(\boldsymbol{\beta}_{is}, \alpha_{is} | \boldsymbol{Z}_{is}) = \sum_j [(\alpha_{is} - 1)\log(z_{ijs}) - \log\Gamma(\alpha_{is}) - \\ \alpha_{is}(\boldsymbol{X}_{ij}^T \boldsymbol{\beta}_{is} - \log(\alpha_{is})) - \frac{z_{ijs}\alpha_{is}}{\exp(\boldsymbol{X}_{ij}^T \boldsymbol{\beta}_{is})}], \end{split} \tag{6.5}$$

which is maximized to estimate $\boldsymbol{\beta}_{is}$ and α_{is} from the data [183]. Numerical methods can be implemented to estimate the maximum likelihood estimates (MLEs). In our model, we fitted the data with the GLM function in *R* using IWLS method and assessed the goodness of fit of gamma distribution by chi-square test of the deviances. Using MLE theory and under mild conditions, it is shown that $\hat{\boldsymbol{\beta}}_{is}$ is asymptotically $N_2(\boldsymbol{\beta}_{is}, (\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}/\alpha_{is})$, where α_{is} here plays a similar role of σ^2 in general linear models [183]. Therefore, the slope parameter β_{is1} in the gamma regression model can be interpreted as the amount of precipitation changes by a factor of $exp(\beta_{is1})$ in every decade, at location i and season s.

Figure 6.6 presents the heat maps of estimated regression slopes ($\hat{\beta}_{is1}, i = 1, \ldots, n, s = 1, 2, 3, 4$) and corresponding Z-scores of precipitation changing. Unlike the consistent increasing trend with historical temperature data, the changing patterns of precipitation varied in four seasons and areas in the state. For example, in Spring, the majority of the states experienced a drier situation in historical periods, while in Fall, most parts in the state have more precipitation, though the effect is not significant (small Z-scores). In fact, overall changing effects of precipitation estimated from the uncorrelated temporal models are not significant as shown in Figure 6.6 with relatively small Z-scores.



(a) (b)

Figure 6.5. Temporal model selection for CA precipitation data. (a) Empirical CDFs of the p values from the Durbin Watson test of autocorrelation. The figure shows that within-seasonal autoregressive model are not needed. (c) Boxplots of coefficient p values from fitting linear regression between consecutive seasons. This figure proves that cross-seasonal correlations are not significant.

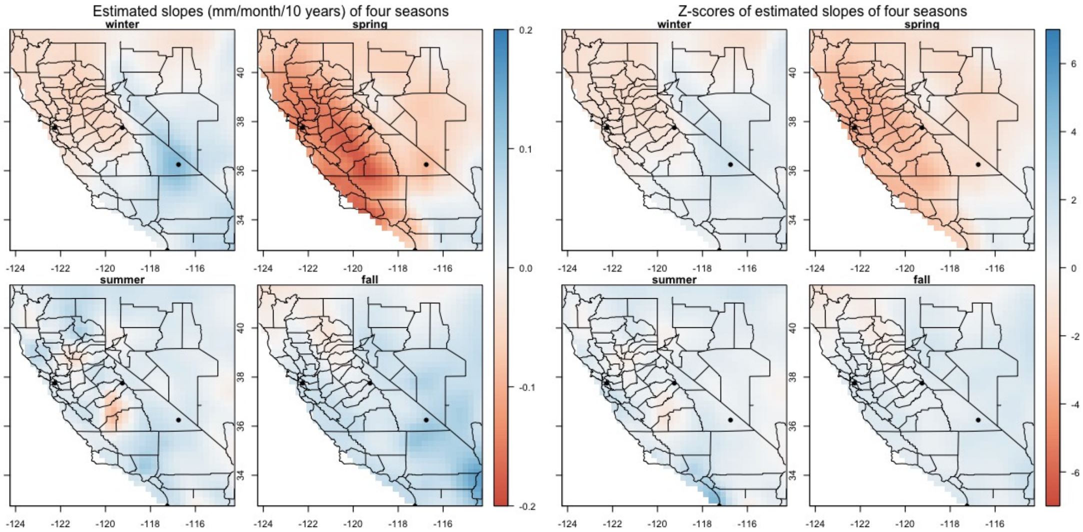

Estimated slopes (mm/month/10 years) of four seasons
Z-scores of estimated slopes of four seasons
winter
spring
summer
fall
-124
-122
-120
-118
-116
34
36
38
40
0.2
0.1
0.0
-0.1
-0.2
6
4
2
0
-2
-4
-6

Figure 6.6. Heat maps of estimated slopes (left) and corresponding Z-scores (right) of precipitation change (mm/month) in a decade for four seasons from GR models, fitted individually at each location.

6.5 Spatial-temporal modeling

6.5.1 Geographically weighted temporal model

With the basic temporal modeling framework, we further considered to extend the model by incorporating geographical correlations. In addition, temporal models disregard covariates related with topological features, such as elevation measures, which may influence climate changing rates. Therefore, we combined proposed MTSR models with framework from geographically weighted regression (GWR), which provides a means of incorporating spatial heterogeneity in linear regression models and modeling spatially varying relations [21]. For scalar outcome and predictor variables, a general form of GWR is given as,

$$y_i = \boldsymbol{x}_i' \boldsymbol{\beta}_i + \varepsilon_i, \tag{6.6}$$

where y_i is the outcome variable at location i (with coordinates $[u_i, v_i]$); $\boldsymbol{x}_i$ is a p-dimensional covariate vector with corresponding coefficients $\boldsymbol{\beta}_i$; ε_i are the normally distributed random noise terms with mean 0 and a common variance σ^2.

In fact, the GWR considers local likelihood and makes a point-wise calibration, i.e. given a specific point $[u_i, v_i]$ in geographical space, the corresponding coefficients $\boldsymbol{\beta}_i$ are more likely to be affected by nearer observations than those farther away. The estimation expression for it is provided as,

$$\hat{\boldsymbol{\beta}}_i = (X'W_iX)^{-1}X'W_i\boldsymbol{y}, \tag{6.7}$$

where X is p-by-n matrix of all covariates stacked by columns; $\boldsymbol{y} = (y_1, \ldots, y_n)'$ is the outcome vector; W_i is a n-by-n diagonal weight matrix denoting the local weight of each observed data for point i.

Kernel functions were commonly considered for constructing the weight matrix W_i, such as (a) linear kernel, (b) bisquare kernel, (c) exponential kernel and (d) gaussian kernels, with the following forms, respectively,

- Linear kernel: $w_{ij} = \begin{cases} 0, & \text{if } d_{ij} \geq l_i \\ 1 - (\frac{d_{ij}}{l_i}), & \text{otherwise} \end{cases}.$
- Bisquare kernel: $w_{ij} = \begin{cases} 0, & \text{if } d_{ij} \geq l_i \\ (1 - (\frac{d_{ij}}{l_i})^2)^2, & \text{otherwise} \end{cases}.$
- Exponential kernel: $w_{ij} = \exp[\frac{-d_{ij}}{l_i}]$.
- Gaussian kernel: $w_{ij} = \exp[-(\frac{d_{ij}}{l_i})^2]$.

where d_{ij} is distance between all observed locations j and a given point i. In our study, the distance was defined as the geometrical distance between grid points, since the location points were interpolated on a common latitude–longitude grid in the study. l_i can be interpreted as either the bandwidth in Gaussian/exponential cases or a threshold distance in linear/bisquare cases, which may be constant or varying among all i's.

Types of kernel function and values of bandwidth/threshold were determined through exploratory regression models for all pairs of locations i and j, using pairwise correlations of residuals from the previous regression models as outcomes and distances as predictors. Let the $\hat{\rho}_{ij}$ be the estimated Pearson correlation between residuals at location i and j, we tested the associations using four types of regression models correspondingly, (a) $\rho_{ij} \sim d_{ij}$, (b) $\rho_{ij} \sim d_{ij}^2$, (c) $\log(\rho_{ij}) \sim d_{ij}$ and (d) $\log(\rho_{ij}) \sim d_{ij}^2$, accounting for four kernel types. Results showed that a linear kernel is appropriate for modeling spatial structure in both temperature and precipitation data. Then an optimal threshold is specified for each time period based on a 10-fold cross-validation (CV) criterion. In addition, we further determined that an adaptive l_i is not necessary. Results related with the kernel selection can be found in Supplementary Materials. In practice, because the GWR provides unbiased estimators compared with uncorrelated models, misspecification of the kernels won't affect the estimation results, which is considered as an advantage of its implementation.

6.5.2 GWMTSR model for temperature

We extended the scalar GWR framework to fit our multivariate time series data and proposed the geographically weighted MTSR (GWMTSR) model. Let $Y_i = (\boldsymbol{y}_{i1}, \ldots, \boldsymbol{y}_{iS})'$ be the $T-1$ be the outcome matrix by stacking all $\boldsymbol{y}_{is} = (y_{i1s}, \ldots, y_{i(T-1)s}), i = 1, \ldots, n, s = 1, \ldots, 4$ in Equation 6.3. Let X_i^w be location-specific design matrix for deterministic process in this weighted model. Specifically, compared with the MTSR model, here we had the elevation h_i included as an additional covariate and the design matrix can be written as $\boldsymbol{X}_i^w = \begin{pmatrix} 1 & t_2 & h_i \\ \ldots & \ldots & \ldots \\ 1 & t_T & h_i \end{pmatrix}'$, for $i = 1, \ldots, n$. The model construction then has a similar format as in Equation 6.3,

$$\boldsymbol{Y}_i = \boldsymbol{X}_i^{w\prime}\boldsymbol{\beta}_i + \boldsymbol{Y}_i^{(1)}\boldsymbol{b}_i + \boldsymbol{\varepsilon}_i, \quad \boldsymbol{\varepsilon}_i \sim N_4(0, \Sigma_i), \tag{6.8}$$

where $\boldsymbol{\beta}_i$ is a 3×4 parameter matrix for deterministic covariates, with row elements $\boldsymbol{\beta}_{is}$. $\boldsymbol{b}_i = \text{diag}(b_{is})$ is a 4×4 diagonal matrix for quantifying the stochastic processes. The weight matrix is defined as $W_i = diag(\boldsymbol{w}_i) \bigotimes I_T$, where $\bigotimes$ is the kronecker product. Denote $\tilde{\boldsymbol{y}}_s = vec(\boldsymbol{y}_{1s}, \ldots, \boldsymbol{y}_{ns})$ and $\boldsymbol{S}_i^w = (\boldsymbol{X}_i^w, \boldsymbol{Y}_i^{(1)})$, the estimated coefficient of each season is computed as,

$$\begin{aligned} \hat{\boldsymbol{B}}_{is} = (\hat{\boldsymbol{\beta}}_{is}, \hat{\boldsymbol{b}}_{is}) &= ((\boldsymbol{S}^w)' W_i \boldsymbol{S}^w)^{-1} (\boldsymbol{S}^w)' W_i \tilde{\boldsymbol{y}}_s \\ &= \left(\sum_{l=1}^{n} w_{il} (\boldsymbol{S}_l^w)' \boldsymbol{S}_l^w \right)^{-1} \left(\sum_{l=1}^{n} w_{il} (\boldsymbol{S}_l^w)' \tilde{\boldsymbol{y}}_s \right). \end{aligned} \tag{6.9}$$

Consequently, the corresponding variance-covariance matrix of estimated coefficients has the form $\hat{\Sigma}_{\boldsymbol{B}_{is}} = C_i C_i' \hat{\Sigma}_i$, where $C_i = ((\boldsymbol{S}^w)' W_i \boldsymbol{S}^w)^{-1} (\boldsymbol{S}^w)' W_i$. Meanwhile, it is still valid to derive the limiting trend parameter under the GWMTSR setting, since the weighted estimates are obtained from the same model.

Model comparison is based on corrected Akaike Information Criterion (AIC) [2, 78], which is defined as,

$$\text{AIC}_c = 2n \log(\frac{\sum_i \hat{\sigma}_{is}^2}{n}) + \frac{n + \text{Tr}(R_s)}{n + 2 - \text{Tr}(R_s)}, s = 1, 2, 3, 4, \tag{6.10}$$

where $\hat{\sigma}_{is}$ is the s- diagonal element of $\hat{\Sigma}_i$; R_s is the hat matrix with the form $R_s = (\boldsymbol{r}_{1s}, \ldots, \boldsymbol{r}_{ns})'$ and $\boldsymbol{r}_{is} = \boldsymbol{S}_i^w ((\boldsymbol{S}^w)' W_i \boldsymbol{S}^w)^{-1} (\boldsymbol{S}^w)' W_i$. We found that the GWMTSR model with elevation yields a smaller AIC value, indicating that adding topological information has improved the model fitting.

Figure 6.7 shows heat maps of GWMTSR estimated linear time trend and their corresponding Z-scores. Compared with the results from previous individually fitted MTSR models

(Figure 6.4), the weighted estimates produced unbiased estimators of these parameters but significantly reduced values of standard errors, especially in Spring and Summer. It can be further illustrated with Z-score heat maps in Figure 6.7 and the histogram for comparing the ranges of Z-scores from both models in Figure 6.8, both showing an overall increased Z-scores in GWMTSR models and demonstrating the necessity of incorporating the geographical relations. In addition, it further indicates that fitting a weighted regression generates a smoother estimated parameter space than estimates from uncorrelated MTSR models, which provides an additional advantage for statistical inference.

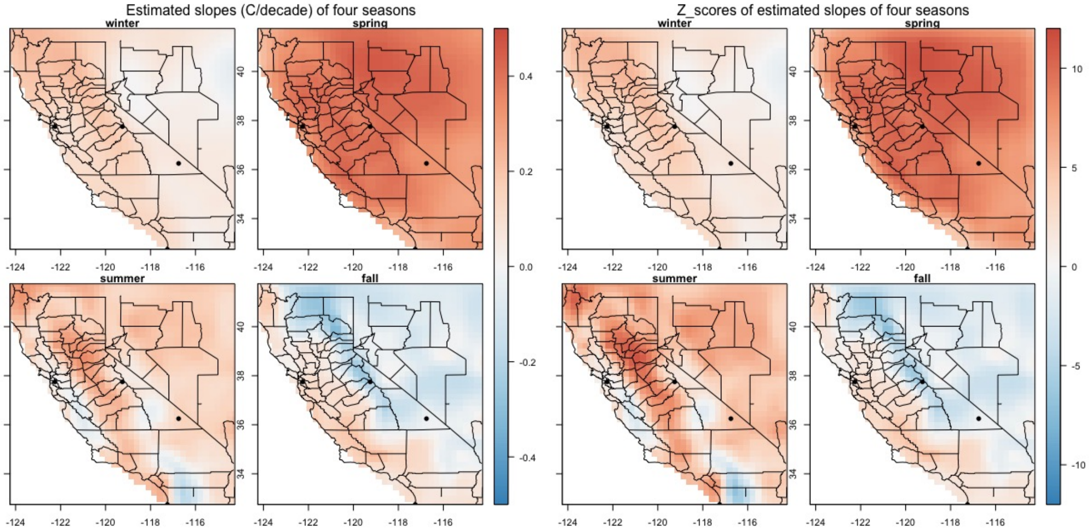


Figure 6.7. Heat maps of estimated linear time trend parameters (left) and corresponding Z-scores (right) of temperature change ($^{\circ}C$) in a decade of four seasons from GWMTSR models.

6.5.3 GWGR model for precipitation

We also incorporated the GWR strategy in GR models for precipitation, referred as the GWGR model. Using the selected weight matrix W_i and design matrix $\boldsymbol{X}_i^w$ at location i in Section 6.5.2, the log-likelihood of a GWGR model is given as,

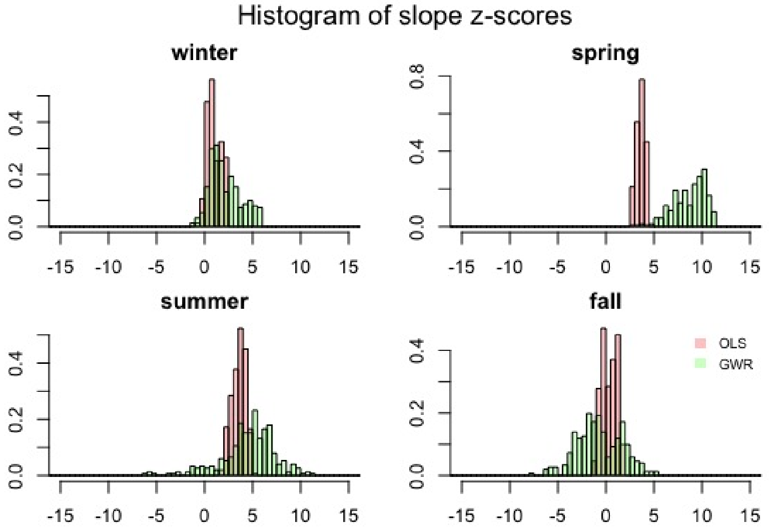


Figure 6.8. Histogram of Z-scores of linear time trend parameters from temperature MTSR and GWMTSR models in four seasons. This figure indicates increasing Z scores from the geographically weighted models. OLS = ordinary least squares estimates; GWR = geographically weighted estimates.

$$
\begin{aligned}
L(\boldsymbol{\beta}_{is}, \alpha_{is}|\boldsymbol{Z}_i) &\propto \sum_{l=1}^{n} w_{il} \sum_{j} L(\boldsymbol{\beta}_{is}, \alpha_{is}|\boldsymbol{Z}_{ij}) \\
&= \sum_{l=1}^{n} w_{il} \sum_{j} [(\alpha_{is}-1)\log(z_{ijs}) - \log\Gamma(\alpha_{is}) - \alpha_{is}((\boldsymbol{X}_i^w)'\boldsymbol{\beta}_{is} - \log(\alpha_{is})) \qquad (6.11) \\
&- \frac{z_{ijs}\alpha_{is}}{\exp((\boldsymbol{X}_i^w)'\boldsymbol{\beta}_{is})}].
\end{aligned}
$$

However, the MLE method could not produce solution analytically with Equation 6.11, thus the numerical optimization method using Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm was employed to estimate the MLEs of the model [7]. Figure 6.9 shows the heat maps of estimated regression slopes and corresponding Z-scores of precipitation using GWGR models. Similar to the results of fitting GWMTSR on temperature data, GWGR models also produce unbiased but more significant estimations of the slopes, indicating the robustness and benefits of the geographically weighted modeling in precipitation modeling. Figure 6.10 provides the histogram of Z-score distributions fitting GR and GWGR models on precipitation data, which further showing smaller standard errors were generated from the GWGR models.

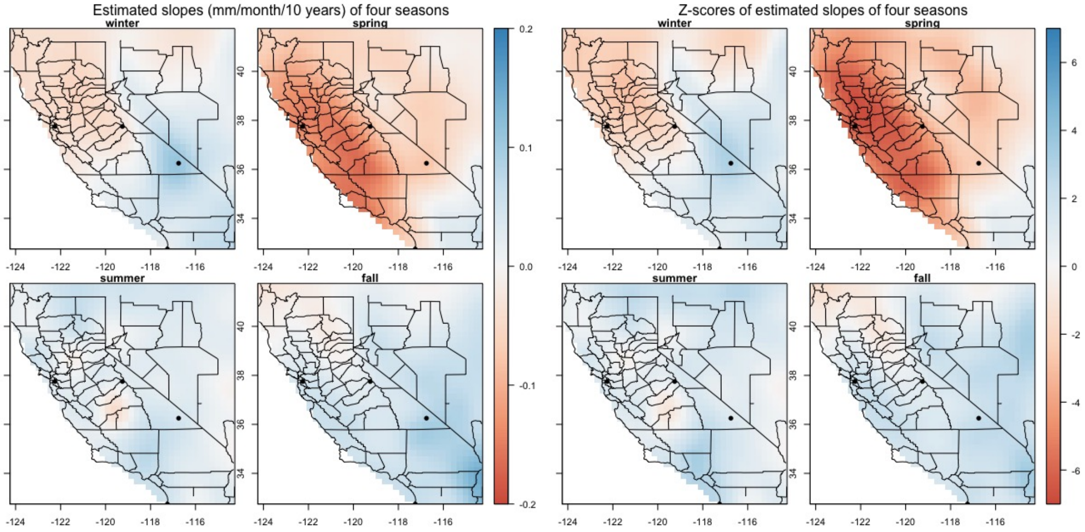

Estimated slopes (mm/month/10 years) of four seasons
Z-scores of estimated slopes of four seasons
winter
spring
summer
fall
0.2
0.1
0.0
-0.1
-0.2
6
4
2
0
-2
-4
-6
-124
-122
-120
-118
-116
34
36
38
40

Figure 6.9. Heat maps of estimated slopes (left) and corresponding Z-scores (right) of precipitation change (mm/month) in a decade for four seasons from GWGR models.

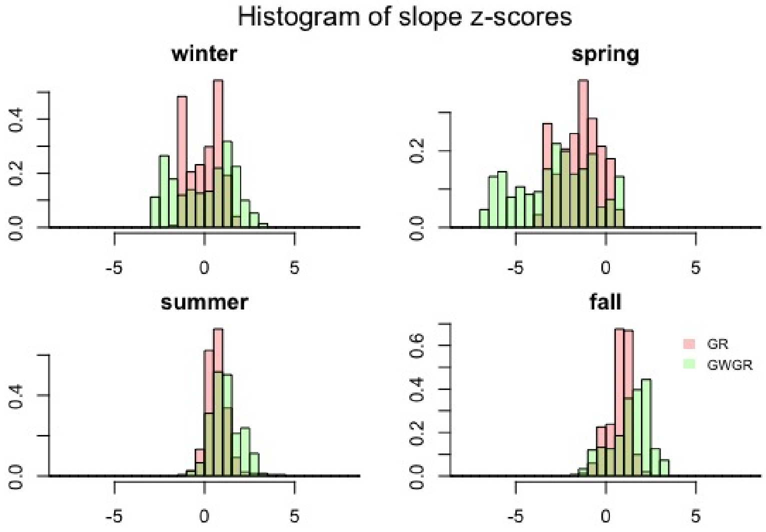


Figure 6.10. Histogram of Z-scores of precipitation change in a decade from GR and GWGR models in four seasons. This figure indicates increasing Z scores from the GWGR models. GR = gamma regression estimates; GWGR = geographically weighted gamma regression estimates.

6.5.4 Simulation Studies for Geographically Weighted Regression Models

To further illustrate the performance of GWR modeling and to compare it with non-weighted regression models, we simulated spatial time series data to conform with data structure of temperature and precipitation, respectively. The basis for generating spatially correlated data is the association between distance and residuals from models in Section 6.4. Specifically, for both temperature and precipitation, we simulated 100 datasets on a regular $n = 10 \times 10$ lattice, where correlations between grid points decay linearly in distance. The error term $\boldsymbol{\varepsilon}$ were generated from a multivariate normal distribution $N(\mathbf{0}, \Sigma)$, where Σ refers to the corvariance matrix constructed from the linearly decayed correlations. Specifically, the generating procedure for simulating temperature data (Scenario 1) is,

$$\boldsymbol{y}_i = \beta_{0i} + \beta_{1i}\boldsymbol{t}_i + \boldsymbol{\varepsilon}_i, \tag{6.12}$$

where $\boldsymbol{t}_i = \{1, 2, \ldots, T\}$ is a time sequence with length of $T = 50$. And for simulating precipitation (Scenario 2), we have,

$$\boldsymbol{z}_i = \exp\{\beta_{0i} + \beta_{1i}\boldsymbol{t}_i\} + \boldsymbol{\varepsilon}_i. \tag{6.13}$$

We simulated β_{0i} and β_{1i} based on the mean estimates and corresponding standard errors from previous results regarding temperature and precipitation, respectively. Several evaluation methods were employed to compare the performance of the two scenarios. The ability of estimating the slope surface is measured by the mean squared error (MSE) of the parameter β_{1i}, i.e. $\frac{1}{n}\sum_{i=1}^{n}(\beta_{1i} - \hat{\beta}_{1i})^2$, where $\hat{\beta}_{1i}$ is the estimated coefficient for location i from either geographically weighted or non-weighted models. The goodness of fit of models is evaluated with the MSE of fitted values $\hat{\boldsymbol{y}}_i$ or $\hat{\boldsymbol{z}}_i$ and simulated $\boldsymbol{y}_i$ or $\boldsymbol{z}_i$, defined as $\frac{1}{n*T}\sum_{i,t_i}(\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i)^2$ or $\frac{1}{n*T}\sum_{i,t_i}(\boldsymbol{z}_i - \hat{\boldsymbol{z}}_i)^2$. Lastly, we compare Z-score distribution of each model for assessing the significance of estimates.

Simulated data regarding the association between correlation and distance, as well as the slope surfaces is visualized in Appendix B, Figure **??**. Figure **??** depicts the MSE values for parameter surfaces of the 100 simulations, showing that the geographically weighted models in both scenarios generate more accurate estimates, compared with non-weighted models. Regarding the MSE values for assessing the goodness of fit, the weighted or non-weighted models produce very similar results in two scenarios, as shown in Figure **??**. Figure **??** conforms with our findings in Figure 6.8 and 6.10, that is, the weighted models for both temperature and precipitation data can enhance the significance of parameter estimates.

6.6 Statistical Inference

6.6.1 CoPE sets

The proposed GWMTSR and GWGR models provide the estimated changing parameters of temperature and precipitation, along with the Z-scores showing pointwise significance of these estimated parameters. However, as it was discussed in 6.1, pointwise tests are not valid when the data are densely sampled on a fine grid, without considering FWER in simultaneous testing. To account for this limitation, we introduced the CoPE method as a effective and feasible substitution.

The CoPE sets approach was proposed by Sommerfeld et al. (2018) [165], to address the inference problem when multiple tests across the spatial domain need to be performed simultaneously. The main idea of this method is to set confidence bounds for regions reflecting time variability. Specifically, for a target function $\mu : S \to \mathbb{R}$ on a spatial domain $S \subset \mathbb{R}^d$, an excursion set $\mu(s)$ above a fixed threshold c is defined as $A_c(\mu) := \{s \in S : \mu(s) \geq c\}$. And based on an estimate $\hat{\mu}_n(s)$ of $\mu(s)$, the confidence regions $\hat{A}_c^{\pm} := \{\hat{A}_c^{+} \subset A_c \subset \hat{A}_c^{-}\}$ are obtained to hold asymptotically above a desired level $1-\alpha$. Such excursion sets $\hat{A}_c^{\pm}$ are referred as CoPE sets. In our analysis, $\hat{\mu}(s)$ refers to each estimated trend parameters from previously fitted models. A straightforward way of interpreting the CoPE results is that, given a significance value α and a fixed threshold level c, the estimated CoPE sets denote that with simultaneous approximate probability of $1-\alpha\%$, slopes in this region are no smaller than the region given by the $\hat{A}_c^+$ boundary and no larger than indicated by the $\hat{A}_c^-$ boundary. The confidence sets are constructed using the multiplier bootstrap method. Detailed background and theories related with the CoPE method can be found in the original paper by Sommerfeld et al. (2018) [165] and we mainly focused on the interpretation of the generated confidence sets, which is more instructive and beneficial for climate changing inference.

Before performing statistical inference, an important aspect is to assess whether certain assumptions are satisfied. One advantage about the CoPE method is that it only requires mild assumptions about the data structure. French et al. 2017 [50] summarized three assumptions need to be checked, including the continuity and Gaussianity of the estimated coefficients in the space domain, as well as independence in time domain of the standardized errors at each location. The first assumption can be assessed with heat maps of estimated coefficients over the space, which should look smooth enough for treating the estimated coefficients fields as continuous. For temperature data, the Gaussianity assumption is valid as long as $\boldsymbol{\varepsilon}_{is}$ at each location, are Gaussian, which can be evaluated similarly using the Shapiro-Wilk test. For precipitation data modeled with Gamma regression models, the Shapiro-Wilk test was performed on deviances at each location for assessing the Gaussianity of estimated parameters. Lastly, the temporal independence was evaluated using the Ljung-Box tests [104], which tests whether the first l autocorrelations in a time series are significant. We took $l = 10$ following the recommendation of Hyndman and Athanasopoulos [66]. The assumption of independence is confirmed if the empirical CDF of the p values is approximately follow with the CDF of a standard uniform distribution. Figures of the assumption assessment were included in the Supplementary Materials.

In this work, considering that geographically weighted models yielded smoother parameter space and smaller standard errors in both temperature and precipitation data, we only implemented the CoPE inference on results from these models. Meanwhile, because the proposed GWMTSR for temperature data includes cross-season effects, corresponding linear time trend coefficients $\boldsymbol{\beta}_{i1}$ can not be interpreted similarly as the slope in a trend analysis. To bridge the gap of interpretation between our proposed model and simple linear regression model in trend analysis, we derived limiting estimates of temperature slope parameters $\boldsymbol{v}_i$ and corresponding standard errors from MTSR and GWMTSR models. $\boldsymbol{v}_i$'s represent the velocity of temperature changing at location i when t is large, and thus could be interpreted as a limiting estimator of the slope parameter in a simple linear trend model. The CoPE approach was then implemented on these estimated limiting slopes from both historical and future data in CA, CO and KS, for comparing the climate changing effects in differing time intervals and regions.

6.6.2 CoPE Sets for Temperature

We began with the CoPE set inference related to temperature changing slopes $\boldsymbol{v}_i$. The significance level α was fixed at 0.1. As it is shown in Figure 6.2, the slopes of temperature in historical periods were less than that in the future. Therefore, we selected two different sets of testing threshold levels c to reflect the discrepancy between these two periods. Bonferroni correction was implemented to account for the number of comparisons (n_t) when performing simultaneous testing, which resulted in an adjusted significance level with value $\alpha = 0.1/n_t$. Specifically, the level set for historical data was specified as $\{0.1, 0.15, 0.2\}°C/$decade and $\{0.5, 0.55, 0.6\}°C/$decade for future data.

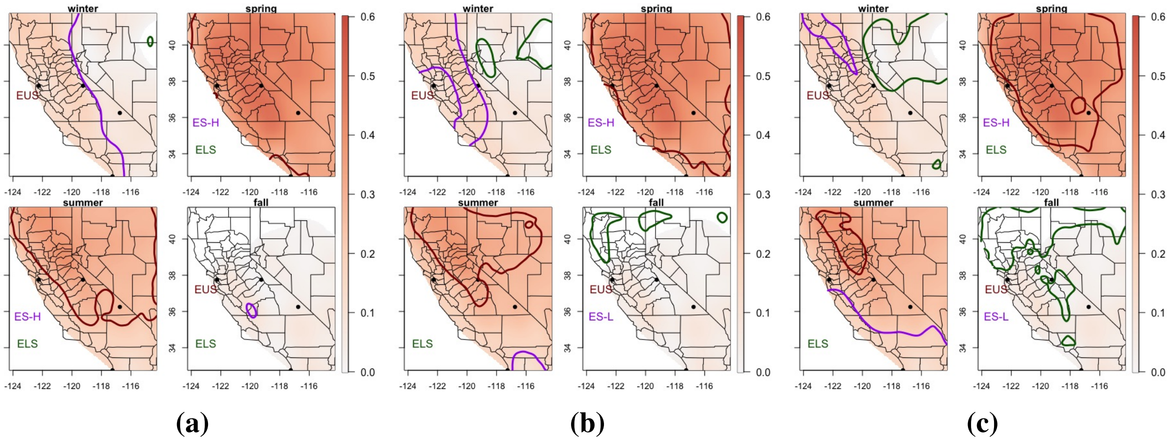
Figure 6.11, 6.12 and 6.13 provide the results of CoPE sets, constructed for slopes in historical temperature data in CA, CO and KS, respectively. The contours of lower set $\hat{A}_c^-$ (green boundary) and upper set $\hat{A}_c^+$ (red boundary) only appear if they are not empty sets, otherwise they were labeled as ELS and EUS. Take $c = 0.15°C$ as an illustrative example, these CoPE plots have the interpretation that with probability 0.9, the region with temperature increased 0.15 $°C$ per decade is no smaller than the region bounded by the red boundary ($\hat{A}_c^+$) and no larger than illustrated by the green boundary ($\hat{A}_c^-$). Therefore, these generated CoPE sets and confidence regions are able to provide assessment of climate change at all locations simultaneously.

Figure 6.11 shows that the historical temperature in CA was consistently increasing in all four seasons with varying rates. Temperature in Spring and Summer significantly increased more than 0.2 $^\circ C$ per decade, for most of areas were either captured within the upper bound or with higher values than that. Temperature changing was much milder and less notable in Fall and Winter, which in fact was in accordance with the relatively smaller Z-score results shown in Figure 6.7. In Figure 6.11(b), for instance, northern areas in Fall and partial areas in Nevada in Winter that are bounded by the lower set (green boundaries), indicating no more than 0.15 $^\circ C$ increase per decade. In CO, Figure 6.12 displays that in a decade, temperature in most of areas significantly increased 0.15 $^\circ C$ in Spring and 0.1 $^\circ C$ in Summer. And in KS, only Summer temperature significantly increased 0.1 $^\circ C$ per decade. Therefore, CA seemed to have a more drastic temperature changing effect in the past fifty-six years than the other two states. It is also interesting to see that Winter temperature even decreased in CO and KS.

We then implemented the proposed GWMTSR models on future data, using the same model format and weight matrix selected from the historical model fitting in each state. Figure 6.14, 6.15 and 6.16 present the CoPE sets results for future temperature changing slopes, with threshold levels set as $c = \{0.5, 0.55, 0.6\}^\circ C/$decade. These constructed CoPE sets further confirm that more extreme temperature increase happens and will continue in the future. Specifically, in most of the areas of CA, the increase in a decade will be over 0.5 $^\circ C$ in Winter and over 0.6 $^\circ C$ in Summer. Compare with their historical trends, inland states will witness relatively more significant temperature increase, for instance, temperature of Springs in both CO and KS are likely to have a over 0.55 $^\circ C$ per decade after 2005.

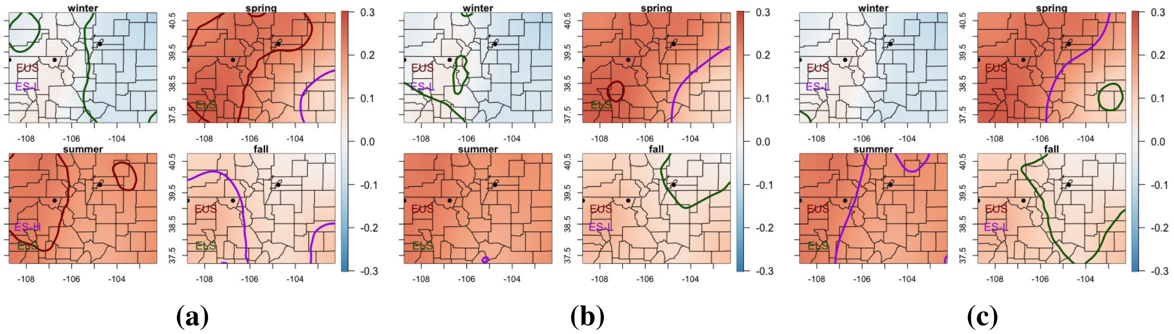
6.6.3 CoPE Sets for Precipitation

The CoPE method was also performed on precipitation results. To conform with the colors in heat maps which reflect the direction of precipitation changing, we use green boundaries to represent the upper bounds indicating more precipitation and red boundaries for lower bounds. In addition, because we assumed a log link in the gamma models, these estimated slopes were interpreted as a factor of $\exp(\beta_i s1)$ change in the precipitation in a decade. Thus, a negative slope



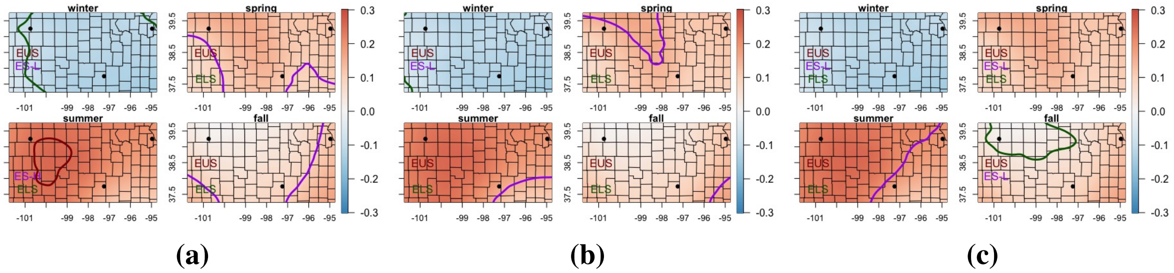
(a) (b) (c)

Figure 6.11. Heat maps of estimated historical temperature changing slopes (°C/decade) of four seasons in CA and corresponding CoPE sets computed at three prespecified levels (a) 0.1, (b) 0.15 and (c) 0.2. The uncertainty in the excursion set estimates $\hat{A}_c$ (purple boundary) is captured by the CoPE set $\hat{A}_c^+$ (red boundary) and $\hat{A}_c^-$ (green boundary) with confidence 0.9. Empty lower (ELS) and upper sets (EUS) are shown in the figures with the same color, representing empty lower and upper bounds. ES-L denotes all estimated slopes are smaller than the specified level c and ES-H denotes the opposite effect.



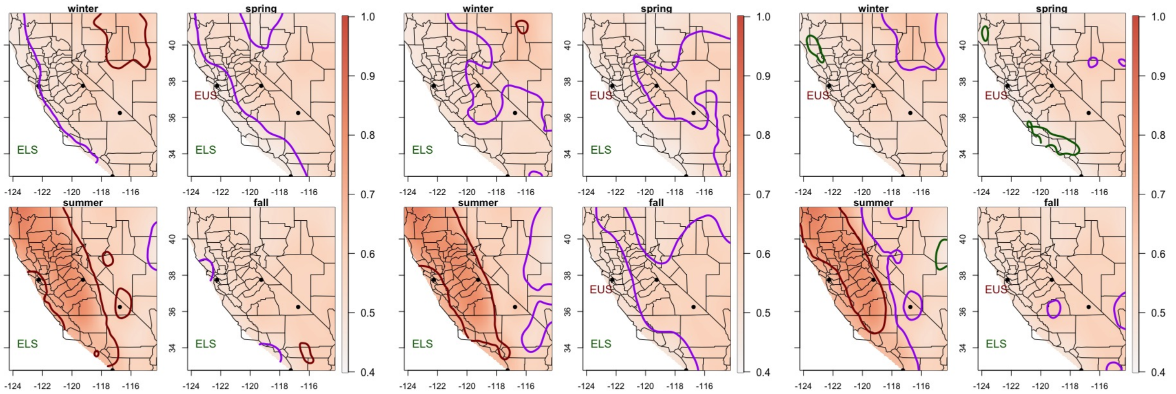
(a) (b) (c)

Figure 6.12. Heat maps of estimated historical temperature changing slopes (°C/decade) of four seasons in CO and corresponding CoPE sets computed at three prespecified levels (a) 0.1, (b) 0.15 and (c) 0.2.



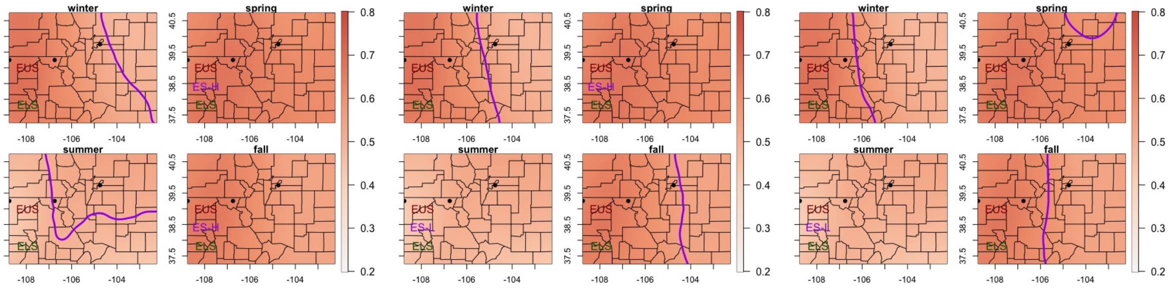
(a) (b) (c)

Figure 6.13. Heat maps of estimated historical temperature changing slopes (°C/decade) of four seasons in KS and corresponding CoPE sets computed at three prespecified levels (a) 0.1, (b) 0.15 and (c) 0.2 (c).



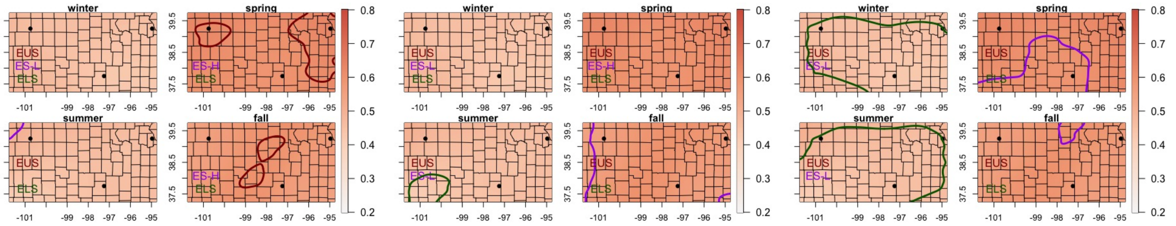
(a) (b) (c)

Figure 6.14. Heat maps of estimated future temperature changing slopes (°C/decade) of four seasons in CA and corresponding CoPE sets computed at three prespecified levels (a) 0.5, (b) 0.55 and (c) 0.6.



(a) (b) (c)

Figure 6.15. Heat maps of estimated future temperature changing slopes (°C/decade) of four seasons in CO and corresponding CoPE sets computed at three prespecified levels (a) 0.5, (b) 0.55 and (c) 0.6.



(a) (b) (c)

Figure 6.16. Heat maps of estimated future temperature changing slopes (°C/decade) of four seasons in KS and corresponding CoPE sets computed at three prespecified levels (a) 0.5, (b) 0.55 and (c) 0.6.

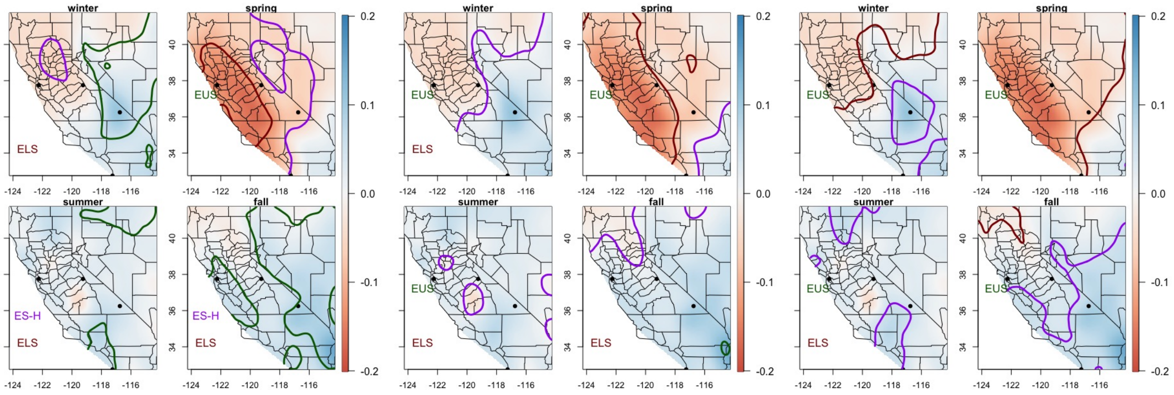
would mean decreasing precipitation and the corresponding CoPE sets have the interpretation that with a simultaneous approximate probability of 90%, the area with the decreasing pattern is no smaller than the region given by the red boundary ($\hat{A}_c^-$) and no larger than indicated by the green boundary ($\hat{A}_c^+$). In addition, the zero level value is set to find regions where the change of precipitation is nonzero with probability 0.9.

Figure 6.17, 6.18 and 6.19 show the results of precipitation modeling from GWGR models using historical data in CA, CO and KS, respectively, estimated with three fixed levels $c = \{-0.05, 0, 0.05\}$ mm/decade. It is noticed in Figure 6.17 that only Springs in CA experienced a significantly drier situation in the past fifty-six years, especially for those areas circled within the lower (red) boundaries in Figure 6.17(a) and (b). For instance, in SF, at least 5% decrease in precipitation could have happened in Springs with 90% confidence. While in Yosemite, we could only conclude that the precipitation was decreasing (less than zero). For other seasons in (a), on the contrary, we inferred with 90% confidence that the decrease of precipitation wasn't more than 5%, especially for those areas circled by upper (green) boundaries. Figures in KS and CO can be interpreted in a similar way, while no significant effect was seen from them.

Figure 6.20, 6.21 and 6.22 display the precipitation changing in the future. In CA, more precipitation is expected in most of areas of CA from Figure 6.20(b) except for Springs. In Springs, drier situations still exist, especially within the area bounded with the lower (red) set, while the estimated decreasing rate is more gentle than that in historical period. A significant increasing pattern of Summer precipitation can be seen in Figure 6.20(c). Most of areas, including San Diego, Yosemite and Death Valley, could have an over 5% increase in precipitation per decade. Similarly as the patterns of temperature change, CO and KS present more notable increasing of precipitation than the past, especially in Winters and Summers.

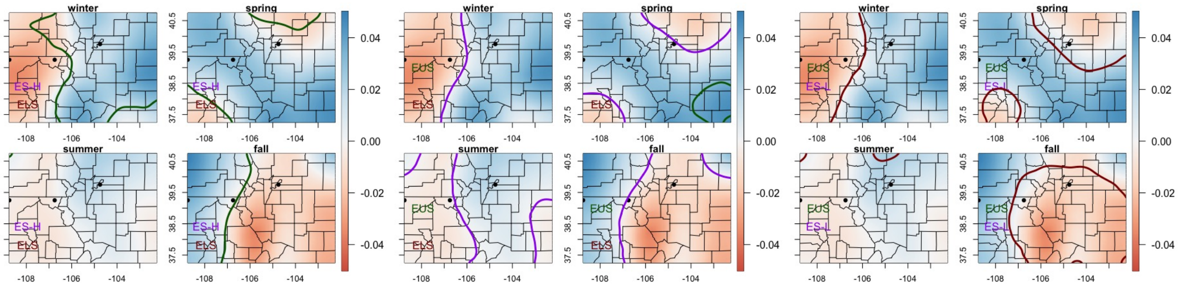
6.7 Discussion

In this paper, we proposed a series of geographically weighted multivariate time series regression models for temperature and precipitation using data from NA-CORDEX program. In addition, the CoPE method was implemented to perform simultaneous inference of estimated climate change rates across the spatial domain. Multivariate time series regression



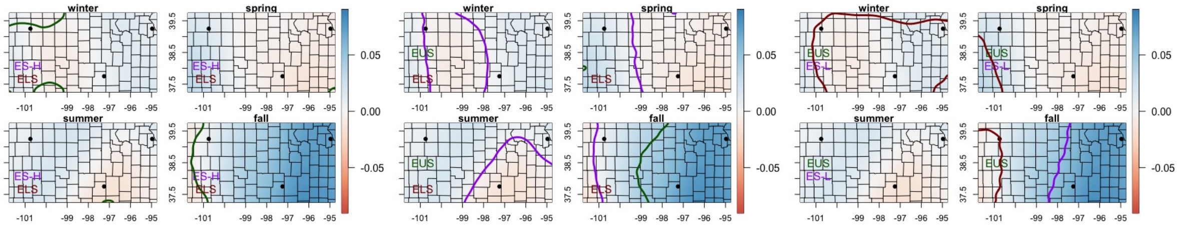
(a) **(b)** **(c)**

Figure 6.17. Heat maps of historical precipitation changing slopes (*mm*/month) of four seasons in CA and corresponding CoPE sets computed at three prespecified levels (a) -0.05, (b) 0 and (c) 0.05. The uncertainty in the excursion set estimates $\hat{A}_c$ (purple boundary) is captured by the CoPE set $\hat{A}_c^+$ (green boundary) and $\hat{A}_c^-$ (red boundary) with confidence 0.9. Empty lower (ELS) and upper sets (EUS) are shown in the figures with the same color, representing empty lower and upper bounds. ES-L denotes all estimated slopes are smaller than the specified level c and ES-H denotes the opposite effect.



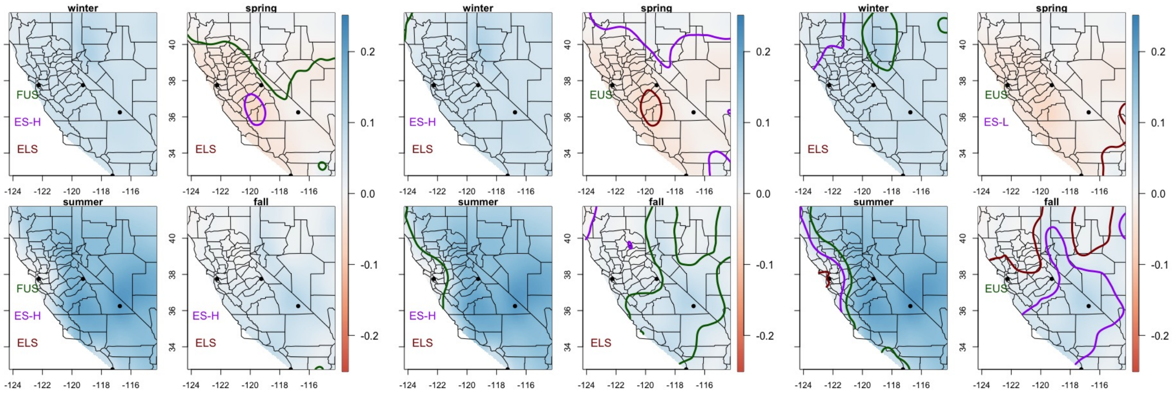
(a) **(b)** **(c)**

Figure 6.18. Heat maps of historical precipitation changing slopes (*mm*/month) of four seasons in CO and corresponding CoPE sets computed at three prespecified levels (a) -0.05, (b) 0 and (c) 0.05.



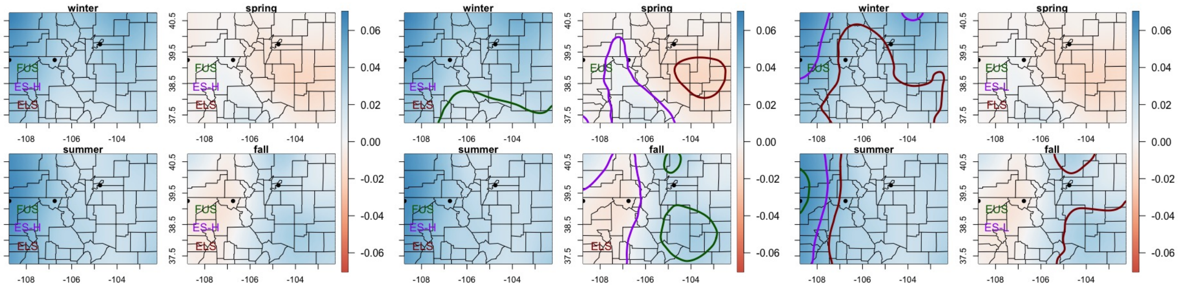
(a) **(b)** **(c)**

Figure 6.19. Heat maps of historical precipitation changing slopes (*mm*/month) of four seasons in KS and corresponding CoPE sets computed at three prespecified levels (a) -0.05, (b) 0 and (c) 0.05.



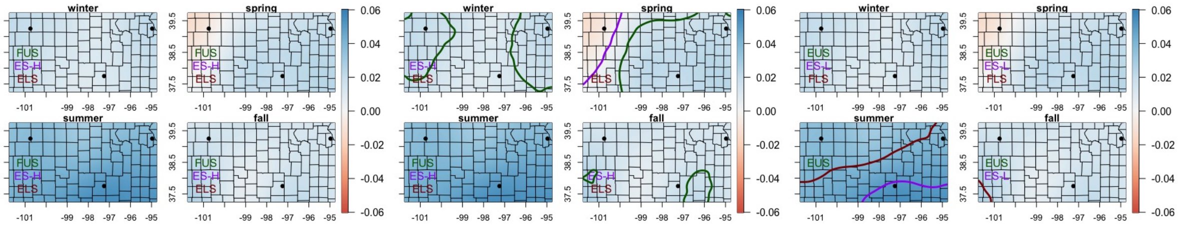
(a) **(b)** **(c)**

Figure 6.20. Heat maps of estimated future precipitation changing slopes (*mm*/month) of four seasons in CA and corresponding CoPE sets computed at three prespecified levels (a) −0.05, (b) 0 and (c) 0.05.



(a) **(b)** **(c)**

Figure 6.21. Heat maps of estimated future precipitation changing slopes (*mm*/month) of four seasons in CO and corresponding CoPE sets computed at three prespecified levels (a) −0.05, (b) 0 and (c) 0.05.



(a) **(b)** **(c)**

Figure 6.22. Heat maps of estimated future precipitation changing slopes (*mm*/month) of four seasons in KS and corresponding CoPE sets computed at three prespecified levels (a) −0.05, (b) 0 and (c) 0.05.

models were constructed for modeling climate data in each season, considering both within- and between-season effects. They were extended to GWMTSR and GWGR, which further include geographical correlation information in modeling multivariate temperature and precipitation sequences. Compared with fitting models independently at each location, these geographically weighted models generated smoother and more significant parameter spaces. These improvements further validated the applicability of CoPE methods, with which we were able to identify areas with significant climate change effects in different seasons and states. As a summary, our work provides a comprehensive two-step analysis procedure for climate change studies, by firstly constructing spatial-temporal models and followed with spatial inference on estimated parameters.

We typically implemented our proposed modeling and inference strategies on three representative states in the US, for comparing climate changing patterns in small regions with varying geographical features. CA is the most geographically diverse state among these three states, which also presented more significant climate changing effects in both temperature and precipitation, especially in the historical period (1950-2005). Using the CoPE method, we identified that most of areas in CA had experienced over 0.15 $^\circ C$ increase in temperature every ten years in Summers and Springs. It is also interesting to witness that the upper bounds gradually separates the coastal areas with the inland parts as we increase the specified testing levels (Figure 6.11). In addition, less precipitation was typically determined in Winters and Springs, which could potentially be associated with the increasing incidence of wildfires in CA in the first six months of the year recently. As for CO and KS, the regional warming temperature and precipitation changing patterns were less significant, illustrating a mitigated climate changing effect in inland areas in the past fifty-six years. On the contrary, results of future climate change (2006-2100) are more consistent in all three states, including significantly higher temperature and more precipitation in most of seasons. Therefore, though the impact of global warming might be delayed in inland areas, it is expected to happen in the near future.

For model construction, we innovatively incorporated the geographically weighted regression (GWR) models to account for the limitation in fitting regression models independently at each location, especially when data were densely recorded within a relatively small spatial domain. Furthermore, we extended the original GWR, which was designed for scalar spatial data, to our multivariate time series regression setting. In particular, we derived a GWMTSR model for temperature and a GWGR model for precipitation data, respectively. The implementation of these models shows promise in terms of a smoother parameter space, reduced spatial autocorrelation in residuals and improved interpretation of estimated coefficients in the spatial domain. These improvements were further validate our simulation studies, in which we mimic the data structure and modeling strategies for both temperature and precipitation.

In addition, we performed statistical inference on the significance of the estimated climate changing rates using the CoPE method (Sommerfeld et al., 2015), which provided simultaneous testing results over the whole spatial domain. Compared with pointwise inference, which was typically done separately at each location, the CoPE methods can control the familywise error rate. As it was discussed in French et al. (2017) [50], because pointwise inference does not make any adjustments for multiple comparisons, it in general detected more significant areas than the CoPE method. Meanwhile, the construction of CoPE sets and their corresponding confidence sets requires mild assumptions and is easy to apply, therefore, the useful CoPE method can be generalized to other types of data that involve with spatial inference, such as brain imaging or infectious disease studies.

Further research is needed to address several limitations. Firstly, we mainly provided a guidance for constructing statistical models and inference process for explaining climate change trend in our work and illustrate its performance with one example dataset. However, in NA-CORDEX program, there are 27 simulations in total and over twenty kinds of climate outcomes that could be further explored. A valid impacts research normally requires the use of the ensemble of simulations regarding the uncertainties and systematic biases included in climate projection. Therefore, it is suggested to test our proposed pipeline using other datasets, especially when the modeling and inference results need to be considered for polity making. In addition, we could consider to have more covariates that might affect the climate change included in our proposed spatial-temporal model, such as distance to the coast, density of population, etc. Last but not least, the CoPE method can be further extended to perform automatically simultaneous testing, without the need of prespecified set of levels and Bonferroni correction.

Appendix A

Additional Results for Chapter 3

A.1 Gibbs Sampler Details

A.1.1 Pólya–Gamma data augmentation approach for multinomial regression

For estimating class-specific parameters in the multinomial latent class regression, we implement algorithms using a family of Pólya–Gamma distribution, introduced by Polson et al.(2013)[143]. The main strategy of the algorithm is to implement Gaussian draws for generating regression parameters and the Pólya–Gamma draws are incorporated for single layer of latent variables. Compared with previous attempts with missing-data strategy to the logit model[72, 53], which is either approximate or complicated, the Pólya–Gamma method is efficient and simpler.

A random variable ω is said to follow Pólya–Gamma distribution with parameters $b > 0$ and $c \in \mathbb{R}$ if

$$f(\omega|b,c) = \frac{1}{\pi^2}\Sigma_{r=1}^{\infty}\frac{g_r}{(r-1/2)^2 + c^2/(4\pi^2)} \tag{A.1}$$

where $g_r \sim Gamma(b,1)$ and ω is denoted as $\omega \sim PG(b,c)$. Polson et al.(2013)[143] proved that for the Bayesian logistic regression model, the Pólya–Gamma family can yield a simple Gibbs sampler and the posterior distribution is a scale mixture of Gaussians. Based on the derivation procedure for binary outcomes, we extend the algorithm to a multinomial setting.

Suppose that a random variable $\omega_{ik}, i = 1,\ldots,n, k = 1,\ldots,K$, follows Pólya–Gamma distribution with parameters $(1,0)$. We denote it by $PG(\omega_{ik}|1,0)$ and assume that the multinomial model of Equation **??** holds. Following Holmes and Held(2006)[72], the likelihood for $\boldsymbol{\gamma}_k$ conditional upon $\boldsymbol{\gamma}_{-k}$, i.e. the parameter matrix with column vector $\boldsymbol{\gamma}_k$ removed, is

$$L(\boldsymbol{\gamma}_k|\boldsymbol{\gamma}_{-k}) = \prod_i^n [\frac{\exp(\boldsymbol{z}_i^T\boldsymbol{\gamma}_k - O_{ik})}{1+\exp(\boldsymbol{z}_i^T\boldsymbol{\gamma}_k - O_{ik})}]^{\mathbf{1}(C_i=k)}[\frac{1}{1+\exp(\boldsymbol{z}_i^T\boldsymbol{\gamma}_k - O_{ik})}]^{\mathbf{1}(C_i\neq k)} \tag{A.2}$$

where $O_{ik} = \log\sum_{k'\neq k}\exp(\boldsymbol{z}_i^T\boldsymbol{\gamma}_k)$. Based on the logistic regression form of conditional likelihood, the contribution of $\boldsymbol{z}_i$ to the likelihood in $\boldsymbol{\gamma}_k$ can be written as

$$\begin{aligned} L(\gamma_k) &\propto \prod_i^n \exp\{\kappa_{ik}(z_i\boldsymbol{\gamma}_k - O_{ik})\}\exp\{-\frac{(\boldsymbol{z}_i^T\boldsymbol{\gamma}_k - O_{ik})^2\omega_{ik}}{2}\}PG(\omega_{ik}|1,0) \\ &\propto \prod_i^n \exp\{\kappa_{ik}(\boldsymbol{z}_i^T\boldsymbol{\gamma}_k - O_{ik}) - \frac{(\boldsymbol{z}_i^T\boldsymbol{\gamma}_k - O_{ik})^2\omega_{ik}}{2}\} \\ &\propto \exp\{-1/2(S_k - (\boldsymbol{Z}\boldsymbol{\gamma}_k - O_k))'\Omega_k(S_k - (\boldsymbol{Z}\boldsymbol{\gamma}_k - O_k))\} \end{aligned} \tag{A.3}$$

where $\kappa_{ik} = \mathbf{1}(C_i = k) - 1/2$, $S_k = \{\kappa_{1k}/\omega_{1k},...,\kappa_{n_kk}/\omega_{n_kk}\}$, and $\Omega_k = diag(\{\omega_{ik}\}_{i=1}^{n_k})$. n_k is the number of subjects that belong to latent class k. $Z = \{z_1,...,z_n\}$ is the aggregated design matrix of class-related covariates. Providing the prior $\boldsymbol{\gamma}_k \sim N(m_{0k}, V_{0k})$, the posterior is given as

$$\begin{aligned} (\boldsymbol{\gamma}_k|\Omega_k) &\sim N(m_k, V_k) \\ (\Omega_k|\gamma_k) &\sim PG(1, Z\gamma_k - O_k) \end{aligned} \tag{A.4}$$

where $V_k = (Z'\Omega_kZ + V_{0k}^{-1})^{-1}$, $m_k = V_k(Z'\Omega_k(\kappa_k + \Omega_kC_k) + V_{0k}^{-1}m_{0k})$ and $O_k = \log\sum_{k'\neq k}\exp(Z\boldsymbol{\gamma}_k)$. Therefore, it allows for Gibbs sampling from the joint posterior distribution without appealing to analytic approximations to the posterior.

A.1.2 Gibbs Sampling Algorithm

- Sample random effects:

For $i = 1,\ldots,n$ and given $C_i = k$, we draw $\boldsymbol{a}_i$ from $N_{2L}(\boldsymbol{\mu}_{a_i}, \boldsymbol{\Sigma}_{a_i})$, where $\boldsymbol{\Sigma}_{a_i} = (T_i' T_i/\sigma_\varepsilon^2 + \boldsymbol{\Sigma}_k^{-1})^{-1}$ and $\boldsymbol{\mu}_{a_i} = (T_i'\tilde{\boldsymbol{y}}_i/\sigma_\varepsilon^2 + \boldsymbol{\Sigma}_k\boldsymbol{\alpha}_k)\boldsymbol{\Sigma}_{a_i}$. The derivation process is expressed as,

$$\begin{aligned} P(\boldsymbol{a}_i|\ldots) &\propto \exp\{-\frac{\sum_{l=1}^{L}\sum_{j=1}^{m_{il}}(y_{ijl}-\xi_{ijl})^2}{2\sigma_\varepsilon^2}\}\cdot\exp\{-\frac{(\boldsymbol{a}_i-\boldsymbol{\alpha}_k)^T\boldsymbol{\Sigma}_k^{-1}(\boldsymbol{a}_i-\boldsymbol{\alpha}_k)}{2}\} \\ &= \exp\{-\frac{1}{2}[(\tilde{\boldsymbol{y}}_i - T_i\boldsymbol{a}_{\boldsymbol{i}})^T(\tilde{\boldsymbol{y}}_i - T_i\boldsymbol{a}_{\boldsymbol{i}})/\sigma_\varepsilon^2 + (\boldsymbol{a}_i-\boldsymbol{\alpha}_k)^T\Sigma_k^{-1}(\boldsymbol{a}_i-\boldsymbol{\alpha}_k)]\} \end{aligned}$$

where $\xi_{ijl} = \boldsymbol{x}_{ijl}^T\boldsymbol{\beta}_l + a_{0il} + a_{1il}t_{ijl}$ and $\tilde{y}_{ijl} = \boldsymbol{y}_{ijl} - \boldsymbol{x}_{ijl}^T\boldsymbol{\beta}_l$; $\tilde{\boldsymbol{y}}_i$ is a stacked vector specified as $\tilde{\boldsymbol{y}}_i = (\tilde{y}_{i11}, \ldots\tilde{y}_{im_{i1}1}, \ldots, \tilde{y}_{i1L}, \ldots\tilde{y}_{im_{iL}L})^T$; T_i is a concatenated block diagonal matrix, specified as $T_i = \begin{pmatrix} (\boldsymbol{1}, \boldsymbol{t}_{i1}), 0, \ldots, 0 \\ 0, (\boldsymbol{1}, \boldsymbol{t}_{i2}), \ldots, 0 \\ \cdots \\ 0, 0, \ldots, (\boldsymbol{1}, \boldsymbol{t}_{iL}) \end{pmatrix}$ with $(\boldsymbol{1}, \boldsymbol{t}_{il}) = \begin{pmatrix} 1, t_{i1l} \\ 1, t_{i2l} \\ \cdots \\ 1, t_{im_{il}l} \end{pmatrix}$.

- Update class-specific covariance matrix:

For $k = 1, 2, \ldots K$, we draw $\boldsymbol{\Sigma}_k$ from Inverse-Wishart (IW) distribution, which includes an additional half-t prior as described in Section 3.3.2:

$$\begin{aligned} P(\boldsymbol{\Sigma}_k|\ldots) \propto& |\boldsymbol{\Sigma}_k|^{-n_k/2}\exp\{-\frac{1}{2}(\boldsymbol{a}_i-\boldsymbol{\alpha}_k)^T\boldsymbol{\Sigma}_k^{-1}(\boldsymbol{a}_i-\boldsymbol{\alpha}_k)\} \\ &\cdot|\boldsymbol{\Sigma}_k|^{\frac{-(\nu+2L+2)}{2}}\exp(-\frac{1}{2}trace(2\nu\Delta\boldsymbol{\Sigma}_k^{-1})) \end{aligned}$$

where Δ is a diagonal matrix with elements λ_l, which are assumed to be independently distributed with $Gamma(\frac{1}{2}, \frac{1}{\psi_l^2})$. Therefore, $\boldsymbol{\Sigma}_k, k = 1, \ldots, K$ and λ_l can be updated from:

$$\boldsymbol{\Sigma}_k \sim IW(n_k+\eta+2L-1, 2\eta\Delta+\sum_{i}^{n_k}(\boldsymbol{a}_i-\boldsymbol{\alpha}_k)(\boldsymbol{a}_i-\boldsymbol{\alpha}_k)')$$

$$\lambda_l \sim Gamma(\frac{\eta+2L}{2}, \frac{1}{\psi_l}+\eta(\boldsymbol{\Sigma}_k^{-1})_l)$$

where $(\Sigma_k^{-1})_l$ is the l^{th} diagonal element of the inverse of Σ_k and n_k is the number of subjects that belongs to latent class k.

- Update class-specific random effect parameters:

  For $k=1,2,...K$, we draw $\boldsymbol{\alpha}_k$ from $N_{2L}(\boldsymbol{\mu}_{\alpha_k},\boldsymbol{\Sigma}_{\alpha_k})$, where $\boldsymbol{\mu}_{\alpha_k}=\boldsymbol{\Sigma}_{\alpha_k}\boldsymbol{\Sigma}_k^{-1}\sum_{i=1}^{n}\boldsymbol{a}_i^{\mathbf{1}(C_i=k)}$ and $\boldsymbol{\Sigma}_{\alpha_k}=(n_k\boldsymbol{\Sigma}_k^{-1}+1/c\mathbf{I})^{-1}$.

- Update fixed effects:

  For $l=1,2,...,L$, we update $\boldsymbol{\beta}_l$ from $N_{Q_x}(\boldsymbol{\mu}_{\beta_l},\boldsymbol{\Sigma}_{\beta_l})$, where $\boldsymbol{\mu}_{\beta_l}=\boldsymbol{\Sigma}_{\beta_l}\sum_{i=1}^{n}\boldsymbol{x}_{il}^T\boldsymbol{y}_{il}/\sigma_\varepsilon^2$ and $\boldsymbol{\Sigma}_{\beta_l}=(\sum_{i=1}^{n}\boldsymbol{x}_{il}^T\boldsymbol{x}_{il}/\sigma_\varepsilon^2+1/c\mathbf{I})^{-1}$. Here $\boldsymbol{x}_{il}=(\boldsymbol{x}_{i1l},\ldots,\boldsymbol{x}_{im_{il}l})$ and $\boldsymbol{y}_{il}=(y_{i1l},\ldots,y_{im_{il}l})$.

- Update variance parameter:

  Draw σ_ε^2 from Inverse-Gamma($\delta_1+\frac{\sum_{i,l}m_{il}}{2}$, $\delta_2+\frac{1}{2}\sum_{i,j,l}\tilde{y}_{ijl}^2$)), where $\tilde{y}_{ijl}=y_{ijl}-\boldsymbol{x}_{ijl}^T\boldsymbol{\beta}_l-a_{0il}-a_{1il}t_{ijl}$.

- Update multinomial logit regression parameters:

  For $k=1,\ldots,K-1$, we draw $\boldsymbol{\gamma}_k$ from $N(m_k,V_k)$ as described in A.1.1, where $V_k=(Z'\Omega_kZ+V_{0k}^{-1})^{-1}$, $m_k=V_k(Z'\Omega_k(\kappa_k+\Omega_kC_k)+V_{0k}^{-1}m_{0k})$ and $O_k=\log\sum_{k'\neq k}\exp(Z\boldsymbol{\gamma}_k)$.

- Update latent class indicators:

  For $i=1,\ldots,n$, we sample C_i from Multinomial$(1,\boldsymbol{\pi}_i)$, where $\boldsymbol{\pi}_i=(\pi_{i1},\ldots,\pi_{iK})$. For $k=1,\ldots,K$, π_{ik} is proportional to

  $\exp(\boldsymbol{z}_i\boldsymbol{\gamma}_k)\exp\{\frac{(\boldsymbol{a}_i-\boldsymbol{\alpha}_k)^T\boldsymbol{\Sigma}_k^{-1}(\boldsymbol{a}_i-\boldsymbol{\alpha}_k)}{2}\}|\boldsymbol{\Sigma}_k|^{-1/2}$. π_{ik} is then scaled to be with sum 1 for all k's.

In addition, to address the common label switching issue when fitting Bayesian mxiture models, we implemented the Stephens' method[169] using the `label.switching` package in R[138]. On the other hand, once the truth is known, for both simulation study and ADNI data, the order constraint can be assigned to one parameter in random effects to avoid the potential label switching.

A.2 Simulation results

Table A.1. Simulation study results of $P = 1$ and $K = 3$. The model was fit to all 200 samples.

Parameter (true values)	**Bias**	**MSE**	**C$_{95}$**
		$k = 1$	
β_1 0.5	-0.0023	0.0013	0.91
β_2 1	-0.0042	0.0001	0.89
α_{011} 0.2	-0.0057	0.0001	0.90
α_{012} 2	-0.0070	0.0042	0.94
α_{013} 10	-0.0380	0.0965	0.94
α_{111} 3	0.0036	0.0011	0.95
α_{112} 1	0.0070	0.0028	0.92
α_{113} -0.5	0.0196	0.0142	0.89
		Covariate 1	
$\boldsymbol{\gamma}_{11}$ 1	-0.0394	0.0933	0.91
$\boldsymbol{\gamma}_{12}$ 2	0.0938	0.1129	0.92
$\boldsymbol{\gamma}_{21}$ 0.5	0.0410	0.0723	0.93
$\boldsymbol{\gamma}_{22}$ -0.5	0.0695	0.0270	0.95
p_{acc}	0.987		

* $K = 3$ is the reference class with $\boldsymbol{\gamma}_3 = 0$.

A.3 Supplementary Results

Table A.2. Simulation study results of $P = 3$ and $K = 2$. The model was fit to all 200 samples.

Parameter (true values)	**Bias**	**MSE**	**C_{95}**	**Bias**	**MSE**	**C_{95}**
	$Q_l = 1$			$Q_l = 2$		
β_{1Q_l} (1,0.5)	-0.0070	0.0030	0.93	0.0007	0.0001	0.95
β_{2Q_l} (3,-0.1)	-0.0015	0.0029	0.93	0.0013	0.0002	0.94
β_{3Q_l} (4,-0.5)	-0.146	0.0081	0.94	0.0043	0.0003	0.91
	$k = 1$			$k = 2$		
α_{0k1} (0.2,2)	-0.0346	0.0967	0.95	0.0207	0.0409	0.91
α_{0k2} (0.5,1)	0.0166	0.0105	0.95	-0.0173	0.0200	0.97
α_{0k3} (3,0.5)	-0.0132	0.0261	0.96	0.0420	0.1048	0.97
α_{1k1} (3,1)	-0.0119	0.0149	0.95	0.0104	0.0714	0.93
α_{1k2} (1,2)	0.0110	0.0223	0.96	-0.0141	0.0502	0.96
α_{1k3} (0.1,4)	0.0354	0.0632	0.94	-0.0568	0.1502	0.95
	Covariate 1			**Covariate 2**		
$\boldsymbol{\gamma}_1$ (1,0.5)	0.0096	0.0270	0.99	0.015	0.0314	0.98
p_{acc}	0.981					

[*] $K = 2$ is the reference class with $\boldsymbol{\gamma}_2 = 0$.

Table A.3. Simulation study results of $P = 3$ and $K = 3$. The model was fit to all 200 samples.

Parameter (true values)	**Bias**	**MSE**	**C_{95}**	**Bias**	**MSE**	**C_{95}**
	$l = 1$			$l = 2$		
β_{l1} (1,3)	-0.008	0.0031	0.92	0.0011	0.0011	0.98
β_{l2} (0.5,-0.1)	-0.0004	0.0002	0.86	-0.0020	0.0001	0.99
α_{01l} (0.2,0.5)	-0.0111	0.0157	0.98	-0.0088	0.0355	0.94
α_{02l} (2,1)	0.0134	0.0096	0.98	0.0258	0.0069	0.98
α_{03l} (10,2)	-0.1358	0.4774	0.92	-0.0734	0.0896	0.99
α_{11l} (3,1)	0.0328	0.0177	0.94	-0.0050	0.0495	1
α_{12l} (1,2)	0.0007	0.0033	0.98	-0.0443	0.0229	0.94
α_{13l} (-0.5,-0.5)	0.0718	0.1183	0.92	0.0520	0.0968	0.99
	$l = 3$					
β_{l1} 4	-0.0194	0.01114	0.90			
β_{l2} -0.5	-0.0024	0.0006	0.85			
α_{01l} 3	-0.0424	0.0364	0.98			
α_{02l} 0.5	0.0138	0.0308	0.94			
α_{03l} -2	0.2017	0.4193	0.94			
α_{11l} 0.1	-0.0826	0.0908	0.88			
α_{12l} 4	0.0063	0.0108	0.96			
α_{13l} 2	0.0667	0.0827	0.94			
	Covariate 1			**Covariate 2**		
$\boldsymbol{\gamma}_1$ (1,0.5)	-0.0720	0.1555	0.90	0.0571	0.1289	0.90
$\boldsymbol{\gamma}_2$ (2,-0.5)	-0.3398	0.3115	0.91	0.0122	0.0827	0.96
p_{acc}	0.95					

[*] $K = 3$ is the reference class with $\boldsymbol{\gamma}_3 = 0$.

Table A.4. Results of model selection regarding different numbers of latent classes *K* assumed.

Latent Class K	WAIC
1	-83771.54
2	-84613.31
3	-85294.46
4	-84929.75

Appendix B

Additional Results for Chapter 4,5

B.1 Theoretical results and proofs

- If $K_U(s,t) = \begin{pmatrix} K_{U_0}(s,t) & K_{U_{01}}(s,t) \\ K_{U_{01}}(t,s) & K_{U_1}(s,t) \end{pmatrix} = \sum_{l=1}^{\infty} \lambda_l^U \phi_l^{(1)}(s)\phi_l^{(1)}(t)'$, where $\phi_l^{(1)}(t) = (\phi_l^{U_0}(t), \phi_l^{U_1}(t))'$, then $K_{U_0}(s,t) = \sum_{l=1}^{\infty} \lambda_l^U \phi_l^{U_0}(t)\phi_l^{U_0}(s)$, $K_{U_1}(s,t) = \sum_{l=1}^{\infty} \lambda_l^U \phi_l^{U_1}(t)\phi_l^{U_1}(s)$ and $K_{U_{01}}(s,t) = \sum_{l=1}^{\infty} \lambda_l^U \phi_l^{U_0}(t)\phi_l^{U_1}(s)$.

  Proof:

$$\begin{aligned}
K_U(s,t) &= \sum_{l=1}^{\infty} \lambda_l^U \phi_l^{(1)}(s)\phi_l^{(1)}(t)' \\
&= \sum_{l=1}^{\infty} \lambda_l^U (\phi_l^{U_0}(s), \phi_l^{U_1}(s))(\phi_l^{U_0}(t), \phi_l^{U_1}(t))' \\
&= \sum_{l=1}^{\infty} \lambda_l^U \begin{pmatrix} \phi_l^{U_0}(s)\phi_l^{U_0}(t) & \phi_l^{U_0}(s)\phi_l^{U_1}(t) \\ \phi_l^{U_0}(t)\phi_l^{U_1}(s) & \phi_l^{U_1}(s)\phi_l^{U_0}(t) \end{pmatrix} \\
&= \sum_{l=1}^{\infty} \begin{pmatrix} \lambda_l^U \phi_l^{U_0}(s)\phi_l^{U_0}(t) & \lambda_l^U \phi_l^{U_0}(s)\phi_l^{U_1}(t) \\ \lambda_l^U \phi_l^{U_0}(t)\phi_l^{U_1}(s) & \lambda_l^U \phi_l^{U_1}(s)\phi_l^{U_0}(t) \end{pmatrix} \\
&= \begin{pmatrix} K_{U_0}(s,t) & K_{U_{01}}(s,t) \\ K_{U_{01}}(t,s) & K_{U_1}(s,t) \end{pmatrix}
\end{aligned}$$

- **Proof of Equation 4.7**. For given eigenfunctions, eigenvalues, the BLUP for principal component scores $\hat{\boldsymbol{\beta}} = (\hat{\xi}_{11}, \dots, \hat{\xi}_{1N_U}, \dots, \hat{\xi}_{N1}, \dots, \hat{\xi}_{NN_U}, \hat{\zeta}_{111}, \dots, \hat{\zeta}_{1J_1 1}, \dots, \hat{\zeta}_{N1N_V}, \dots, \hat{\zeta}_{Nn_N N_V})$ has a usual form as,

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\Lambda} \mathbf{Z}'(\mathbf{Z}\boldsymbol{\Lambda}\mathbf{Z}')^{-1}\boldsymbol{X}$$

where $\boldsymbol{\Lambda} = blockdiag\{\boldsymbol{I}_N \otimes diag(\lambda_1^U, \dots, \lambda_{N_U}^U), \boldsymbol{I}_n \otimes diag(\lambda_1^V, \dots, \lambda_{N_V}^V)\}$, $n = \sum_{i=1}^N n_i$. When $D > N_U + n_i * N_V$, $\mathbf{Z}\boldsymbol{\Lambda}\mathbf{Z}'$ is not invertible and only the generalized inverse of $\mathbf{Z}\boldsymbol{\Lambda}\mathbf{Z}'$ can be used[200]. For our implementation, N_U and N_V are in general small numbers and the length of grid points of time is always significantly larger. In this case, $\hat{\boldsymbol{\beta}} = \boldsymbol{\Lambda}\mathbf{Z}'(\mathbf{Z}\boldsymbol{\Lambda}\mathbf{Z}')^{-1}\boldsymbol{X} = \boldsymbol{\Lambda}^{1/2}(\boldsymbol{\Lambda}^{1/2}\mathbf{Z}'\mathbf{Z}\boldsymbol{\Lambda}^{1/2})^{-1}\boldsymbol{\Lambda}^{1/2}\mathbf{Z}'\boldsymbol{X} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\boldsymbol{X}$ [200]. Thus we proved the expression in Equation 7 is the BLUP for $\boldsymbol{\beta}$.

- **Proof of Equation 5.15** With iterated expectation, we have

$$Var\{Y_{ijk}(t)\} = Var\{E(Y_{ijk}(t)|T_{ij})\} + E\{Var(Y_{ijk}(t)|T_{ij})\}$$

With $E\{Var(Y_{ijk}(t)|T_{ij})\} = 0$, the equation is written as,

$$\int_0^1 Var\{Y_{ijk}(t)\}dt = \int_0^1 E\{\sum_l \xi_{il}(\phi_l^{U_0}(t) + T_{ij}\phi_l^{U_1}(t))^2 + \sum_m \zeta_{ijm}(\phi_m^V(t))^2 + \sum_r \eta_{ijkr}\phi_r^W(t)^2\}dt$$

If $E(T_{ij}) = 0$ and $Var(T_{ij}) = 1$, then,

$$
\begin{aligned}
\int_0^1 Var\{Y_{ijk}(t)\}dt &= \sum_l \xi_{il} \int_0^1 (\phi_l^{U_0}(t))^2 + (\phi_l^{U_1}(t))^2 dt + \sum_m \zeta_{ijm} \int_0^1 (\phi_m^V(t))^2 dt \\
&+ \sum_r \eta_{ijkr} \int_0^1 (\phi_r^W(t))^2 dt \\
&= \sum_l \xi_{il} + \sum_m \zeta_{ijm} + \sum_r \eta_{ijkr}
\end{aligned}
$$

B.2 Simulation Results

We simulate data based on ideas implemented in [60]. For each simulation setting, we generate 100 replicates with $n = 100$ subjects in each dataset. We assume a balanced design with $n_i = 3$ visits for each subject and the time variable T_{ij} is generated by standardizing the visits (j=1,2,3) to have unit variance. $M = 300$ is the total number of observations in each simulation replicate. The functional curves $X_{ij}(t)$ with length of 600 are generated as follows,

$$
\begin{cases}
X_{ij}(t) & = \sum_{l=1}^{N_U} \xi_{il}\phi_l^{(U_0)}(t) + \sum_{l=1}^{N_U} T_{ij}\xi_{il}\phi_l^{(U_1)}(t) + \sum_{m=1}^{N_V} \zeta_{ijm}\phi_m^{(2)}(t), t \in \mathscr{D} \\
\xi_{il} & \overset{\text{i.i.d.}}{\sim} \mathscr{N}(0.\lambda_l^U), \zeta_{ijm} \overset{\text{i.i.d.}}{\sim} \mathscr{N}(0.\lambda_m^V)
\end{cases}
$$

where the number of eigenfunctions $N_U = N_V = 4$ and the scores ξ_{il}'s and ζ_{ijm}'s are mutually independent. The eigenfunctions bases are set as,

$$
\begin{array}{lll}
\phi_1^{U_0}(t) = \sin(2\pi t), & \phi_1^{U_1}(t) = 1/2, & \phi_1^V(t) = 1 \\
\phi_2^{U_0}(t) = \cos(2\pi t), & \phi_2^{U_1}(t) = \sin(6\pi t), & \phi_2^V(t) = \sqrt{3}(2t-1) \\
\phi_3^{U_0}(t) = \sin(4\pi t), & \phi_3^{U_1}(t) = \cos(6\pi t), & \phi_3^V(t) = \sqrt{5}(6d^2-6d+1) \\
\phi_4^{U_0}(t) = \cos(4\pi t), & \phi_4^{U_1}(t) = \sin(8\pi t), & \phi_4^V(t) = \sqrt{7}(20d^3-30d^2+12d-1)
\end{array}
$$

where $\phi_l^{U_0}$ and $\phi_l^{U_1}$ are orthogonal but they are correlated with ϕ_m^V if $m \neq 1$.

The true eigenvalues have two scenarios,

1. level 1: $\lambda_1^U = 4, \lambda_l^U = 0.5^{l-1}, l = 2,3,4$ and level 2: $\lambda_m^V = 0.5^{m-1}, m = 1,2,3,4$

2. level 1: $\lambda_l^U = 0.5^{l-1}, l = 1,2,3,4$ and level 2: $\lambda_1^V = 4, \lambda_m^V = 0.5^{m-1}, m = 2,3,4$

For each of the 100 simulated datasets, we implemented the longitudinal FPCA, as described in section 3.2 to estimate the eigenfunctions, eigenvalues, scores and predicted functional trajectories. As a first assessment of model estimation accuracy, we computed the normalized errors between the estimated and true values of subject-level $(\hat{\xi}_{il} - \xi_{il})/\sqrt{\lambda_l^U}$ and visit-level $(\hat{\zeta}_{ijm} - \zeta_{ijm})/\sqrt{\lambda_m^V}$ scores. The results are displayed in B.1, which show that the score parameters are unbiasedly estimated. The simulation results demonstrate the agreement with simulation results in Greven et al. (2010)[60], which provided a more complete list of simulation examples.

As a second assessment of accuracy, we calculated three ways of *residual* $(R_{ij}(t))$ MSE described in section 3.2, which is defined as the total mean squared count difference per observation between the predicted and observed activity curves, i.e.$\frac{1}{M}\sum_{i,j}(\sum_t |R_{ij}(t)|)^2$. The results are displayed in B.2, and the findings are discussed in section 3.2.

B.3 Sensitivity Analysis

To assess, if averaging daily records impacted on our findings, we conducted three sensitivity analyses. In particular, we took (i) a random day, (ii) the first three-day average and the (iii) up to the first seven-day average for each participant at each visit, and then applied the same longitudinal FPCA on these new inputs. The results are provided B.1, B.2 and B.3. Compared with results in the original manuscript (Table 2 and B.4), averaging over daily inputs does not meaningfully alter results of either functional PCA or regression procedures. Specifically, although the variation explained by level-1 PCs is lower when only considering a random day, the overall variance explained by both level 1 and level 2 PCs are similar in general. Also, in the regression models, while point estimates vary, the findings are consistent across approaches, namely that higher person-level PC scores 1 and 2 are associated with lower (log)insulin and

BMI. In fact, the averaged inputs could reduce the influence of random movements or activity on a single day, thus reducing noise in the data.

Table B.1. Percentages of average variance explained by different levels of principal components and regression results of log(insulin) and BMI on the first two level 1 and level 2 principal component scores using a-random-day data.

# Component	$\phi_l^{U_0}$	$\phi_l^{U_1}$	ϕ_m^2	cumulative variation explained
1	0.0764	0.0120	0.2543	0.3427
2	0.0200	0.0034	0.1606	0.5267
3	0.0125	0.0017	0.0836	0.6245
	0.1089	0.0171	0.4985	0.6245
Outcome	Predictor	Coefficient estimate	SE	Confidence interval
Log(insulin)	PC11	-0.09	0.03	(-0.15, -0.04)
	PC12	-0.07	0.03	(-0.12, -0.01)
	PC21	-0.04	0.02	(-0.07, -0.01)
	PC22	0.00	0.01	(-0.03, 0.03)
BMI	PC11	-0.45	0.22	(-0.87, -0.03)
	PC12	-0.39	0.22	(-0.82, 0.04)
	PC21	-0.10	0.08	(-0.25, 0.05)
	PC22	0.10	0.07	(-0.03, 0.24)

B.4 Additional Results

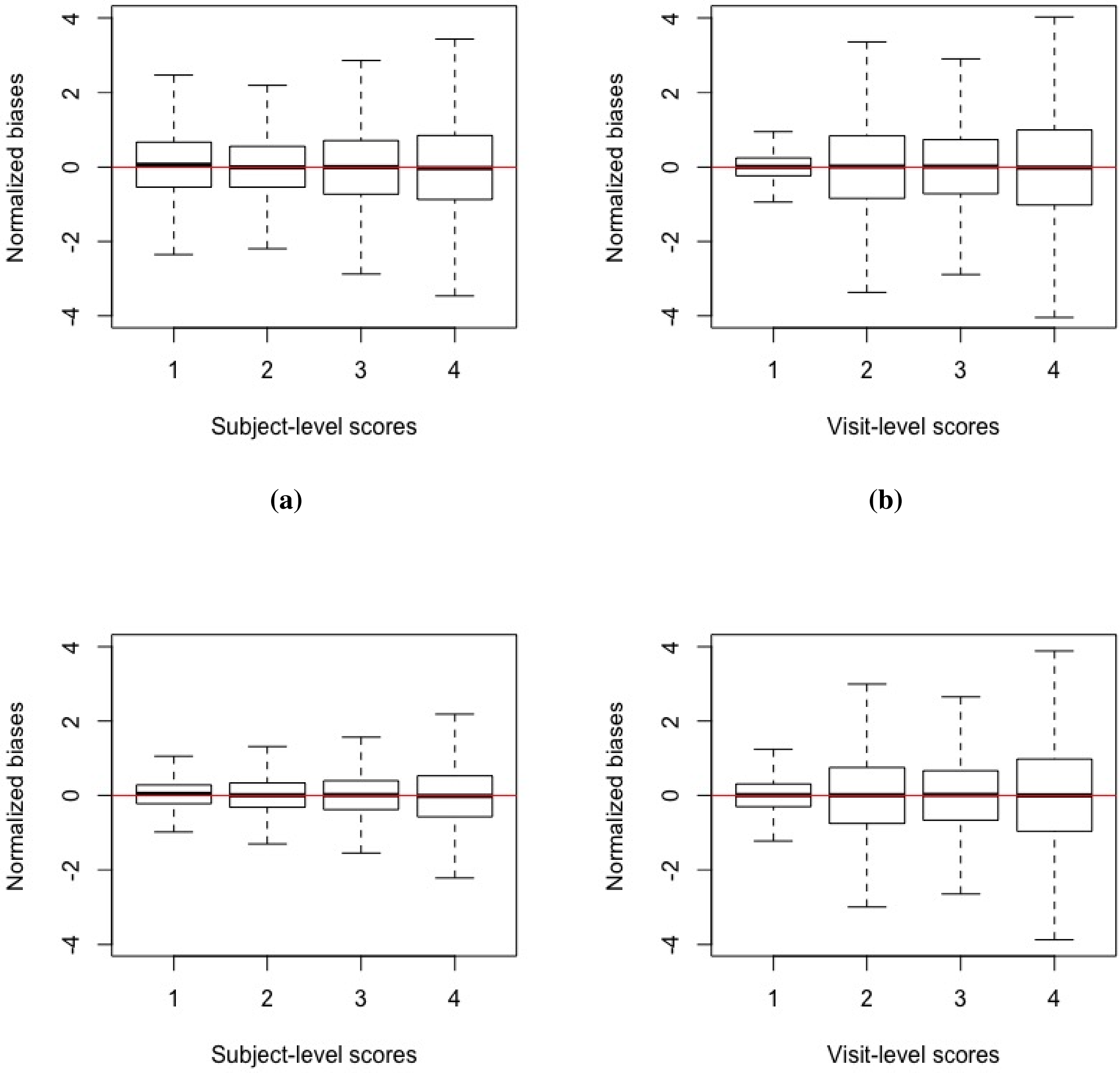
B.3 provides two examples from our dataset of step-wise reconstruction of the activity curves, after the eigen-decomposition. The first individual example has a large first level 1 principal component score but a small first level 2 principal component score, and vice versa for the second individual example. It further illustrates that principal component scores can inform explained variation at subject- and visit-level. Specifically, minimal between-visit difference is witnessed in the first example, i.e., adding in the visit-level component to the subject-level curves does not improve the fit materially, indicating the majority of variation is explained at subject-level. While in the second example which has a larger level 2 principal component score, the figure presents more significant variation between the two visits, and thus the visit-level component is needed to recapitulate the trends of the original data.

Table B.2. Percentages of average variance explained by different levels of principal components and regression results of log(insulin) and BMI on the first two level 1 and level 2 principal component scores using three-day average data.

# Component	$\phi_l^{U_0}$	$\phi_l^{U_1}$	ϕ_m^2	cumulative variation explained
1	0.1507	0.0104	0.2160	0.3771
2	0.0391	0.0142	0.1322	0.5626
3	0.0206	0.0052	0.0898	0.6782
4	0.0152	0.0013	0.0585	0.7532
	0.2256	0.0311	0.4965	0.7532
Outcome	Predictor	Coefficient estimate	SE	Confidence interval
Log(insulin)	PC11	-0.11	0.03	(-0.16, -0.05)
	PC12	-0.10	0.02	(-0.14, -0.06)
	PC21	-0.03	0.02	(-0.06, 0.00)
	PC22	0.01	0.02	(-0.02, 0.04)
BMI	PC11	-0.86	0.21	(-1.28, -0.44)
	PC12	-0.12	0.18	(-0.47, 0.22)
	PC21	-0.13	0.08	(-0.28, 0.02)
	PC22	0.13	0.02	(-0.12, 0.03)

Table B.3. Percentages of average variance explained by different levels of principal components and regression results of log(insulin) and BMI on the first two level 1 and level 2 principal component scores using up to seven-day average data.

# Component	$\phi_l^{U_0}$	$\phi_l^{U_1}$	ϕ_m^2	cumulative variation explained
1	0.2302	0.0220	0.1551	0.4073
2	0.0852	0.0167	0.1030	0.6122
3	0.0363	0.0044	0.0843	0.7372
4	0.0192	0.0013	0.0484	0.8061
	0.3709	0.0444	0.3908	0.8061
Outcome	Predictor	Coefficient estimate	SE	Confidence interval
Log(insulin)	PC11	-0.14	0.04	(-0.22, -0.06)
	PC12	-0.09	0.03	(-0.15, -0.03)
	PC21	-0.13	0.04	(-0.20, -0.06)
	PC22	0.02	0.03	(-0.03, 0.08)
BMI	PC11	-0.78	0.29	(-1.34, -0.22)
	PC12	-0.39	0.24	(-0.86, 0.07)
	PC21	-0.19	0.19	(-0.56, 0.18)
	PC22	0.00	0.14	(-0.26, 0.27)



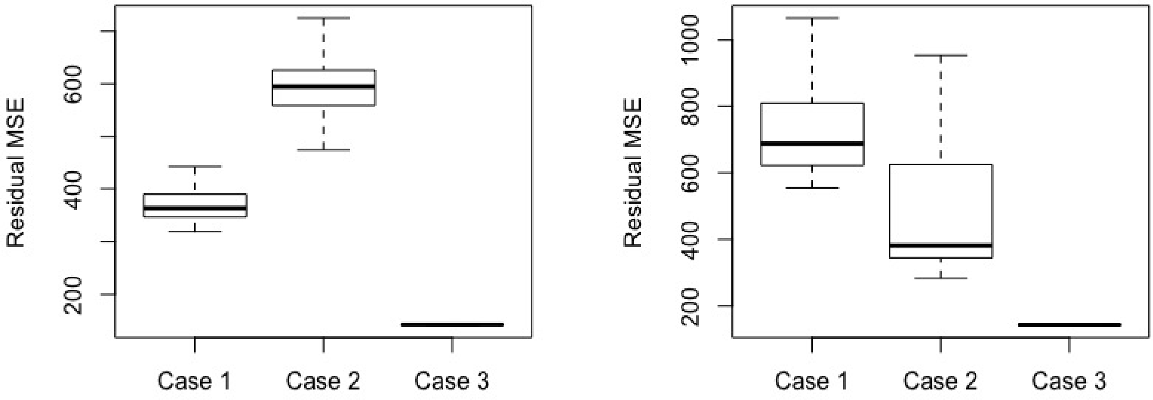
(a)

(b)

(c)

(d)

Figure B.1. Boxplots of the normalized biases of estimated principal component scores for subject-level process $(\hat{\xi}_{il} - \xi_{il})/\sqrt{\lambda_l^U}$ (left) and visit-level $(\hat{\zeta}_{ijm} - \zeta_{ijm})/\sqrt{\lambda_m^V}$ (right) based on simulation scenario 1 (top) and scenario 2 (bottom) with 100 replicates. Red line represents the zero.

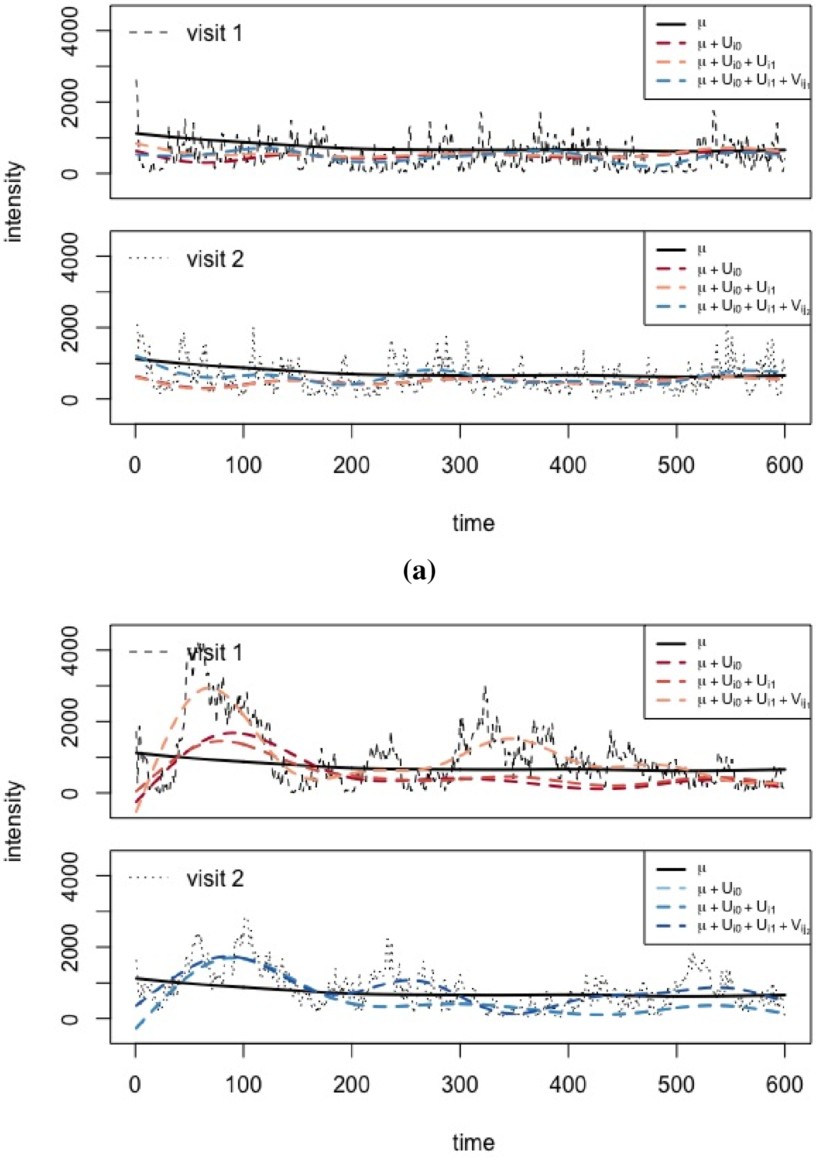


(a) **(b)**

Figure B.2. Boxplots of residual MSE from stepwise decomposition based on simulation scenario 1 (top) and scenario 2 (bottom) with 100 replicates. From left to right, the mean squared errors are acquired from three ways of computing residuals: $X_{ij}(t) - U_i(t)$ (Case 1), $X_{ij}(t) - V_{ij}(t)$ (Case 2), $X_{ij}(t) - U_i(t) - V_{ij}(t)$ (Case 3)

Table B.4. Percentages of average variance explained by different levels of principal components. The cumulative variation is the sum of row entries for the current row. The last row presents the cumulative variance for each column.

# Component	$\phi_l^{U_0}$	$\phi_l^{U_1}$	$\phi_m^{(2)}$	**cumulative variation explained**
1	0.2301	0.0228	0.2530	0.5059
2	0.0863	0.0163	0.1026	0.7111
3	0.0365	0.0048	0.0412	0.7935
4	0.0196	0.0015	0.0210	0.8356
5	0.0131	0.0022	0.0153	0.8662
	0.3855	0.0476	0.4331	0.8662



(a)

(b)

Figure B.3. Stepwise decomposition of two examples of PA records with raw count inputs (black) and estimated curves at each visit (red, blue): (a) is an example with a large first level 1 principal component score but a small first level 2 principal component score; (b) is an example with a small first level 1 principal component score but a large first level 2 principal component score.

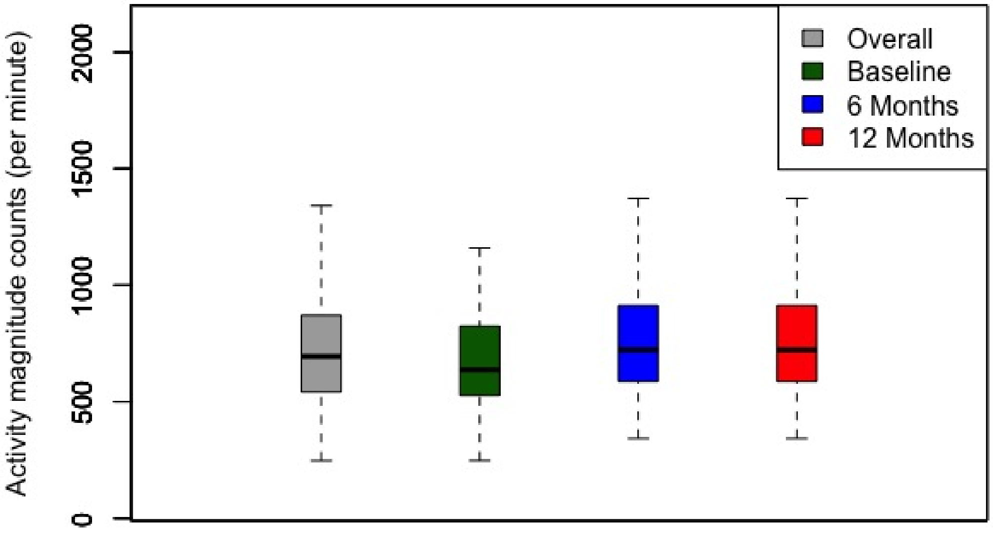


Figure B.4. Boxplot of individual daily-average activity magnitude (sum of activity magnitudes over 600 minutes divided by 600) overall and at baseline, 6 months, 12 months. It illustrates an increase in PA magnitudes after baseline visits.

Table B.5. Linear mixed effect regression results of health outcomes on scaled total activity counts and MVPA respectively. It presents that both total activity counts and MVPA both exhibit a negative association with health outcomes, which supports the PCR results.

Outcome	Total Activity Counts	MVPA
Log(insulin)	-0.13 (-0.17, -0.09)	-0.12 (-0.16,-0.08)
Log(CRP)	-0.09 (-0.17, -0.01)	-0.09 (-0.17,-0.02)
BMI	-0.69 (-0.9, -0.48)	-0.63 (-0.84,-0.43)

*Adjusted for baseline age, ethnicity, smoking history and visit indicator.

Appendix C

Additional Results for Chapter 6

C.1 Deriving Trend Parameters (Slopes) in MTSR Models

To better interpret the linear time trend parameter $\boldsymbol{\beta}_{i1}$ in MTSR models for temperature data, we provided an analytical expansion of the original model. For the purpose of clear representation, we use $\boldsymbol{y}_{ij}^{(k)}$, $\boldsymbol{\beta}_{ij}^{(k)}$, $\boldsymbol{b}_{ij}^{(k)}$ and $\boldsymbol{\varepsilon}_{ij}^{(k)}$ to denote the k-lagged covariates, parameters and residuals, with $k = 0, 1, 2, 3$. $\boldsymbol{y}_{ij}$ can be computed as,

$$\begin{aligned}
\boldsymbol{y}_{ij} &= \boldsymbol{X}'_{ij}\boldsymbol{\beta}_i^{(0)} + \boldsymbol{y}_{ij}^{(1)}\boldsymbol{b}_i^{(0)} + \boldsymbol{\varepsilon}_{ij}^{(0)} \\
&= \boldsymbol{X}'_{ij}\boldsymbol{\beta}_i^{(0)} + (\boldsymbol{X}'_{ij}\boldsymbol{\beta}_i^{(1)} + \boldsymbol{y}_{ij}^{(2)}\boldsymbol{b}_i^{(1)} + \boldsymbol{\varepsilon}_{ij}^{(1)})\boldsymbol{b}_i^{(0)} + \boldsymbol{\varepsilon}_{ij}^{(0)} \\
&= \boldsymbol{X}'_{ij}(\boldsymbol{\beta}_i^{(0)} + \boldsymbol{\beta}_i^{(1)}\boldsymbol{b}_i^{(0)}) + \boldsymbol{y}_{ij}^{(2)}\boldsymbol{b}_i^{(1)}\boldsymbol{b}_i^{(0)} + \boldsymbol{\varepsilon}_{ij}^{(0)} + \boldsymbol{\varepsilon}_{ij}^{(1)}\boldsymbol{b}_i^{(0)} \\
&= \dots \\
&= \boldsymbol{X}'_{ij}\sum_{k=0}^{3}(\prod_{k'<k}\boldsymbol{b}_i^{(k'-1)})\boldsymbol{\beta}_i^{(k)} + \boldsymbol{y}_{i(j-1)}\prod_{k=0}^{3}\boldsymbol{b}_i^{(k)} + \sum_{k=0}^{3}(\prod_{k'<k}\boldsymbol{b}_i^{(k'-1)})\boldsymbol{\varepsilon}_{ij}^{(k)}
\end{aligned}$$

where $k' \geq -1$ and $b_i^{(-1)} = \mathtt{1}$. Similarly, we can write out $\boldsymbol{y}_{i(j-1)}$ with the form,

$$\boldsymbol{y}_{i(j-1)} = \boldsymbol{X}'_{ij-1}\sum_{k=0}^{3}(\prod_{k'<k}\boldsymbol{b}_i^{(k'-1)})\boldsymbol{\beta}_i^{(k)} + \boldsymbol{y}_{i(j-2)}\prod_{k=0}^{3}\boldsymbol{b}_i^{(k)} + \sum_{k=0}^{3}(\prod_{k'<k}\boldsymbol{b}_i^{(k'-1)})\boldsymbol{\varepsilon}_{ij-1}^{(k)}$$

Take the difference of the above two equations and denote $\Delta(\boldsymbol{y}_{ij}) = \boldsymbol{y}_{ij} - \boldsymbol{y}_{ij-1}$, $\Delta_t = t_j - t_{j-1}$ and $\Delta(\boldsymbol{\varepsilon}_{ij}^{(k)}) = \boldsymbol{\varepsilon}_{ij}^{(k)} - \boldsymbol{\varepsilon}_{ij-1}^{(k)}$, a difference model is provided as,

$$\Delta(\boldsymbol{y}_{ij}) = \Delta_t \sum_{k=0}^{3} (\prod_{k'<k} \boldsymbol{b}_i^{(k'-1)}) \boldsymbol{\beta}_{i1}^{(k)} + \Delta(\boldsymbol{y}_{ij-1}) \prod_{k=0}^{3} \boldsymbol{b}_i^{(k)} + \sum_{k=0}^{3} (\prod_{k'<k} \boldsymbol{b}_i^{(k'-1)}) \Delta(\boldsymbol{\varepsilon}_{ij}^{(k)})$$

In our model setting, $\Delta_t = 0.1$ is a fixed number. Therefore, when we divide the two sides of the equation with Δ_t and let $\boldsymbol{v}_{ij} = \frac{\Delta(\boldsymbol{y}_{ij})}{\Delta_t}$, the equation becomes,

$$\boldsymbol{v}_{ij} = \sum_{k=0}^{3} \boldsymbol{a}_{i0}^{(k)} \boldsymbol{\beta}_{i1}^{(k)} + \boldsymbol{a}_{i1} \boldsymbol{v}_{ij-1} + \sum_{k=0}^{3} \boldsymbol{a}_{i0}^{(k)} \Delta(\boldsymbol{\varepsilon}_{ij}^{(k)}) / \Delta_t$$

where $\boldsymbol{a}_{i0}^{(k)} = \prod_{k'<k} \boldsymbol{b}_i^{(k'-1)}$ and $\boldsymbol{a}_{i1} = \prod_{k=0}^{3} \boldsymbol{b}_i^{(k)}$. $\boldsymbol{v}_{ij}$ is in fact the temperature changing rate per decade, and hence has the same interpretation as the slope parameter if we fit a simple linear trend model with time as the only predictor. Meanwhile, if $-1 < \boldsymbol{a}_{i1} < 1$, $\boldsymbol{v}_{ij}$, itself is a autoregressive model with the long-run mean $E(\boldsymbol{v}_{ij}) = \frac{\sum_{k=0}^{3} \boldsymbol{a}_{i0}^{(k)} \boldsymbol{\beta}_{i1}^{(k)}}{1 - \boldsymbol{a}_{i1}}$.

The corresponding standard deviance of $\boldsymbol{v}_{ij}$ is derived from,

$$Var(\boldsymbol{v}_{ij}) = \boldsymbol{a}_{i1}^2 Var(\boldsymbol{v}_{ij-1}) + \frac{1}{\Delta_t^2} Var(\sum_{k=0}^{3} \boldsymbol{a}_{i0}^{(k)} \Delta(\boldsymbol{\varepsilon}_{ij}^{(k)}))$$

With the time-independent assumption of $\boldsymbol{\varepsilon}_{ij}^{(k)}$, i.e. $Cov(\boldsymbol{\varepsilon}_{ij}^{(k)}, \boldsymbol{\varepsilon}_{ij-1}^{(k)}) = 0$, and let $\Sigma_i^{(k_1,k_2)} = Cov(\boldsymbol{\varepsilon}_{ij}^{(k_1)}, \boldsymbol{\varepsilon}_{ij}^{(k_2)}), k_1, k_2 = 0,1,2,3$, we have $Var(\sum_{k=0}^{3} \boldsymbol{a}_{i0}^{(k)} \Delta(\boldsymbol{\varepsilon}_{ij}^{(k)})) = 2\sum_{k_1,k_2} \boldsymbol{a}_{i0}^{(k_1)} \boldsymbol{a}_{i0}^{(k_2)} \Sigma_i^{(k_1,k_2)}$. Therefore, the variance of the $\boldsymbol{v}_{ij}$ has the form,

$$Var(\boldsymbol{v}_{ij}) = \frac{\Sigma_{v_i}}{1 - \boldsymbol{a}_{i1}^2}$$

where $\Sigma_{v_i} = \frac{2}{\Delta_t^2}\sum_{k_1,k_2} \boldsymbol{a}_{i0}^{(k_1)}\boldsymbol{a}_{i0}^{(k_2)}\Sigma_i^{(k_1,k_2)}$. Thus we have $\mathrm{sd}(\boldsymbol{v}_{ij}) = \sqrt{\frac{\Sigma_{v_i}}{1-\boldsymbol{a}_{i1}^2}}$.

We further estimated the standard errors of $\hat{\boldsymbol{v}}_{ij}$. It can be estimated from the standard error of $\frac{\sum_{k=0}^{3}\hat{\boldsymbol{a}}_{i0}^{(k)}\hat{\boldsymbol{\beta}}_{i1}^{(k)}}{1-\hat{\boldsymbol{a}}_{i1}}$ via the multivariate delta method, that is, $\sqrt{T-1}(h(\hat{\boldsymbol{B}}_i)-h(\boldsymbol{B}_i)) \xrightarrow{d} N(0, \nabla h(\boldsymbol{B}_i)'\Sigma_{\boldsymbol{B}_i}\nabla h(\boldsymbol{B}_i))$, where $h(\boldsymbol{B}_i) = \frac{\sum_{k=0}^{3}\boldsymbol{a}_{i0}^{(k)}\boldsymbol{\beta}_{i1}^{(k)}}{1-\boldsymbol{a}_{i1}}$ and $\nabla h(\boldsymbol{B}_i)$ is the vector notation for the gradient.

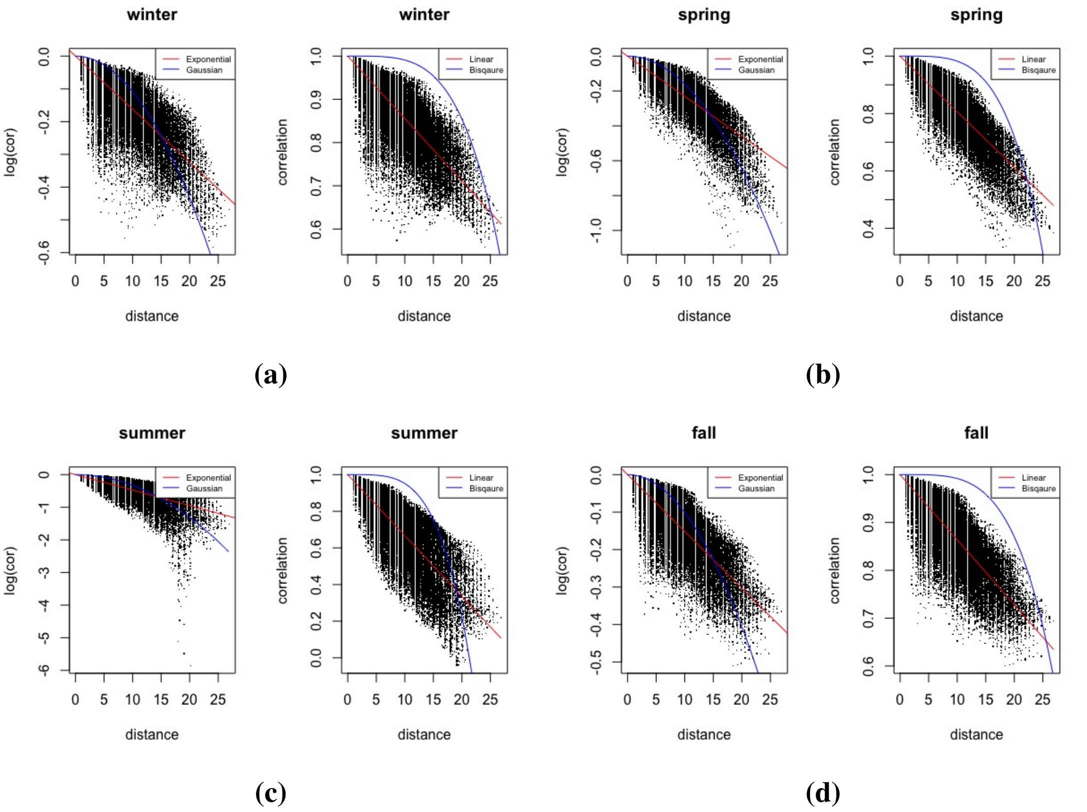
For the simplicity of representation, we remove all i and j's in this part. The denominator $1-\boldsymbol{a}_{i1}^2$ is a number close to 1, therefore we treated it as a constant in our computation. These gradients were computed as,

$$
\begin{aligned}
\frac{\nabla v_1}{\nabla \boldsymbol{B}} &= (1, b_1b_4b_3, b_1b_4, b_1, \beta_4b_2+\beta_3b_2b_4, \beta_1+\beta_4b_1+\beta_3b_1b_4, 0, \beta_3b_2b_1)' \\
\frac{\nabla v_2}{\nabla \boldsymbol{B}} &= (b_2, 1, b_2b_1b_4, b_2b_1, \beta_4b_3b_2, \beta_1b_3+\beta_4b_3b_1, \beta_2+\beta_1b_2+\beta_4b_2b_1, 0)' \\
\frac{\nabla v_3}{\nabla \boldsymbol{B}} &= (b_3b_2, b_3, 1, b_3b_2b_1, 0, \beta_1b_4b_3, \beta_2b_4+\beta_1b_4b_2, \beta_3+\beta_2b_3+\beta_1b_3b_2)' \\
\frac{\nabla v_4}{\nabla \boldsymbol{B}} &= (b_4b_3b_2, b_4b_3, b_4, 1, \beta_4+\beta_3b_4+\beta_2b_4b_3, 0, \beta_2b_1b_4, \beta_3b_1+\beta_2b_1b_3)'
\end{aligned}
$$

C.2 Supplementary Materials for CA

In this section, we provided supplementary figures and tables for analysis in California. Figure C.1 and Table C.1 illustrates the results of selecting an appropriate kernel for constructing the weight matrix in geographically weighted models, which shows that the linear kernel has the best performance. The regression lines demonstrate a linear association between pairwise residual correlation and distance, and the results are supported by the R-square results. FigureC.2 presents the results of selecting a threshold value for constructing the weight matrix. The mean squared errors (MSE) were computed based on a 10-fold cross-validation process, which performed the fitting procedure a total of ten times. Each fit was performed on a training set consisting of 90% of locations selected at random, with the remaining 10% used as a hold out set for validation. The MSE reflected the averaged fitting performance from the ten folds and we selected the threshold l which has the least MSE.

Figure C.3 and C.4 provide the assessment of CoPE method assumptions for temperature and precipitation, respectively. For temperature data (Figure C.3(a)), we performed Shapiro-Wilk test on regression residuals at all locations and seasons and computed corresponding p-values. For precipitation (Figure C.3(b)), the Shapiro-Wilk test was performed on deviance from the gamma regression models. The normality assumption were valid for all scenarios since the empirical cumulative distribution function of these p values approximately follow the CDF of a standard uniform distribution. As for assessing the assumption of independence, the Ljung–Box tests were performed on residuals from both temperature and precipitation results (Figure C.4). In general, the independence assumptions were valid from the CDF results.



(a)

(b)

(c)

(d)

Figure C.1. Selecting appropriate kernels in modeling CA climate data in four seasons. The y-axis provides the pairwise residual correlations for all locations with the form of either log(ρ) or *rho*.The x-axis provide the pariwise distance in grid.

Table C.1. R-squares of regression models for selecting appropriate kernels in modeling CA climate data in four seasons.

	Winter	Spring	Summer	Fall
Exponential kernel	0.89	0.91	0.81	0.90
Gaussian kernel	0.80	0.89	0.77	0.81
Linear kernel	0.90	0.94	0.90	0.91
Bisquare lernel	0.79	0.88	0.80	0.81

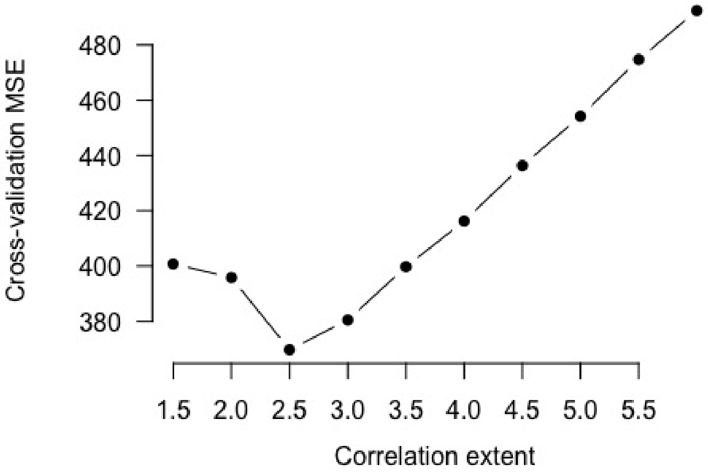
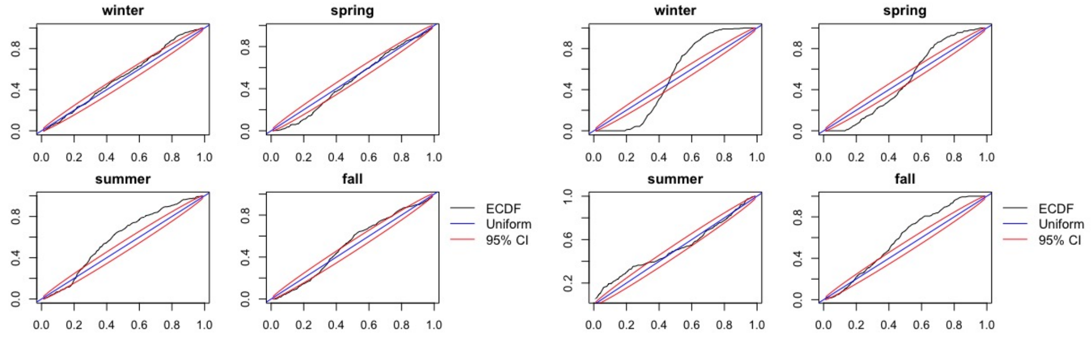
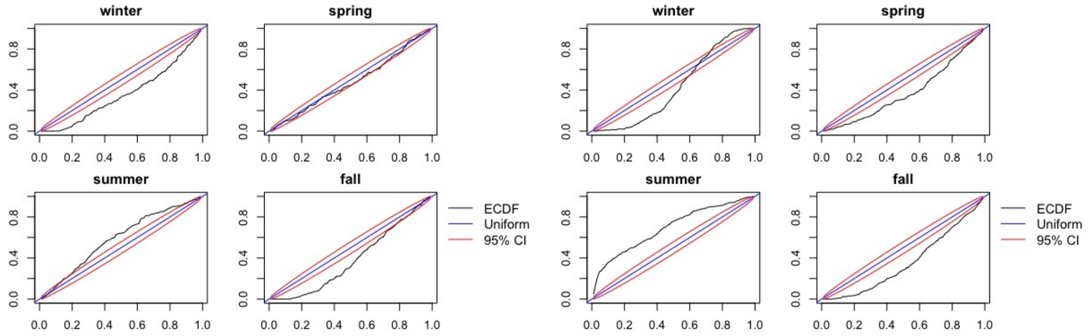


Figure C.2. 10-fold cross-validation results for selecting a threshold value in constructing the weight matrix of geographically weighted models.



(a) **(b)**

Figure C.3. Assessment of CoPE assumptions for temperature parameters. (a) Empirical CDFs of the p values from the Shapiro-Wilk test of normality of the residuals for all locations. (b) Empirical CDFs of the p values from the Ljung–Box test of independence of the residuals for all locations. The black line is the CDF of a standard uniform distribution.



(a) (b)

Figure C.4. Assessment of CoPE assumptions for precipitation parameters. (a) Empirical CDFs of the p values from the Shapiro-Wilk test of normality of the deviance for all locations. (b) Empirical CDFs of the p values from the Ljung–Box test of independence of the residuals for all locations. The black line is the CDF of a standard uniform distribution.

Bibliography

[1] IPCC Adopted. Climate change 2014 synthesis report. *IPCC: Geneva, Szwitzerland*, 2014.

[2] Hirotogu Akaike. Information theory and an extension of the maximum likelihood principle. In *Selected papers of hirotugu akaike*, pages 199–213. Springer, 1998.

[3] Marilyn S Albert, Steven T DeKosky, Dennis Dickson, Bruno Dubois, Howard H Feldman, Nick C Fox, Anthony Gamst, David M Holtzman, William J Jagust, Ronald C Petersen, et al. The diagnosis of mild cognitive impairment due to alzheimer's disease: recommendations from the national institute on aging-alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease. *Alzheimer's & dementia*, 7(3):270–279, 2011.

[4] Eugenia C Argiropoulou, Maria Michalopoulou, Nikolaos Aggeloussis, and Andreas Avgerinos. Validity and reliability of physical activity measures in greek high school age children. *Journal of sports science & medicine*, 3(3):147, 2004.

[5] Amogne Asfaw, Belay Simane, Ali Hassen, and Amare Bantider. Variability and time series trend analysis of rainfall and temperature in northcentral ethiopia: A case study in woleka sub-basin. *Weather and climate extremes*, 19:29–41, 2018.

[6] Nicole H Augustin, Calum Mattocks, Julian J Faraway, Sonja Greven, and Andy R Ness. Modelling a response as a function of high-frequency count data: The association between physical activity and fat mass. *Statistical methods in medical research*, 26(5):2210–2226, 2017.

[7] Mordecai Avriel. *Nonlinear programming: analysis and methods*. Courier Corporation, 2003.

[8] Jiawei Bai, Jeff Goldsmith, Brian Caffo, Thomas A Glass, and Ciprian M Crainiceanu. Movelets: A dictionary of movement. *Electronic journal of statistics*, 6:559, 2012.

[9] Rachel Ballard-Barbash, Stephanie M George, Catherine M Alfano, and Kathryn Schmitz. Physical activity across the cancer continuum. *Oncology*, 27(6):589–589, 2013.

[10] Sudipto Banerjee, Alan E Gelfand, Andrew O Finley, and Huiyan Sang. Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(4):825–848, 2008.

[11] David R Bassett. Device-based monitoring in physical activity and public health research. *Physiological measurement*, 33(11):1769, 2012.

[12] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*, 2014.

[13] Mohamed Amine Benadjaoud, Mehdi Menai, Vincent T van Hees, Vadim Zipunnikov, Jean-Philippe Regnaux, Mika Kivimäki, Archana Singh-Manoux, and Séverine Sabia. The association between accelerometer-assessed physical activity and respiratory function in older adults differs between smokers and non-smokers. *Scientific reports*, 9(1):1–9, 2019.

[14] Jūratė šaltytė Benth, Fred Espen Benth, and Paulius Jalinskas. A spatial-temporal model for temperature with seasonal variance. *Journal of Applied Statistics*, 34(7):823–841, 2007.

[15] L Mark Berliner, Ralph F Milliff, and Christopher K Wikle. Bayesian hierarchical modeling of air-sea interaction. *Journal of Geophysical Research: Oceans*, 108(C4), 2003.

[16] Jorge L Bernal-Rusiel, Douglas N Greve, Martin Reuter, Bruce Fischl, Mert R Sabuncu, Alzheimer's Disease Neuroimaging Initiative, et al. Statistical analysis of longitudinal neuroimage data with linear mixed effects models. *Neuroimage*, 66:249–260, 2013.

[17] Milton C Biagioni and James E Galvin. Using biomarkers to improve detection of alzheimer's disease. *Neurodegenerative disease management*, 1(2):127–139, 2011.

[18] Aviroop Biswas, Paul I Oh, Guy E Faulkner, Ravi R Bajaj, Michael A Silver, Marc S Mitchell, and David A Alter. Sedentary time and its association with risk for disease incidence, mortality, and hospitalization in adults: a systematic review and meta-analysis. *Annals of internal medicine*, 162(2):123–132, 2015.

[19] Jan Christian Brønd and Daniel Arvidsson. Sampling frequency affects the processing of actigraph raw acceleration data to activity counts. *Journal of Applied Physiology*, 120(3):362–369, 2016.

[20] Babette A Brumback and John A Rice. Smoothing spline models for the analysis of nested and crossed samples of curves. *Journal of the American Statistical Association*, 93(443):961–976, 1998.

[21] Chris Brunsdon, Stewart Fotheringham, and Martin Charlton. Geographically weighted regression. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47(3):431–443, 1998.

[22] Flavia Bürgi, Ursina Meyer, Urs Granacher, Christian Schindler, Pedro Marques-Vidal, Susi Kriemler, and Jardena J Puder. Relationship of physical activity with motor skills, aerobic fitness and body fat in preschool children: a cross-sectional and longitudinal study (ballabeina). *International journal of obesity*, 35(7):937–944, 2011.

[23] V Carson, RL Rinaldi, B Torrance, K Maximova, GDC Ball, SR Majumdar, RC Plotnikoff, P Veugelers, NG Boulé, P Wozny, et al. Vigorous physical activity and longitudinal associations with cardiometabolic risk factors in youth. *International journal of obesity*, 38(1):16–21, 2014.

[24] CDC. Lack of Physical Activity. https://www.cdc.gov/chronicdisease/resources/publications/factsheets/physical-activity.htm, 2019. Accessed: 2022-03-21.

[25] CDC. Physical activity for a healthy weight, 2020.

[26] Leena Choi, Zhouwen Liu, Charles E Matthews, and Maciej S Buchowski. Validation of accelerometer wear and nonwear time classification algorithm. *Medicine and science in sports and exercise*, 43(2):357, 2011.

[27] Leena Choi, Zhouwen Liu, Charles E Matthews, Maciej S Buchowski, and Maintainer Leena Choi. Package 'physicalactivity'. 2018.

[28] Mei Sian Chong and Suresh Sahadevan. Preclinical alzheimer's disease: diagnosis and prediction of progression. *The Lancet Neurology*, 4(9):576–579, 2005.

[29] Lubomír Civín and Luboš Smutka. Vulnerability of european union economies in agro trade. *Sustainability*, 12(12):5210, 2020.

[30] Rachel C Colley, Didier Garriguet, Ian Janssen, Cora L Craig, Janine Clarke, and Mark S Tremblay. Physical activity of canadian adults: accelerometer results from the 2007 to 2009 canadian health measures survey. *Health reports*, 22(1):7, 2011.

[31] Ciprian M Crainiceanu, Ana-Maria Staicu, and Chong-Zhi Di. Generalized multilevel functional regression. *Journal of the American Statistical Association*, 104(488):1550–1561, 2009.

[32] Clara Deser, Adam Phillips, Vincent Bourdette, and Haiyan Teng. Uncertainty in climate change projections: the role of internal variability. *Climate dynamics*, 38(3):527–546, 2012.

[33] Chong-Zhi Di, Ciprian M Crainiceanu, Brian S Caffo, and Naresh M Punjabi. Multilevel functional principal component analysis. *The annals of applied statistics*, 3(1):458, 2009.

[34] Dmitry Divine. Climate time series analysis: Classical statistical and bootstrap methods, m. mudelsee, springer, dordrecht (2010), isbn: 978-90-481-9482-7, 2012.

[35] Michael C Donohue, Hélène Jacqmin-Gadda, Mélanie Le Goff, Ronald G Thomas, Rema Raman, Anthony C Gamst, Laurel A Beckett, Clifford R Jack Jr, Michael W Weiner, Jean-François Dartigues, et al. Estimating long-term multivariate progression from short-term data. *Alzheimer's & Dementia*, 10:S400–S410, 2014.

[36] Michael C Donohue, Reisa A Sperling, Ronald Petersen, Chung-Kai Sun, Michael W Weiner, Paul S Aisen, Alzheimer's Disease Neuroimaging Initiative, et al. Association between elevated brain amyloid and subsequent cognitive decline among cognitively normal persons. *Jama*, 317(22):2305–2316, 2017.

[37] Bruno Dubois, Howard H Feldman, Claudia Jacova, Steven T DeKosky, Pascale Barberger-Gateau, Jeffrey Cummings, André Delacourte, Douglas Galasko, Serge Gauthier, Gregory Jicha, et al. Research criteria for the diagnosis of alzheimer's disease: revising the nincds–adrda criteria. *The Lancet Neurology*, 6(8):734–746, 2007.

[38] Dorothea Dumuid, Željko Pedišić, Javier Palarea-Albaladejo, Josep Antoni Martín-Fernández, Karel Hron, and Timothy Olds. Compositional data analysis in time-use epidemiology: what, why, how. *International journal of environmental research and public health*, 17(7):2220, 2020.

[39] James Durbin and Geoffrey S Watson. Testing for serial correlation in least squares regression: I. *Biometrika*, 37(3/4):409–428, 1950.

[40] John J Dziak, Donna L Coffman, Matthew Reimherr, Justin Petrovich, Runze Li, Saul Shiffman, and Mariya P Shiyko. Scalar-on-function regression for predicting distal outcomes from intensively gathered longitudinal data: Interpretability for applied scientists. *Statistics surveys*, 13:150, 2019.

[41] Robert H Eckel, John M Jakicic, Jamy D Ard, Janet M de Jesus, Nancy Houston Miller, Van S Hubbard, I-Min Lee, Alice H Lichtenstein, Catherine M Loria, Barbara E Millen, et al. 2013 aha/acc guideline on lifestyle management to reduce cardiovascular risk: a report of the american college of cardiology/american heart association task force on practice guidelines. *Journal of the American college of cardiology*, 63(25 Part B):2960–2984, 2014.

[42] Bradley Efron and Robert J Tibshirani. *An introduction to the bootstrap*. CRC press, 1994.

[43] FDA. FDA Grants Accelerated Approval for Alzheimer's Drug. https://www.fda.gov/news-events/press-announcements/fda-grants-accelerated-approval-alzheimers-drug, 2021. Accessed: 2022-03-21.

[44] Garrett M Fitzmaurice, Nan M Laird, and James H Ware. *Applied longitudinal analysis*. John Wiley & Sons, 2012.

[45] A Stewart Fotheringham, Ricardo Crespo, and Jing Yao. Geographical and temporal weighted regression (gtwr). *Geographical Analysis*, 47(4):431–452, 2015.

[46] Caroline S Fox, Sherita Hill Golden, Cheryl Anderson, George A Bray, Lora E Burke, Ian H De Boer, Prakash Deedwania, Robert H Eckel, Abby G Ershow, Judith Fradkin, et al. Update on prevention of cardiovascular disease in adults with type 2 diabetes mellitus in light of recent evidence: a scientific statement from the american heart association and the american diabetes association. *Circulation*, 132(8):691–718, 2015.

[47] Patty Freedson, Heather R Bowles, Richard Troiano, and William Haskell. Assessment of physical activity using wearable monitors: recommendations for monitor calibration and use in the field. *Medicine and science in sports and exercise*, 44(1 Suppl 1):S1, 2012.

[48] Patty Freedson, David Pober, Kathleen F Janz, et al. Calibration of accelerometer output for children. *Medicine and science in sports and exercise*, 37(11):S523, 2005.

[49] Joshua P French. Confidence regions for the level curves of spatial data. *Environmetrics*, 25(7):498–512, 2014.

[50] Joshua P French, Seth McGinnis, and Armin Schwartzman. Assessing narccap climate model effects using spatial confidence regions. *Advances in statistical climatology, meteorology and oceanography*, 3(2):67, 2017.

[51] Joshua P French and Stephan R Sain. Spatio-temporal exceedance locations and confidence regions. *The Annals of Applied Statistics*, 7(3):1421–1449, 2013.

[52] Giovanni B Frisoni, Nick C Fox, Clifford R Jack, Philip Scheltens, and Paul M Thompson. The clinical use of structural mri in alzheimer disease. *Nature Reviews Neurology*, 6(2):67–77, 2010.

[53] Sylvia Frühwirth-Schnatter and Helga Wagner. Stochastic model specification search for gaussian and partial non-gaussian state space models. *Journal of Econometrics*, 154(1):85–100, 2010.

[54] Cheryl D Fryar, Margaret D Carroll, and Cynthia L Ogden. Prevalence of overweight, obesity, and severe obesity among children and adolescents aged 2–19 years: United states, 1963–1965 through 2015–2016. 2018.

[55] Andrew Gelman et al. Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper). *Bayesian analysis*, 1(3):515–534, 2006.

[56] M Geraci. pawacc: Physical activity with accelerometers. 2012.

[57] Jan Gertheiss, Jeff Goldsmith, Ciprian Crainiceanu, and Sonja Greven. Longitudinal scalar-on-functions regression with application to tractography data. *Biostatistics*, 14(3):447–461, 2013.

[58] Filippo Giorgi, Colin Jones, Ghassem R Asrar, et al. Addressing climate information needs at the regional level: the cordex framework. *World Meteorological Organization (WMO) Bulletin*, 58(3):175, 2009.

[59] Jeff Goldsmith, Ciprian M Crainiceanu, Brian Caffo, and Daniel Reich. Longitudinal penalized functional regression for cognitive outcomes on neuronal tract measurements. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 61(3):453–469, 2012.

[60] Sonja Greven, Ciprian Crainiceanu, Brian Caffo, and Daniel Reich. Longitudinal functional principal component analysis. In *Recent Advances in Functional Data Analysis and Related Topics*, pages 149–154. Springer, 2011.

[61] Craig M Hales, Cheryl D Fryar, Margaret D Carroll, David S Freedman, Yutaka Aoki, and Cynthia L Ogden. Differences in obesity prevalence by demographic characteristics and urbanization level among adults in the united states, 2013-2016. *Jama*, 319(23):2419–2429, 2018.

[62] Bernard J Hanseeuw, Rebecca A Betensky, Heidi IL Jacobs, Aaron P Schultz, Jorge Sepulcre, J Alex Becker, Danielle M Orozco Cosio, Michelle Farrell, Yakeel T Quiroz, Elizabeth C Mormino, et al. Association of amyloid and tau with cognition in preclinical alzheimer disease: a longitudinal study. *JAMA neurology*, 76(8):915–924, 2019.

[63] Sarah K Harding, Angie S Page, Catherine Falconer, and Ashley R Cooper. Longitudinal changes in sedentary time and physical activity during adolescence. *International Journal of Behavioral Nutrition and Physical Activity*, 12(1):1–7, 2015.

[64] A Dewi Hartkamp, Kirsten De Beurs, Alfred Stein, and Jeffrey W White. Interpolation techniques for climate variables, 1999.

[65] William L Haskell, I-Min Lee, Russell R Pate, Kenneth E Powell, Steven N Blair, Barry A Franklin, Caroline A Macera, Gregory W Heath, Paul D Thompson, and Adrian Bauman. Physical activity and public health: updated recommendation for adults from the american college of sports medicine and the american heart association. *Circulation*, 116(9):1081, 2007.

[66] Hossein Hassani and Mohammad Reza Yeganegi. Selecting optimal lag order in ljung–box test. *Physica A: Statistical Mechanics and its Applications*, 541:123700, 2020.

[67] Trevor J Hastie and Robert J Tibshirani. *Generalized additive models*. Routledge, 2017.

[68] Hendrik Hendrik JF Helmerhorst, Søren Brage, Janet Warren, Herve Besson, and Ulf Ekelund. A systematic review of reliability and objective criterion-related validity of physical activity questionnaires. *International Journal of Behavioral Nutrition and Physical Activity*, 9(1):1–55, 2012.

[69] Barbara Hennemuth, S Bender, K Bülow, N Dreier, E Keup-Thiel, O Krüger, C Mudersbach, C Radermacher, and R Schoetter. Statistical methods for the analysis of simulated and observed climate data: applied in projects and institutions dealing with climate change impact and adaptation. CSC, 2013.

[70] Tamás Illy Hennemuth, Daniela Jacob, Elke Keup-Thiel, Sven Kotlarski, Grigory Nikulin, Juliane Otto, and G Szépszó. Guidance for euro-cordex climate projections data use. *Version1. 0-2017.08. Retrieved on*, 6:2019, 2017.

[71] Richard A Hickman, Arline Faustin, and Thomas Wisniewski. Alzheimer disease and its growing epidemic: risk factors, biomarkers, and the urgent need for therapeutics. *Neurologic clinics*, 34(4):941–953, 2016.

[72] Chris C Holmes, Leonhard Held, et al. Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian analysis*, 1(1):145–168, 2006.

[73] Cheryl A Howe, John W Staudenmayer, and Patty S Freedson. Accelerometer prediction of energy expenditure: vector magnitude versus vertical axis. *Med Sci Sports Exerc*, 41(12):2199–2206, 2009.

[74] Chi-yuan Hsu, Charles E McCulloch, Carlos Iribarren, Jeanne Darbinian, and Alan S Go. Body mass index and risk for end-stage renal disease. *Annals of internal medicine*, 144(1):21–28, 2006.

[75] Alan Huang, Matthew P Wand, et al. Simple marginally noninformative prior distributions for covariance matrices. *Bayesian Analysis*, 8(2):439–452, 2013.

[76] Bo Huang, Bo Wu, and Michael Barry. Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *International Journal of Geographical Information Science*, 24(3):383–401, 2010.

[77] Charles P Hughes, Leonard Berg, Warren Danziger, Lawrence A Coben, and Ronald L Martin. A new clinical scale for the staging of dementia. *The British journal of psychiatry*, 140(6):566–572, 1982.

[78] Clifford M Hurvich, Jeffrey S Simonoff, and Chih-Ling Tsai. Smoothing parameter selection in nonparametric regression using an improved akaike information criterion. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(2):271–293, 1998.

[79] Core Writing Team IPCC. *Climate change 2007: synthesis report*. IPCC Geneva, Switzerland, 2007.

[80] TAR IPCC. Climate change 2001: synthesis report. *Intergovernmental Panel on Climate Change (IPCC), Geneva, Switzerland*, 2001.

[81] Clifford R Jack Jr, Marilyn S Albert, David S Knopman, Guy M McKhann, Reisa A Sperling, Maria C Carrillo, Bill Thies, and Creighton H Phelps. Introduction to the recommendations from the national institute on aging-alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease. *Alzheimer's & dementia*, 7(3):257–262, 2011.

[82] Clifford R Jack Jr, David S Knopman, William J Jagust, Ronald C Petersen, Michael W Weiner, Paul S Aisen, Leslie M Shaw, Prashanthi Vemuri, Heather J Wiste, Stephen D Weigand, et al. Update on hypothetical model of alzheimer's disease biomarkers. *Lancet neurology*, 12(2):207, 2013.

[83] Clifford R Jack Jr, David S Knopman, William J Jagust, Leslie M Shaw, Paul S Aisen, Michael W Weiner, Ronald C Petersen, and John Q Trojanowski. Hypothetical model of dynamic biomarkers of the alzheimer's pathological cascade. *The Lancet Neurology*, 9(1):119–128, 2010.

[84] Clifford R Jack Jr, Heather J Wiste, Stephen D Weigand, Terry M Therneau, Val J Lowe, David S Knopman, Jeffrey L Gunter, Matthew L Senjem, David T Jones, Kejal Kantarci, et al. Defining imaging biomarker cut points for brain aging and alzheimer's disease. *Alzheimer's & Dementia*, 13(3):205–216, 2017.

[85] Dinesh John and Patty Freedson. Actigraph and actical physical activity monitors: a peek under the hood. *Medicine and science in sports and exercise*, 44(1 Suppl 1):S86, 2012.

[86] Richard Arnold Johnson, Dean W Wichern, et al. *Applied multivariate statistical analysis*, volume 6. Pearson London, UK:, 2014.

[87] Ian T Jolliffe. *Principal component analysis for special types of data*. Springer, 2002.

[88] Kari Karhunen. *Ueber lineare Methoden in der Wahrscheinlichkeitsrechnung*. Soumalainen Tiedeakatemia, 1947.

[89] Gary G Koch. Some further remarks concerning "a general approach to the estimation of variance components". *Technometrics*, 10(3):551–558, 1968.

[90] Hubert Kolb, Michael Stumvoll, Werner Kramer, Kerstin Kempf, and Stephan Martin. Insulin translates unfavourable lifestyle into obesity. *BMC medicine*, 16(1):1–10, 2018.

[91] Rao Kotamarthi, Linda Mearns, Katharine Hayhoe, Christoper L Castro, and Donald Wuebbles. Use of climate information for decision-making and impacts research: State of our understanding. Technical report, Argonne National Laboratory Argonne United States, 2016.

[92] Oliver Krueger and Jin-Song Von Storch. A simple empirical model for decadal climate prediction. *Journal of climate*, 24(4):1276–1283, 2011.

[93] Lawrence H Kushi, Colleen Doyle, Marji McCullough, Cheryl L Rock, Wendy Demark-Wahnefried, Elisa V Bandera, Susan Gapstur, Alpa V Patel, Kimberly Andrews, Ted Gansler, et al. American cancer society guidelines on nutrition and physical activity for cancer prevention: reducing the risk of cancer with healthy food choices and physical activity. *CA: a cancer journal for clinicians*, 62(1):30–67, 2012.

[94] Dongbing Lai, Huiping Xu, Daniel Koller, Tatiana Foroud, and Sujuan Gao. A multivariate finite mixture latent trajectory model with application to dementia studies. *Journal of applied statistics*, 43(14):2503–2523, 2016.

[95] Nan M Laird and James H Ware. Random-effects models for longitudinal data. *Biometrics*, pages 963–974, 1982.

[96] Susan M Landau, Danielle Harvey, Cindee M Madison, Robert A Koeppe, Eric M Reiman, Norman L Foster, Michael W Weiner, William J Jagust, Alzheimer's Disease Neuroimaging Initiative, et al. Associations between cognitive, functional, and fdg-pet measures of decline in ad and mci. *Neurobiology of aging*, 32(7):1207–1218, 2011.

[97] Tran Le, Shirley W Flatt, Loki Natarajan, Bilge Pakiz, Elizabeth L Quintana, Dennis D Heath, Brinda K Rana, and Cheryl L Rock. Effects of diet composition and insulin resistance status on plasma lipid levels in a weight loss intervention in women. *Journal of the American Heart Association*, 5(1):e002771, 2016.

[98] Seungyong Lee, George Wolberg, and Sung Yong Shin. Scattered data interpolation with multilevel b-splines. *IEEE transactions on visualization and computer graphics*, 3(3):228–244, 1997.

[99] Sik-Yum Lee. *Structural equation modeling: A Bayesian approach*. John Wiley & Sons, 2007.

[100] Jeannie-Marie S Leoutsakos, Sarah N Forrester, Christopher D Corcoran, Maria C Norton, Peter V Rabins, Martin I Steinberg, Joann T Tschanz, and Constantine G Lyketsos. Latent classes of course in alzheimer's disease and predictors: the cache county dementia progression study. *International journal of geriatric psychiatry*, 30(8):824–832, 2015.

[101] Feng V Lin, Xixi Wang, Rachel Wu, George W Rebok, Benjamin P Chapman, Alzheimer's Disease Neuroimaging Initiative, et al. Identification of successful cognitive aging in the alzheimer's disease neuroimaging initiative study. *Journal of Alzheimer's Disease*, 59(1):101–111, 2017.

[102] Chia-Chen Liu, Takahisa Kanekiyo, Huaxi Xu, and Guojun Bu. Apolipoprotein e and alzheimer disease: risk, mechanisms and therapy. *Nature Reviews Neurology*, 9(2):106–118, 2013.

[103] Ying Liu, Lan Tan, Hui-Fu Wang, Yong Liu, Xiao-Ke Hao, Chen-Chen Tan, Teng Jiang, Bing Liu, Dao-Qiang Zhang, and Jin-Tai Yu. Multiple effect of apoe genotype on clinical and neuroimaging biomarkers across alzheimer's disease spectrum. *Molecular neurobiology*, 53(7):4539–4547, 2016.

[104] Greta M Ljung and George EP Box. On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303, 1978.

[105] Yungtai Lo, Nancy R Mendell, and Donald B Rubin. Testing the number of components in a normal mixture. *Biometrika*, 88(3):767–778, 2001.

[106] Iryna Lobanova and Adnan I Qureshi. The association between cardiovascular risk factors and progressive hippocampus volume loss in persons with alzheimer's disease. *Journal of vascular and interventional neurology*, 7(5):52, 2014.

[107] M. Loeve. *Probability Theory II*. F.W.Gehring P.r.Halmos and C.c.Moore. Springer, 1978.

[108] Paul D Loprinzi and Eveleen Sng. The effects of objectively measured sedentary behavior on all-cause mortality in a national sample of adults with diabetes. *Preventive medicine*, 86:55–57, 2016.

[109] Amy Luke, Lara R Dugas, Ramon A Durazo-Arvizu, Guichan Cao, and Richard S Cooper. Assessing physical activity and its relationship to cardiovascular risk factors: Nhanes 2003-2006. *BMC Public Health*, 11(1):1–11, 2011.

[110] Joanne R Lupton, JA Brooks, NF Butte, B Caballero, JP Flatt, SK Fried, et al. Dietary reference intakes for energy, carbohydrate, fiber, fat, fatty acids, cholesterol, protein, and amino acids. *National Academy Press: Washington, DC, USA*, 5:589–768, 2002.

[111] Merryn J Mathie, Adelle CF Coster, Nigel H Lovell, and Branko G Celler. Accelerometry: providing an integrated, practical method for long-term, ambulatory monitoring of human movement. *Physiological measurement*, 25(2):R1, 2004.

[112] Charles E Matthews, Kong Y Chen, Patty S Freedson, Maciej S Buchowski, Bettina M Beech, Russell R Pate, and Richard P Troiano. Amount of time spent in sedentary behaviors in the united states, 2003–2004. *American journal of epidemiology*, 167(7):875–881, 2008.

[113] Mayo Clinic Staff. Alzheimer's disease. https://www.mayoclinic.org/diseases-conditions/alzheimers-disease/symptoms-causes/syc-20350447, 2022. Accessed: 2022-03-21.

[114] Linda K McEvoy and James B Brewer. Biomarkers for the clinical evaluation of the cognitively impaired elderly: amyloid is not enough. *Imaging in medicine*, 4(3):343, 2012.

[115] Guy M McKhann, David S Knopman, Howard Chertkow, Bradley T Hyman, Clifford R Jack Jr, Claudia H Kawas, William E Klunk, Walter J Koroshetz, Jennifer J Manly, Richard Mayeux, et al. The diagnosis of dementia due to alzheimer's disease: Recommendations from the national institute on aging-alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease. *Alzheimer's & dementia*, 7(3):263–269, 2011.

[116] Linda O Mearns, Ray Arritt, Sébastien Biner, Melissa S Bukovsky, Seth McGinnis, Stephan Sain, Daniel Caya, James Correia, Dave Flory, William Gutowski, et al. The north american regional climate change assessment program: overview of phase i results. *Bulletin of the American Meteorological Society*, 93(9):1337–1362, 2012.

[117] LO Mearns, S McGinnis, D Korytina, R Arritt, S Biner, M Bukovsky, HI Chang, O Christensen, D Herzmann, Y Jiao, et al. The na-cordex dataset, version 1.0. *NCAR Climate Data Gateway. Boulder (CO): The North American CORDEX Program*, 10:D6SJ1JCH, 2017.

[118] Jerry M Melillo, TT Richmond, G Yohe, et al. Climate change impacts in the united states. *Third national climate assessment*, 52, 2014.

[119] JM Melillo, TC Richmond, and GW Yohe. Our changing climate. *Climate Change Impacts in the United States: The Third National Climate Assessment*, 19:67, 2014.

[120] Expert Panel Members, Michael D Jensen, Donna H Ryan, Karen A Donato, Caroline M Apovian, Jamy D Ard, Anthony G Comuzzie, Frank B Hu, Van S Hubbard, John M Jakicic, et al. Executive summary: guidelines (2013) for the management of overweight and obesity in adults: a report of the american college of cardiology/american heart association task force on practice guidelines and the obesity society published by the obesity society and american college of cardiology/american heart association task force on practice guidelines. based on a systematic review from the the obesity expert panel, 2013. *Obesity*, 22(S2):S5–S39, 2014.

[121] James Mercer. Xvi. functions of positive and negative type, and their connection the theory of integral equations. *Philosophical transactions of the royal society of London. Series A, containing papers of a mathematical or physical character*, 209(441-458):415–446, 1909.

[122] SC Michaelides, FS Tymvios, and T Michaelidou. Spatial and temporal characteristics of the annual rainfall frequency distribution in cyprus. *Atmospheric Research*, 94(4):606–615, 2009.

[123] Jeffrey S Morris, Cassandra Arroyo, Brent A Coull, Louise M Ryan, Richard Herrick, and Steven L Gortmaker. Using wavelet-based functional mixed models to characterize population heterogeneity in accelerometer profiles: a case study. *Journal of the American Statistical Association*, 101(476):1352–1364, 2006.

[124] Jeffrey S Morris and Raymond J Carroll. Wavelet-based functional mixed models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(2):179–199, 2006.

[125] John C Morris. The clinical dementia rating (cdr): Current version and. *Young*, 41:1588–1592, 1991.

[126] John C Morris, Kaj Blennow, Lutz Frölich, Agneta Nordberg, Hilkka Soininen, Gunhild Waldemar, L-O Wahlund, and B Dubois. Harmonized diagnostic criteria for alzheimer's disease: recommendations. *Journal of internal medicine*, 275(3):204–213, 2014.

[127] Manfred Mudelsee. Trend analysis of climate time series: A review of methods. *Earth-science reviews*, 190:310–322, 2019.

[128] Susanne G Mueller, Michael W Weiner, Leon J Thal, Ronald C Petersen, Clifford R Jack, William Jagust, John Q Trojanowski, Arthur W Toga, and Laurel Beckett. Ways toward an early diagnosis in alzheimer's disease: the alzheimer's disease neuroimaging initiative (adni). *Alzheimer's & Dementia*, 1(1):55–66, 2005.

[129] Akihiko Murata, Shun-ichi I Watanabe, Hidetaka Sasaki, Hiroaki Kawase, and Masaya Nosaka. Assessing goodness of fit to a gamma distribution and estimating future projection on daily precipitation frequency using regional climate model simulations over japan with and without the influence of tropical cyclones. *Journal of Hydrometeorology*, 21(12):2997–3010, 2020.

[130] Bengt Muthén and Linda K Muthén. Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and experimental research*, 24(6):882–891, 2000.

[131] Bengt Muthén and Kerby Shedden. Finite mixture modeling with mixture outcomes using the em algorithm. *Biometrics*, 55(2):463–469, 1999.

[132] Radford M Neal. *Probabilistic inference using Markov chain Monte Carlo methods*. Department of Computer Science, University of Toronto Toronto, ON, Canada, 1993.

[133] John R Nesselroade. Interindividual differences in intraindividual change. 1991.

[134] Ninh T Nguyen, Cheryl P Magno, Karen T Lane, Marcelo W Hinojosa, and John S Lane. Association of hypertension, diabetes, dyslipidemia, and metabolic syndrome with obesity: findings from the national health and nutrition examination survey, 1999 to 2004. *Journal of the American College of Surgeons*, 207(6):928–934, 2008.

[135] Robert Nisticò and John Joseph Borg. Aducanumab for alzheimer's disease: A regulatory perspective. *Pharmacological Research*, 171:105754, 2021.

[136] World Health Organization. Physical Activity. https://www.who.int/news-room/fact-sheets/detail/physical-activity, 2020. Accessed: 2022-03-21.

[137] Henning Palmefors, Smita DuttaRoy, Bengt Rundqvist, and Mats Börjesson. The effect of physical activity or exercise on key biomarkers in atherosclerosis–a systematic review. *Atherosclerosis*, 235(1):150–161, 2014.

[138] Panagiotis Papastamoulis. label. switching: An r package for dealing with the label switching problem in mcmc outputs. *arXiv preprint arXiv:1503.02271*, 2015.

[139] Ronald C Petersen. Mild cognitive impairment as a diagnostic entity. *Journal of internal medicine*, 256(3):183–194, 2004.

[140] Ronald Carl Petersen, PS Aisen, Laurel A Beckett, MC Donohue, AC Gamst, Danielle J Harvey, CR Jack, WJ Jagust, LM Shaw, AW Toga, et al. Alzheimer's disease neuroimaging initiative (adni): clinical characterization. *Neurology*, 74(3):201–209, 2010.

[141] Robert H Pietrzak, Yen Ying Lim, David Ames, Karra Harrington, Carolina Restrepo, Ralph N Martins, Alan Rembach, Simon M Laws, Colin L Masters, Victor L Villemagne, et al. Trajectories of memory decline in preclinical alzheimer's disease: results from the australian imaging, biomarkers and lifestyle flagship study of ageing. *Neurobiology of aging*, 36(3):1231–1238, 2015.

[142] Judes Poirier, P Bertrand, S Kogan, S Gauthier, J Davignon, and D Bouthillier. Apolipoprotein e polymorphism and alzheimer's disease. *The Lancet*, 342(8873):697–699, 1993.

[143] Nicholas G Polson, James G Scott, and Jesse Windle. Bayesian inference for logistic models using pólya–gamma latent variables. *Journal of the American statistical Association*, 108(504):1339–1349, 2013.

[144] Stéphanie A Prince, Kristi B Adamo, Meghan E Hamel, Jill Hardt, Sarah Connor Gorber, and Mark Tremblay. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. *International journal of behavioral nutrition and physical activity*, 5(1):1–24, 2008.

[145] Nilam Ram and Kevin J Grimm. Methods and measures: Growth mixture modeling: A method for identifying differences in longitudinal change among unobserved groups. *International journal of behavioral development*, 33(6):565–576, 2009.

[146] James O Ramsay. Functional data analysis. *Encyclopedia of Statistical Sciences*, 4, 2004.

[147] James O Ramsay and CJ Dalzell. Some tools for functional data analysis. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(3):539–561, 1991.

[148] James O Ramsay and Bernard W Silverman. *Fitting differential equations to functional data: Principal differential analysis*. Springer, 2005.

[149] James O Ramsay and Bernhard W Silverman. Functional data analysis. *İnternet Adresi: http*, 2008.

[150] Philip T Reiss and R Todd Ogden. Functional principal component regression and functional partial least squares. *Journal of the American Statistical Association*, 102(479):984–996, 2007.

[151] Leandro Fornias Machado de Rezende, Maurício Rodrigues Lopes, Juan Pablo Rey-López, Victor Keihan Rodrigues Matsudo, and Olinda do Carmo Luiz. Sedentary behavior and health outcomes: an overview of systematic reviews. *PloS one*, 9(8):e105620, 2014.

[152] John A Rice. Functional and longitudinal data analysis: perspectives on smoothing. *Statistica Sinica*, pages 631–647, 2004.

[153] John A Rice and Bernard W Silverman. Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(1):233–243, 1991.

[154] Cheryl L Rock, Shirley W Flatt, Bilge Pakiz, Elizabeth L Quintana, Dennis D Heath, Brinda K Rana, and Loki Natarajan. Effects of diet composition on weight loss, metabolic factors and biomarkers in a 1-year weight loss intervention in obese women examined by baseline insulin resistance status. *Metabolism*, 65(11):1605–1613, 2016.

[155] Wilma G Rosen, Richard C Mohs, and Kenneth L Davis. A new rating scale for alzheimer's disease. *The American journal of psychiatry*, 1984.

[156] Donald B Rubin. Bayesian data analysis. 2013.

[157] Alok Kumar Samantaray, Adway Mitra, Meenu Ramadas, and Rabindra Kumar Panda. Regionalization of hydroclimatic variables using markov random field model for climate change impact assessment. *Journal of Hydrology*, 596:126071, 2021.

[158] Jeffer E Sasaki, Dinesh John, and Patty S Freedson. Validation and comparison of actigraph activity monitors. *Journal of science and medicine in sport*, 14(5):411–416, 2011.

[159] JF Scinocca, VV Kharin, Y Jiao, MW Qian, M Lazare, L Solheim, GM Flato, S Biner, M Desgagne, and B Dugas. Coordinated global and regional climate modeling. *Journal of Climate*, 29(1):17–35, 2016.

[160] Mait Sepp and Jaak Jaagus. Frequency of circulation patterns and air temperature variations in europe. *Boreal Environment Research*, 7(3):273–280, 2002.

[161] Francesco Sera, Lucy J Griffiths, Carol Dezateux, Marco Geraci, and Mario Cortina-Borja. Using functional data analysis to understand daily activity levels and patterns in primary school-aged children: cross-sectional analysis of a uk-wide study. *PLoS one*, 12(11):e0187677, 2017.

[162] Han Lin Shang. A survey of functional principal component analysis. *AStA Advances in Statistical Analysis*, 98(2):121–142, 2014.

[163] Samuel Sanford Shapiro and Martin B Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611, 1965.

[164] Haochang Shou, Vadim Zipunnikov, Ciprian M Crainiceanu, and Sonja Greven. Structured functional principal component analysis. *Biometrics*, 71(1):247–257, 2015.

[165] Max Sommerfeld, Stephan Sain, and Armin Schwartzman. Confidence regions for spatial excursion sets from repeated random field observations, with an application to climate. *Journal of the American Statistical Association*, 113(523):1327–1340, 2018.

[166] Max Sommerfeld, Stephen Sain, and Armin Schwartzman. Confidence regions for excursion sets in asymptotically gaussian random fields, with an application to climate. *arXiv preprint arXiv:1501.07000*, 2015.

[167] Reisa A Sperling, Paul S Aisen, Laurel A Beckett, David A Bennett, Suzanne Craft, Anne M Fagan, Takeshi Iwatsubo, Clifford R Jack Jr, Jeffrey Kaye, Thomas J Montine, et al. Toward defining the preclinical stages of alzheimer's disease: Recommendations from the national institute on aging-alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease. *Alzheimer's & dementia*, 7(3):280–292, 2011.

[168] Fiona Steele, Paul Clarke, George Leckie, Julia Allan, and Derek Johnston. Multilevel structural equation models for longitudinal data where predictors are measured more frequently than outcomes: An application to the effects of stress on the cognitive function of nurses. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 180(1):263–283, 2017.

[169] Matthew Stephens. Dealing with label switching in mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(4):795–809, 2000.

[170] Hossein Tabari, Meron Teferi Taye, and Patrick Willems. Statistical assessment of precipitation trends in the upper blue nile river basin. *Stochastic environmental research and risk assessment*, 29(7):1751–1761, 2015.

[171] Andrew Tait, Roddy Henderson, Richard Turner, and Xiaogu Zheng. Thin plate smoothing spline interpolation of daily rainfall for new zealand using a climatological rainfall surface. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 26(14):2097–2115, 2006.

[172] Richard P Troiano, David Berrigan, Kevin W Dodd, Louise C Masse, Timothy Tilert, Margaret McDowell, et al. Physical activity in the united states measured by accelerometer. *Medicine and science in sports and exercise*, 40(1):181, 2008.

[173] Richard P Troiano, Caroline A Macera, and Rachel Ballard-Barbash. Be physically active each day. how can we know? *The Journal of nutrition*, 131(2):451S–460S, 2001.

[174] Richard P Troiano, James J McClain, Robert J Brychta, and Kong Y Chen. Evolution of accelerometer methods for physical activity research. *British journal of sports medicine*, 48(13):1019–1023, 2014.

[175] Stewart G Trost, Russell R Pate, Patty S Freedson, James F Sallis, and Wendell C Taylor. Using objective physical activity measures with youth: how many days of monitoring are needed? *Medicine & Science in Sports & Exercise*, 32(2):426, 2000.

[176] Catrine Tudor-Locke, Meghan M Brashear, William D Johnson, and Peter T Katzmarzyk. Accelerometer profiles of physical activity and inactivity in normal weight, overweight, and obese us men and women. *International Journal of Behavioral Nutrition and Physical Activity*, 7(1):1–11, 2010.

[177] Scott E Umbaugh. *Digital image processing and analysis: human and computer vision applications with CVIPtools*. CRC press, 2010.

[178] P Vemuri, HJ Wiste, SD Weigand, LM Shaw, JQ Trojanowski, MW Weiner, David S Knopman, Ronald Carl Petersen, CR Jack, et al. Mri and csf biomarkers in normal, mci, and ad subjects: predicting future clinical change. *Neurology*, 73(4):294–301, 2009.

[179] Prashanthi Vemuri and Clifford R Jack. Role of structural mri in alzheimer's disease. *Alzheimer's research & therapy*, 2(4):1–10, 2010.

[180] J Walsh, D Weuebbles, K Hayhoe, J Kossin, K Kunkel, G Stephens, et al. Chapter 2: Our changing climate in: Melillo j, richmond t, yohe g, editors. climate change impacts in the united states: The third national climate assessment. washington, dc: Us global change research program; 2014, 2017.

[181] Jane-Ling Wang, Jeng-Min Chiou, and Hans-Georg Müller. Functional data analysis. *Annual Review of Statistics and Its Application*, 3:257–295, 2016.

[182] Amal A Wanigatunga, Eleanor M Simonsick, Vadim Zipunnikov, Adam P Spira, Stephanie Studenski, Luigi Ferrucci, and Jennifer A Schrack. Perceived fatigability and objective physical activity in mid-to late-life. *The Journals of Gerontology: Series A*, 73(5):630–635, 2018.

[183] Mohammad Wasef Hattab. A derivation of prediction intervals for gamma regression. *Journal of Statistical Computation and Simulation*, 86(17):3512–3526, 2016.

[184] Sumio Watanabe and Manfred Opper. Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of machine learning research*, 11(12), 2010.

[185] Alette M Wessels, SA Dowsett, and JR Sims. Detecting treatment group differences in alzheimer's disease clinical trials: a comparison of alzheimer's disease assessment scale-cognitive subscale (adas-cog) and the clinical dementia rating-sum of boxes (cdr-sb). *The journal of prevention of Alzheimer's disease*, 5(1):15–20, 2018.

[186] Christopher K Wikle, Andrew Zammit-Mangion, and Noel Cressie. *Spatio-temporal Statistics with R*. Chapman and Hall/CRC, 2019.

[187] John Wiley. Alzheimer's disease facts and figures. *Alzheimers Dement*, 17:327–406, 2021.

[188] Daniel S Wilks and Keith L Eggleston. Estimating monthly and seasonal precipitation distributions using the 30-and 90-day outlooks. *Journal of climate*, 5(3):252–259, 1992.

[189] Simon N Wood. *Generalized additive models: an introduction with R*. chapman and hall/CRC, 2006.

[190] Simon N Wood. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(1):3–36, 2011.

[191] Simon N Wood. *Generalized additive models: an introduction with R*. CRC press, 2017.

[192] Luo Xiao, Vadim Zipunnikov, David Ruppert, and Ciprian Crainiceanu. Fast covariance estimation for high-dimensional functional data. *Statistics and computing*, 26(1):409–421, 2016.

[193] Selene Yue Xu, Sandahl Nelson, Jacqueline Kerr, Suneeta Godbole, Eileen Johnson, Ruth E Patterson, Cheryl L Rock, Dorothy D Sears, Ian Abramson, and Loki Natarajan. Modeling temporal variation in physical activity using functional principal components analysis. *Statistics in Biosciences*, 11(2):403–421, 2019.

[194] Che-Chang Yang and Yeh-Liang Hsu. A review of accelerometry-based wearable motion detectors for physical activity monitoring. *Sensors*, 10(8):7772–7788, 2010.

[195] Fang Yao, Hans-Georg Müller, and Jane-Ling Wang. Functional data analysis for sparse longitudinal data. *Journal of the American statistical association*, 100(470):577–590, 2005.

[196] Liming Ye, Guixia Yang, Eric Van Ranst, and Huajun Tang. Time-series modeling and prediction of global monthly absolute temperature for environmental decision making. *Advances in Atmospheric Sciences*, 30(2):382–396, 2013.

[197] Sheng Yue, Paul Pilon, and George Cavadias. Power of the mann–kendall and spearman's rho tests for detecting monotonic trends in hydrological series. *Journal of hydrology*, 259(1-4):254–271, 2002.

[198] Yukun Zhang, Haocheng Li, Sarah Kozey Keadle, Charles E Matthews, and Raymond J Carroll. A review of statistical analyses on physical activity data collected from accelerometers. *Statistics in biosciences*, 11(2):465–476, 2019.

[199] Vadim Zipunnikov, Brian Caffo, David M Yousem, Christos Davatzikos, Brian S Schwartz, and Ciprian Crainiceanu. Multilevel functional principal component analysis for high-dimensional data. *Journal of Computational and Graphical Statistics*, 20(4):852–873, 2011.

[200] Vadim Zipunnikov, Sonja Greven, Haochang Shou, Brian Caffo, Daniel S Reich, and Ciprian Crainiceanu. Longitudinal high-dimensional principal components analysis with application to diffusion tensor imaging of multiple sclerosis. *The annals of applied statistics*, 8(4):2175, 2014.