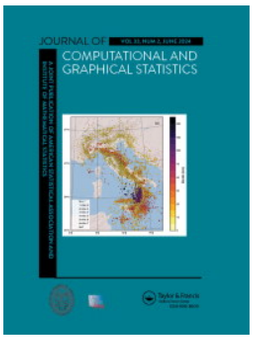

Journal of Computational and Graphical Statistics

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/ucgs20

Dynamic Survival Prediction Using Sparse Longitudinal Images via Multi-Dimensional Functional Principal Component Analysis

Haolun Shi, Shu Jiang, Da Ma, Mirza Faisal Beg & Jiguo Cao

To cite this article: Haolun Shi, Shu Jiang, Da Ma, Mirza Faisal Beg & Jiguo Cao (23 May 2024): Dynamic Survival Prediction Using Sparse Longitudinal Images via Multi-Dimensional Functional Principal Component Analysis, Journal of Computational and Graphical Statistics, DOI: 10.1080/10618600.2024.2335182

To link to this article: https://doi.org/10.1080/10618600.2024.2335182

JOURNAL OF COMPUTATIONAL AND GRAPHICAL STATISTICS
2024, VOL. 00, NO. 0, 1–12
https://doi.org/10.1080/10618600.2024.2335182

Taylor & Francis
Taylor & Francis Group

Dynamic Survival Prediction Using Sparse Longitudinal Images via Multi-Dimensional Functional Principal Component Analysis

Haolun Shi[a], Shu Jiang[b], Da Ma[c], Mirza Faisal Beg[d], and Jiguo Cao[a]

[a]Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, Canada; [b]Division of Public Health Sciences, Washington University School of Medicine in St. Louis, St. Louis, MO; [c]School of Medicine, Wake Forest University, Winston-Salem, NC; [d]School of Engineering, Simon Fraser University, Burnaby, Canada

ABSTRACT

Our work is motivated by predicting the progression of Alzheimer's disease (AD) based on a series of longitudinally observed brain scan images. Existing works on dynamic prediction for AD focus primarily on extracting predictive information from multivariate longitudinal biomarker values or brain imaging data at the baseline; whereas in practice, the subject's brain scan image represented by a multi-dimensional data matrix is collected at each follow-up visit. It is of great interest to predict the progression of AD directly from a series of longitudinally observed images. We propose a novel multi-dimensional functional principal component analysis based on alternating regression on tensor-product B-spline, which circumvents the computational difficulty of doing eigendecomposition, and offers the flexibility of accommodating sparsely and irregularly observed image series. We then use the functional principal component scores as features in the Cox proportional hazards model. We further develop a dynamic prediction framework to provide a personalized prediction that can be updated as new images are collected. Our method extracts visibly interpretable images of the functional principal components and offers an accurate prediction of the conversion to AD. We examine the effectiveness of our method via simulation studies and illustrate its application on the Alzheimer's Disease Neuroimaging Initiative data. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received April 2023
Accepted March 2024

KEYWORDS
Dynamic prediction; Eigendecomposition; Functional data analysis; Functional principal component analysis; Multi-dimensional process

1. Introduction

Alzheimer's Disease (AD) is one of the most prevalent neurodegenerative diseases worldwide. As the disease progresses, it eventually leads to severe cognitive impairment such as language and memory loss. It has no effective disease-modifying treatment yet. Due to the lack of cure, it is of paramount interest to be able to detect and predict the onset of AD early. This is especially relevant for people diagnosed with mild cognitive impairment (MCI), which is defined as a transitional stage between the cognitive normal (CN) state and the dementia state which is of a much more severe degree. Massive amounts of data are being collected with the goal of identifying significant biomarkers or covariates that may be related to the progression of AD. In the Alzheimer's Disease Neuroimaging Initiative (ADNI), the focus was on the collection of longitudinal assessments of magnetic resonance imaging (MRI), where in each follow-up visit, the patient's MRI scan was collected and the patient would be diagnosed as one of the three categories: cognitive normal (CN), MCI or AD. A series of longitudinal MRI scans as well as demographic and prognostic covariates were collected and might serve as significant predictors for predicting the neurodegenerative pathology due to AD.

Various existing methods have been proposed to use the longitudinal biomarkers for the dynamic prediction of the time-to-event outcome of AD. Welsh, Lin, and Carroll (2002) adopted nonparametric smoothing methods to obtain denoised smoothed values of the biomarkers, which are then used for prediction. More recently, functional data analysis and in particular the functional principal component analysis (FPCA) has been often used for analyzing the trajectories and surface data. FPCA seeks to decompose the underlying process into a linear combination of functional principal components (FPCs). The earlier works on FPCA focus on one-dimensional densely and regularly observed curves, see, for example, Ramsay and Dalzell (1991), and Silverman (1996).

More recently, with the availability of multidimensional or multivariate functional data, various methods have been developed to extract the FPCs. The majority of these approaches focus on estimating and extending the covariance function of the multivariate random process. Lin, Li, and Luo (2021) developed the multivariate FPCA by extending the classical Karhunen-Loeve expansion to multivariate setting via normalization on the covariance operator. Chen and Jiang (2017) extended the FPCA to analyze functional data on a general multi-dimensional domain, and used local linear smoothing to achieve large-scale efficient estimation of the multi-dimensional covariance function. Lin, Wang, and Cao (2016) introduced a penalty-based method to derive FPCs that exhibit nonzero values exclusively within intervals where their significance is observed. Sang, Wang, and Cao (2017) developed a parametric FPCA method

CONTACT Jiguo Cao *jiguo_cao@sfu.ca* Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, Canada.
Supplementary materials for this article are available online. Please go to *www.tandfonline.com/r/JCGS*.

to estimate the top FPCs using simple parametric functions in order to enhance their interpretability for users. Happ and Greven (2018) established the theoretical framework of the multivariate Karhunen-Loeve theorem and allowed the dimensions of functional data to be definable over different domains. Nie et al. (2018) introduced a supervised FPCA approach that accounts for the correlation between the functional predictor and the response variable. Lin, Li, and Luo (2021) developed a unique extension of Karhunen-Loeve expansion for multi-dimensional functional electroencephalography data to capture the longitudinal trend of multi-level process. Nie and Cao (2020) proposed a sparse FPCA method in a new regression framework. Lin, Li, and Luo (2021) further developed a novel covariance-based multivariate sparse FPCA that can capture cross-correlations between functions. Shi et al. (2021) proposed the informatively missing FPCA method for cases where the longitudinal trajectories are subject to informative missingness. Nie, Wang, and Cao (2022) introduced a sparse orthonormal approximation (SOAP) method for estimating FPCs, which avoids estimating the covariance function. Shi et al. (2022) developed a two-dimensional FPCA method for extracting FPCs and achieving dimensionality reduction for images.

For dynamic prediction of AD, various methods have been proposed in recent years. For example, Kong et al. (2018) proposed a functional Cox proportional hazards model that uses functional principal component analysis (FPCA) to extract FPC scores from brain imaging data at the baseline as features in the model. Li and Luo (2019a) proposed a dynamic prediction framework using multiple longitudinal biomarkers under a proportional hazards model with the multivariate FPC scores developed by Happ and Greven (2018), and Li and Luo (2019b, 2019c) explored using the multivariate FPC scores under a joint model framework. Jiang, Xie, and Colditz (2020) and Lin, Li, and Luo (2021) proposed to integrate random survival forest with multivariate FPCA to predict the progression of AD. Zou et al. (2021, 2023) developed a Bayesian extension of the functional mixed model.

All of the aforementioned methods focus on the modeling time-to-event outcome based on either multivariate longitudinal biomarker values or brain imaging data at the baseline. However, in the practical setting, at each visit, the subject's brain MRI scan image is collected, represented by a multi-dimensional data matrix. It is of great interest to predict the progression of AD directly from such a series of longitudinally observed images. For illustration purpose, Figure 1 illustrates the longitudinal brain scans of two subjects, which are collected at month 0, 6, 12, 18, and 24; subject 1 was diagnosed with AD at around month 24, with a missed collection at month 18, and subject 2 was still in the MCI state at his/her latest visit in month 18. It is also worth noting that the actual visit time is not precisely spaced on a 6/12-month grid and may differ by the nominal time by several months.

Our objective is to develop an effective method for incorporating predictive information from the longitudinal image series, which, to the best of our knowledge, has never been explored in the literature. Compared with the models using multivariate biomarkers (Li and Luo 2019a; Jiang, Xie, and Colditz 2020), the features extracted from longitudinal images may provide additional information for the accurate prediction of AD. Compared with the models based on brain imaging data at the baseline (Kong et al. 2018), our model extended the capacity for handling more than one image observed sparsely across multiple time points.

However, several challenges may arise in the problem of predicting AD from the longitudinal images. First of all, the longitudinal images are represented by multi-dimensional data matrices. On top of the row/column dimension of an image matrix, for each subject the data series is characterized by an additional dimension, the time of visit, which further adds to the complexity of the data. Conventionally, it might be feasible (albeit with heavy computational cost) to use the multidimensional FPCA for handling such functional data, so long as they are observed from a dense sampling scheme, that is, where the time of follow-up visits and diagnosis are exactly spaced on an equidistant grid. This typically requires eigendecomposition of a six-dimensional covariance function, a task that is computationally heavy. However, in the ADNI studies, the sampling scheme is sparse and irregular, as the time of visits may differ by a few months from subject to subject. This renders such conventional approaches futile which rely upon the assumption of a dense sampling scheme.

To tackle these challenges, in this article, we develop a novel approach for dynamic prediction that is capable of accommodating multidimensional longitudinal images under the Cox regression framework. The proposed strategy consists of two main steps. First, we consider the longitudinal image process as a stochastic function over a multi-dimensional support, and use a new regression-based FPCA to obtain the FPCs from all the subjects' longitudinal images that characterize their major mode of variation. Second, we use the FPC scores extracted from these subjects' longitudinal images as predictors in a Cox proportional hazards model to capture the relationship between the time-to-event outcome and the longitudinal images.

Our approach has several key contributions. First, to the best of our knowledge, our method is the first attempt handling sparse and multi-dimensional functional data. Compared with the majority of existing methods that focus on multivariate longitudinal biomarkers, our proposed method greatly expands the methodological horizon of the dynamic prediction model with functional predictors. Compared with the multivariate FPCA (Happ and Greven 2018; Li and Luo 2019a), our method is tackling the problem from a different perspective: rather than having multivariate trajectories, in our problem, the domain/support of the stochastic process is multi-dimensional, which is rarely studied in the existing literature. Second, our approach to FPCA circumvents the computational necessity of estimating a high-dimensional covariance function, as it directly estimates the optimal empirical FPCs and their corresponding basis coefficients via tensor product B-spline under a regression framework. In comparison, the conventional approach for multi-dimensional functional data is contingent upon a dense sampling scheme and eigen-decomposition on the sample covariance function, which requires heavy computation. Our method provides a useful alternative to the conventional approaches. Third, our method is capable of handling sparse and irregular multi-dimensional data, as it relies on a flexible regression framework. While there are existing methodologies

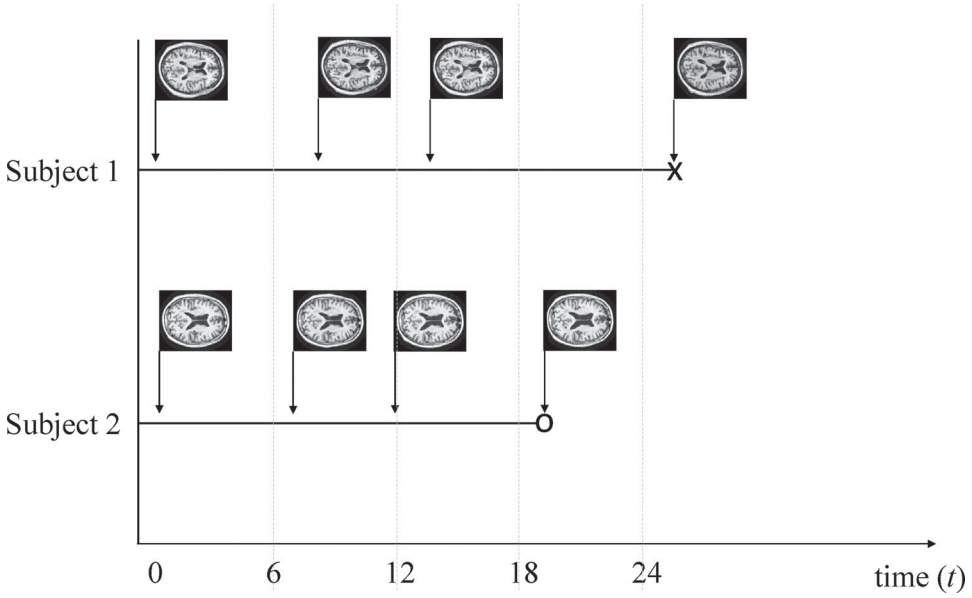

Figure 1. Illustration of the longitudinal images of two subjects during their follow-ups in the ADNI study, where the AD event is denoted by a cross and the censoring event denoted by a circle. Note that the visit times of the two subjects are irregular and do not coincide.

that focus on either the topic of sparsity (Yao, Müller, and Wang 2005) or the topic of multi-dimensionality, a method that simultaneously accommodates both sparsity and multi-dimensionality is rarely studied in the literature of functional data analysis and our method aims to fill such a research gap. The computing codes for the implementation of the method are available at *https://github.com/haoluns/dynamicMFPCA*.

The rest of the article is organized as follows. In Section 2, we lay out the notation and describe the details of our model. In particular, we present the proposed methodology of our regression-based multidimensional FPCA for longitudinal images and construct the dynamic prediction model based on the FPC scores. In Section 3, we conduct simulation studies to assess the finite sample performance of the survival prediction. In Section 4, we apply the proposed method to the ADNI datasets and present several interesting findings. Finally, Section 5 concludes the article with a discussion.

2. Multi-Dimensional Functional Principal Component Analysis

We first set up the functional framework for modeling multiple images. Let u and v denote the two-dimensional positional offsets (horizontal and vertical) of an image, and let t denote the time when the image is observed. Let $X_i(u, v, t)$ denote underlying three-dimensional function for the ith subject, which equals the grayscale value of a pixel at (u, v) in an image sampled at time t.

Let t^* denote the maximum time period within which the longitudinal images are used for modeling. Over the follow-up of subject i, we assume there are $n_{i,t}$ time points at which an image is observed. Each image is of size $n_{i,u}$-by-$n_{i,v}$, that is, $n_{i,u}$ pixels along the horizontal axis and $n_{i,v}$ pixels along the vertical axis. For a pixel on an image, we index it as (j_u, j_v), which correspond to the j_u/j_vth pixel along the horizontal/vertical axis. Furthermore, we index the time points by j_t, where $j_t = 1, \ldots, n_{i,t}$. For subject i, there are a total of $n_{i,t}$ longitudinally observed $n_{i,u}$-by-$n_{i,v}$ images, with the total number of pixels in all the observed images being $n_i = n_{i,t}n_{i,u}n_{i,v}$. In the application of our method on the ADNI study, longitudinal image are still available after the subject's event outcome.

Let y_{i,j_u,j_v,j_t} denote the observed grayscale value of a pixel on $X_i(\cdot)$. We assume that y_{i,j_u,j_v,j_t} is sampled as,

$$y_{i,j_u,j_v,j_t} = X_i(u_{i,j_u,j_t}, v_{i,j_v,j_t}, t_{i,j_t}) + \epsilon_{i,j_u,j_v,j_t},$$

where u_{i,j_u,j_t} and v_{i,j_v,j_t} are the positional offsets of the j_u and j_vth horizontal/vertical pixel of an image observed at the j_tth time point t_{i,j_t}, with $j_u = 1, \ldots, n_{i,u}$, $j_v = 1, \ldots, n_{i,v}$, and $j_t = 1, \ldots, n_{i,t}$. Without loss of generality, we assume that the height and width of an image are equal to I, that is, $u_{i,j_u,j_t}, v_{i,j_v,j_t} \in [0, I]$, and $t_{i,j_t} \in [0, t^*]$. The error term ϵ_{i,j_u,j_v,j_t} is assumed to have mean zero and variance σ^2.

To simplify notation, we reduce the three-dimensional indices j_u, j_v, and j_t into a single index, defined simply as $j = 1, \ldots, n_i$ for every pixel of all the observed images of subject i. The equation above can then be simplified as,

$$y_{ij} = X_i(u_{ij}, v_{ij}, t_{ij}) + \epsilon_{ij},$$

where $i = 1, \ldots, n$ and $j = 1, \ldots, n_i$, with $n_i = n_{i,t}n_{i,u}n_{i,v}$ being the total number of pixels of all the longitudinally observed images.

The three-dimensional random function $X_i(u, v, t)$ can be expressed in terms of the Mercer expansion of a series of orthonormal basis functions, that is,

$$X_i(u, v, t) = \mu(u, v, t) + \sum_{m=1}^{\infty} \xi_{im}\psi_m(u, v, t),$$

where $\xi_{im} = \int_0^{t^*}\int_0^I\int_0^I \{X_i(u, v, t) - \mu(u, v, t)\}\psi_m(u, v, t)dudvdt$ is the mth functional principal component (FPC) score for subject i, and the function $\psi_m(\cdot)$'s are the functional principal components that have a norm of 1 and are orthogonal to each other.

Each longitudinal image process $X_i(u,v,t)$ can be characterized by the sequence of FPC scores ξ_{im}, where $m = 1,\ldots,\infty$. In practice, we retain the first M scores where M can be determined by the total variance explained in a similar way as a multivariate principal component analysis. Dimensionality reduction is achieved as the infinite-dimensional function $X_i(u,v,t)$ is summarized by the information of the first M FPC scores.

To extract the FPC scores, we propose a multi-dimensional functional principal component analysis, which estimates the first M functional principal components $\psi_m(u,v,t)$ as well as the scores ξ_{im} as the minimizer of the objective function,

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{n_i}\sum_{j=1}^{n_i}\left\{y_{ij}-\mu(u_{ij},v_{ij},t_{ij})-\sum_{m=1}^{M}\xi_{im}\psi_m(u_{ij},v_{ij},t_{ij})\right\}^2.$$

These M FPCs and FPC scores are obtained sequentially, that is, the mth FPC is computed based on estimates of the first $m-1$ FPCs.

In Section 2.1, we will focus on estimating the first functional principal component. In Section 2.2, we will then introduce how to estimate the subsequent functional principal components

2.1. Estimating the First Functional Principal Component

The first FPC $\psi_1(u,v,t)$ is obtained by minimizing

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{n_i}\sum_{j=1}^{n_i}\left\{y_{ij}-\widehat{\mu}(u_{ij},v_{ij},t_{ij})-\xi_{i1}\psi_1(u_{ij},v_{ij},t_{ij})\right\}^2, \quad (1)$$

subject to $\|\psi_1\|^2 = 1$, where $\widehat{\mu}(u_{ij},v_{ij},t_{ij})$ is the estimated mean function by pooling all the data together and performing a spline regression (details can be found in the supplementary material).

We use the tensor product B-spline as a means of basis expansion to model the three-dimensional function $\psi_1(u,v,t)$,

$$\psi_1(u,v,t)=\sum_{i=1}^{s_u}\sum_{j=1}^{s_v}\sum_{l=1}^{s_t}\beta_{1,ijl}b_i^{(1)}(s)b_j^{(2)}(v)b_l^{(3)}(t).$$

Let $\mathbf{b}_u = (b_1^{(1)},\ldots,b_{s_u}^{(1)})^\top$, $\mathbf{b}_v = (b_1^{(2)},\ldots,b_{s_v}^{(2)})^\top$ and $\mathbf{b}_t = (b_1^{(2)},\ldots,b_{s_t}^{(2)})^\top$ denote the spline basis functions for the dimension u, v, and t, respectively, where s_u, s_v and s_t are the numbers of bases in each dimension. We can further reorganize and simplify the summation above as

$$\psi_1(u,v,t)=\boldsymbol{\beta}_1^\top\mathbf{b}(u,v,t),$$

where

$$\mathbf{b}(u,v,t)=\mathrm{vec}(\mathbf{b}_u(u)\otimes\mathbf{b}_v(v)\otimes\mathbf{b}_t(t)),$$

and $\boldsymbol{\beta}_1$ is the vector of corresponding tensor product spline coefficients $\beta_{1,ijl}$. Here, $\otimes$ denotes Kronecker product and the $\mathrm{vec}(\cdot)$ is the vectorization operator, which performs stacking of all the columns of a matrix.

Regarding the choice of s_u, s_v, and s_t, the optimal number of splines in each dimension would ideally be chosen over cross-validation and grid-search. To be specific, we may loop through all the candidate combinations of s_u, s_v, and s_t over a grid, compute the FPCs on the training portions in the cross-validation folds, calculate the average of the objective functions in (1) on the holdout testing samples based on the fitted FPC and scores, and finally choose the parameter combination that has the smallest objective function. For empirical convenience, another approach with less computational burden would be to apply the univariate cross-validation procedure for choosing the number of splines. For example, to choose the optimal value of s_t, the number of splines in the time dimension, we may rearrange the data into univariate curves that corresponds to the longitudinal trend in the values of each pixel over time, and apply univariate cross-validation procedure on these curves.

To estimate $\boldsymbol{\xi}_1 = (\xi_{11},\ldots,\xi_{n1})^\top$ and the coefficient vector $\boldsymbol{\beta}_1$, we propose an iterative optimization procedure, which works as follows: given the estimate of $\boldsymbol{\xi}_1$, update the estimate of $\boldsymbol{\beta}_1$; given the new estimate of $\boldsymbol{\beta}_1$, update the estimate of $\boldsymbol{\xi}_1$; iterate between these two steps until convergence. To be more specific, given the current estimate $\boldsymbol{\beta}_1^{(\ell)}$, we obtain the estimate of the FPC score $\xi_{i1}^{(\ell+1)}$ by minimizing (1) via least-square as

$$\xi_{i1}^{(\ell+1)}=(\boldsymbol{\psi}_{i1}^\top\boldsymbol{\psi}_{i1})^{-1}\boldsymbol{\psi}_{i1}^\top\mathbf{y}_i, \qquad i=1,\ldots,n,$$

where $\boldsymbol{\psi}_{i1} = (\psi_1^{(\ell)}(u_{i1},v_{i1},t_{i1}),\ldots,\psi_1^{(\ell)}(u_{in_i},v_{in_i},t_{in_i}))^\top$, $\mathbf{y}_i = (y_{i1}^*,\ldots,y_{in_i}^*)^\top$, and $\psi_1^{(\ell)}(u,v,t) = \boldsymbol{\beta}_1^{(\ell)\top}\mathbf{b}(u,v,t)$ is the estimated functional principal component.

Next, given the estimate $\boldsymbol{\xi}_1^{(\ell+1)} = (\xi_{11}^{(\ell+1)},\ldots,\xi_{n1}^{(\ell+1)})^\top$, we obtain the estimate of $\boldsymbol{\beta}_1$ by minimizing (1) via least-square, subject to $\|\psi_1\|^2 = 1$,

$$\boldsymbol{\beta}_1^{(\ell+1)}=\frac{\widetilde{\boldsymbol{\beta}}_1^{(\ell+1)}}{\left\|\widetilde{\psi}_1^{(\ell+1)}\right\|},$$

where $\widetilde{\boldsymbol{\beta}}_1^{(\ell+1)}$ is the unconstrained estimate of $\boldsymbol{\beta}_1$.

We repeat the steps above until convergence. In terms of the convergence criterion, the algorithm is deemed to have reached convergence if the elementwise maximum of $|\boldsymbol{\beta}_1^{(\ell+1)} - \boldsymbol{\beta}_1^{(\ell)}|$ is smaller than a prespecified threshold δ. Typically, δ takes a very small value such as 0.00001.

2.2. Estimating Subsequent Functional Principal Components

We obtain subsequent functional principal components in a sequential manner. The Jth functional principal component is computed based on the estimated FPC spline coefficient $\widehat{\boldsymbol{\beta}}_m$ from the first $J-1$ steps. Given $\widehat{\boldsymbol{\beta}}_m$, where $m = 1,\ldots,J-1$, the Jth functional principal component ψ_J is estimated by minimizing

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{n_i}\sum_{j=1}^{n_i}\Bigg\{y_{ij}-\widehat{\mu}(u_{ij},v_{ij},t_{ij}) -\sum_{m=1}^{J-1}\xi_{im}\widehat{\psi}_m(u_{ij},v_{ij},t_{ij})-\xi_{iJ}\psi_J(u_{ij},v_{ij},t_{ij})\Bigg\}^2, \quad (2)$$

subject to $\|\psi_J\|^2 = 1$ and $\langle\widehat{\psi}_m,\psi_J\rangle = 1$ if $m = J$, and 0 otherwise.

Let $\boldsymbol{\xi}_i = (\xi_{i1},\ldots,\xi_{iJ})^\top$ denote the FPC scores for all the FPCs in the ith subject and $\boldsymbol{\beta}_J$ denote the vector of spline coefficients

for the Jth FPC. We propose an optimization procedure that iteratively alternates between estimating the FPC scores and the FPC spline coefficients, which are detailed as follows.

Given the current estimate $\boldsymbol{\beta}_J^{(\ell)}$, update the estimate of $\boldsymbol{\xi}_i$ by minimizing (2) via least-square as

$$\boldsymbol{\xi}_i^{(\ell+1)} = (\boldsymbol{\psi}_i^\top \boldsymbol{\psi}_i)^{-1} \boldsymbol{\psi}_i^\top \boldsymbol{y}_i, \qquad i = 1, \ldots, n, \tag{3}$$

where $\boldsymbol{\psi}_i = (\boldsymbol{\psi}_{i1}, \ldots, \boldsymbol{\psi}_{iJ})$, $\boldsymbol{\psi}_{im} = (\psi_m^{(\ell)}(u_{i1}, v_{i1}, t_{i1}), \ldots, \psi_m^{(\ell)}(u_{in_i}, v_{in_i}, t_{in_i}))^\top$, $\boldsymbol{y}_i = (y_{i1}^*, \ldots, y_{in_i}^*)^\top$, and

$$\psi_m^{(\ell)}(u, v, t) = \widehat{\boldsymbol{\beta}}_m^\top \mathbf{b}(u, v, t), \quad m = 1, \ldots, J-1,$$
$$\psi_J^{(\ell)}(u, v, t) = \boldsymbol{\beta}_J^{(\ell)\top} \mathbf{b}(u, v, t).$$

Next, given the current estimates $\boldsymbol{\xi}_i^{(\ell+1)}$, update the estimate of $\boldsymbol{\beta}_J$ as the minimizer of (2), subject to $\|\psi_J\|^2 = 1$, and $\langle \widehat{\psi}_m, \psi_J \rangle = 1$ if $m = J$, and 0 otherwise. This can be casted into a least-square problem with constraints (Lawson and Hanson 1974). Let $w_{ij} = y_{ij}^* - \sum_{m=1}^{J-1} \xi_{im}^{(\ell+1)} \widehat{\boldsymbol{\beta}}_m^\top \mathbf{b}(u_{ij}, v_{ij}, t_{ij})$. The estimate of $\boldsymbol{\beta}_J$ is updated as

$$\boldsymbol{\beta}_J^{(\ell+1)} = \arg\min_{\boldsymbol{\beta}_J} \sum_{i=1}^{n} \frac{1}{n_i} \sum_{j=1}^{n_i} \left\{ w_{ij} - \xi_{ij}^{(\ell+1)} \boldsymbol{\beta}_J^\top \mathbf{b}(u_{ij}, v_{ij}, t_{ij}) \right\}^2,$$

subject to

$$\int_0^{t^*} \int_0^I \int_0^I \left\{ \widehat{\boldsymbol{\beta}}_m^\top \mathbf{b}(u, v, t) \right\} \left\{ \boldsymbol{\beta}_J^\top \mathbf{b}(u, v, t) \right\} du\, dv\, dt = 0, \qquad m = 1, \ldots, J-1,$$

and

$$\int_0^{t^*} \int_0^I \int_0^I \left\{ \boldsymbol{\beta}_J^\top \mathbf{b}(u, v, t) \right\}^2 du\, dv\, dt = 1.$$

After obtaining the FPC spline coefficient estimate $\widehat{\boldsymbol{\beta}}_m$ for $m = 1, \ldots, J$, we can construct the estimate for $\widehat{\psi}_m(u, v, t)$. As both the FPC $\widehat{\psi}_m(u, v, t)$ and the FPC scores ξ_{im} are the minimizer of the objective function,

$$\frac{1}{n} \sum_{i=1}^{n} \frac{1}{n_i} \sum_{j=1}^{n_i} \left\{ y_{ij} - \mu(u_{ij}, v_{ij}, t_{ij}) - \sum_{m=1}^{J} \xi_{im} \psi_m(u_{ij}, v_{ij}, t_{ij}) \right\}^2,$$

we can estimate the FPC scores $\boldsymbol{\xi}_i$ via least-square as

$$\widehat{\boldsymbol{\xi}}_i = (\widehat{\boldsymbol{\psi}}_i^\top \widehat{\boldsymbol{\psi}}_i)^{-1} \widehat{\boldsymbol{\psi}}_i^\top \boldsymbol{y}_i, \qquad i = 1, \ldots, n,$$

where $\widehat{\boldsymbol{\psi}}_i = (\widehat{\boldsymbol{\psi}}_{i1}, \ldots, \widehat{\boldsymbol{\psi}}_{iJ})$, $\widehat{\boldsymbol{\psi}}_{im} = (\widehat{\psi}_m(u_{i1}, v_{i1}, t_{i1}), \ldots, \widehat{\psi}_m(u_{in_i}, v_{in_i}, t_{in_i}))^\top$, $\boldsymbol{y}_i = (y_{i1}^*, \ldots, y_{in_i}^*)^\top$, and $\widehat{\psi}_m(u, v, t) = \widehat{\boldsymbol{\beta}}_m^\top \mathbf{b}(u, v, t)$, $m = 1, \ldots, J$, are the estimated FPCs.

We repeat the steps above until convergence. In terms of the convergence criterion, the algorithm is deemed to have reached convergence if the elementwise maximum of $|\boldsymbol{\beta}_J^{(\ell+1)} - \boldsymbol{\beta}_J^{(\ell)}|$ is smaller than a prespecified threshold δ.

3. Dynamic Survival Prediction via FPC Score

We build the Cox proportional hazards model using the training dataset. For the ith subject, a Cox proportional hazard model can be formulated based on the FPC scores $\widehat{\boldsymbol{\xi}}_i$ and the time-independent covariates $\mathbf{Z}_i$, which specifies the hazard function as

$$h_i(t) = h_0(t) \exp\{\widehat{\boldsymbol{\xi}}_i^\top \boldsymbol{\alpha} + \mathbf{Z}_i^\top \boldsymbol{\gamma}\},$$

where $h_0(t)$ is the baseline hazard function, $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ are the vector of regression coefficients for the FPC scores and the time-independent covariates, respectively.

For a new subject $(n+1)$ who is event free and has a series of longitudinal images up to time t^*, let the series of the pixel values of subject $(n+1)$ be denoted as $\{(y_j, u_j, v_j, t_j); j = 1, \ldots, n^*)\}$, where n^* is the total number of pixels of all the images of subject $(n+1)$. The FPC scores $\{\xi_{(n+1)m}; m = 1, \ldots, J\}$ can be estimated as the minimizer of the following objective function,

$$\sum_{j=1}^{n^*} \left\{ y_j - \widehat{\mu}(u_j, v_j, t_j) - \sum_{m=1}^{J} \xi_{(n+1)m} \widehat{\psi}_m(u_j, v_j, t_j) \right\}^2, \tag{4}$$

where the $\widehat{\mu}(\cdot)$ and $\widehat{\psi}_m(\cdot)$ are obtained from all the n subjects in the cohort, and thus the objective function above is the same as the squared loss in least-squared estimation, which can be minimized easily via least-squared regression fit. Hence, the FPC scores $\{\xi_{(n+1)m}; m = 1, \ldots, J\}$ can be estimated by regressing all of the observed mean-centered pixels of the $(n+1)$th subject on the FPCs $\widehat{\psi}_m(u, v, t)$.

It is worth noting that the FPC scores $\{\xi_{(n+1)m}; m = 1, \ldots, J\}$ can be updated as the time horizon t^* increases up to the maximum follow-up time of the longitudinal images in the cohort, and is thus dependent upon t^*. We denote the FPC score vector $(\xi_{(n+1)1}, \ldots, \xi_{(n+1)J})^\top$ under time horizon t^* as $\widehat{\boldsymbol{\xi}}_{n+1}(t^*)$. Under the dynamic survival prediction framework, we predict the survival probability at $t^* + \Delta t$ as

$$\widehat{S}_{n+1}\left(t^* + \Delta t \mid t^*\right) = \frac{\widehat{S}_{n+1}\left(t^* + \Delta t \mid \mathbf{Z}_{n+1}, \widehat{\boldsymbol{\xi}}_{n+1}(t^*)\right)}{\widehat{S}_{n+1}\left(t^* \mid \mathbf{Z}_{n+1}, \widehat{\boldsymbol{\xi}}_{n+1}(t^*)\right)} = \left\{ \frac{\widehat{S}_0\left(t^* + \Delta t\right)}{\widehat{S}_0(t^*)} \right\}^{\exp\left\{\widehat{\boldsymbol{\xi}}_{n+1}(t^*)^\top \widehat{\boldsymbol{\alpha}} + \mathbf{Z}_{n+1}^\top \widehat{\boldsymbol{\gamma}}\right\}}.$$

4. Data Application

4.1. ADNI Brain MRI Scans

To illustrate the usefulness of the dynamic prediction model with multi-dimensional FPCA, we apply the method on the publicly available Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset. In the ADNI study, a subject's MRI scans are collected longitudinally every 6 months in the first 24 months and subsequently every 12 months until the end of their follow-up period, which usually spans over 10 years, with occasional missed follow-ups. At each follow-up visit, a subject is diagnosed as one of three types: cognitive normal, mild cognitive impairment (MCI), and Alzheimer's disease (AD). We treat the time

Table 1. The regression results from Cox proportional hazards model using FPC scores, age, gender and APOE4 as covariates and time to conversion to AD as the time-to-event outcome in the ADNI brain scan dataset.

Variable	Estimate	SE	*p*-value
FPC1	−0.3586	0.0636	<**0.0001**
FPC2	−0.2563	0.0642	<**0.0001**
FPC3	0.0761	0.0638	0.2327
FPC4	−0.0568	0.0678	0.4023
FPC5	0.0893	0.0619	0.1489
Age	−0.0186	0.0108	0.0863
Gender	0.2115	0.1414	0.1348
APOE4	1.0337	0.1294	<**0.0001**

Table 1 shows the results of the regression coefficients, standard errors and p-values under the Cox regression model, where the p-values below 0.001 are highlighted in bold.

until the subject is diagnosed as AD for the first time as the time-to-event outcome. We focus on the 674 subjects who are free of AD at baseline, out of which 286 have eventually progressed into AD. The average follow-up period is 47.9 (sd = 28.8) months and the average number of follow-up visits is 5.6 (sd = 2.2).

To prepare the longitudinal images, we adopt the following image processing procedures for the T1-weighted structural brain MRI data collected at each follow-up visit. First, we transform the volumetric images of all the baseline scans into the same stereotaxis space by affinely registering them to the same template using the image processing pipelines based on FreeSurfer version 5.3.0 (Ma et al. 2018; Popuri et al. 2020). Specifically, we apply automatic rigid registration to the original 3D MRI images which align all the images to the same stereotaxis reference space, as well as conform the original data to the same isotropic resolution. Such process ensures that the extracted 2D slice from each subjects represent same anatomical locations and are measured in the same scale. We use the middle axial slice in the reference space for all the registered images for the sake of its representative anatomical location which contains most of the AD-related functional brain regions from both hemispheres, for example, ventricle, hippocampus, para-hippocampal region, amygdala, and neocortex (including fronto- and temporal- lobe). We have experimented with a number of other candidate axial slices and noted similar performances. Finally, images are represented as grayscale matrices and are de-centered by subtracting the mean value at each pixel.

4.2. Multi-Dimensional FPCA

We first apply the proposed multi-dimensional FPCA on all the longitudinal images, and use the first 5 FPC scores as predictors in the Cox regression model. Table 1 shows the results of the regression coefficients, standard errors and p-values under the Cox regression model. To compare the effect size of the coefficient in the table, we standardize the FPC scores to be centered around 0 with a standard deviation of 1. We see that the first two FPC scores have significant association with the risk of conversion to AD, with p-values being less than 0.0001. The other FPC scores appear to be nonsignificant and we exclude them from our further analysis. We additionally include age, gender and the indicator of the presence of allele of the apolipoprotein E4 gene (APOE4) as covariates. The APOE4 gene is known to associated with increased risk of developing late-onset Alzheimer's disease. We note that the variable APOE4 is also significant with a p-value less than 0.0001, while the variables gender and age are not.

Figure 2 shows the plots of the first 2 FPCs with a snapshot taken at time points 0, 6, 12, 18, and 24. As the FPCs are longitudinal images, we present the plot of FPCs at baseline, as well as the change from the baseline FPCs (denoted as $\Delta\text{FPC}_1(t)$ and $\Delta\text{FPC}_2(t)$). The first FPC has no regional variation in the brain and is of a uniform color, which represents an overall difference in the grayscale density of the images. Moreover, there is almost no change in the first FPC since the baseline as the subsequent plots of $\Delta\text{FPC}_1(t)$ are almost transparent. The second FPC exhibits a regional contrast in the middle part (blue) and the two peripheral regions (red), which corresponds to roughly where the hippocampus is located. There are some evident changes in subsequent months; for example, a positive change at months 6 and 24, and a negative change at months 12 and 18.

4.3. Model Comparison

Moreover, for comparison, we consider a functional Cox proportional hazards model with a univariate biomarker as the functional predictor (Kong et al. 2018). The model uses the average grayscale density in the hippocampus region as the longitudinal biomarker, and conducts FPCA on the univariate biomarker trajectories before using the obtained FPC scores as covariates in the Cox proportional hazards model. In addition to the FPC scores, the model incorporates the baseline information of subjects, such as age, gender and APOE4.

We compare the predictive performance of our model with that of the univariate functional Cox proportional hazards model, in terms of the AUC and the expected Brier score (BS), at various time points in the whole study period of ADNI. For the sake of robustness in the comparison, we conduct a 5-fold cross-validation, where we estimate the first three FPCs and the Cox model from training set, and apply the model in the testing set by first computing the FPC scores for the new subjects and estimating their survival probabilities using the Cox model from the training stage. For landmark time points t^* =6, 12, 18, and 24, we predict the conditional survival probability at $t^* + \Delta t$, for time window $\Delta t = 24$. The AUC, denoted as $\text{AUC}(t^*, t^* + 24)$, and expected Brier scores, denoted as $\text{BS}(t^*, t^* + 24)$, are computed based on the predicted probabilities of the subjects in the testing set.

Figure 3 shows the AUC and Brier score evaluated dynamically for each time window $(t^*, t^* + 24)$ given the data observed up to $t^* = 6, 12, 18$, and 24. It is evident that the proposed model that uses the information in the longitudinal multi-dimensional images achieves better predictive performance than the univariate functional Cox model over all time points. This indicates that the information extracted from the multi-dimensional FPCA leads to improved model discrimination and calibration.

4.4. Illustration of Dynamic Prediction

Finally, we demonstrate the model's capability of dynamic prediction for a new subject. As an illustration, we present the dynamic updating of a subject's FPC scores and prediction of the event-free probability. After being enrolled in the study, the subject's follow-up visits occur at around months 6, 12,

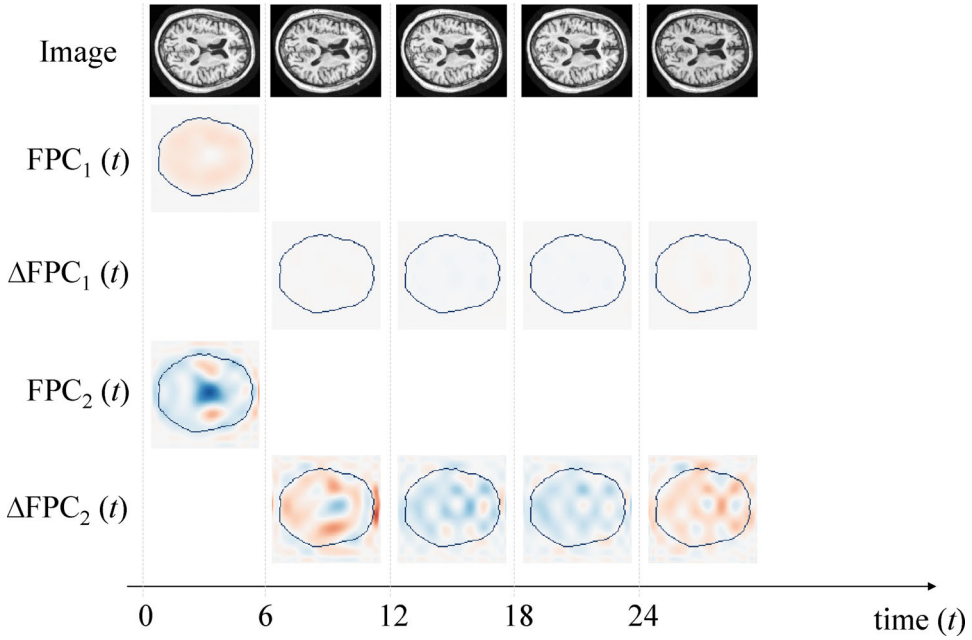

Figure 2. The plots of the first 2 FPCs with a snapshot taken at time points 0, 6, 12, 18, and 24 in the analysis of ADNI brain MRI scans. As the FPCs are longitudinal images, we present the plot of FPCs at baseline, as well as the change from the baseline FPCs, which are denoted as $\Delta\text{FPC}_1(t)$ and $\Delta\text{FPC}_2(t)$. The magnitudes of positive and negative values are shaded by a gradual color scheme in red and blue, respectively.

18, and 24, and his/her longitudinal brain scans are collected. At each landmark time point, we re-estimate the FPC scores. Specifically, at time t^*, we updated the subject's first and second FPC scores by regressing all the observed mean-centered pixels of the series of images observed up to time t^*, on the FPCs $\widehat{\psi}_1(u,v,t)$ and $\widehat{\psi}_2(u,v,t)$ obtained from the trained model. With new information being gleaned, the subject's FPC scores as well as the predicted hazard ratio to the baseline are updated.

Figures 4 and 5 show the dynamic updating of the FPC scores and the resultant contributive factor from the FPC, that is, the product between the FPC and the subject's FPC scores. We observe that the first FPC score remains relatively stable in the range between -600 and -400. On the other hand, the second FPC score decreases from 565.2 at month 6 to -389.5 at month 24. The change in the second FPC score is evidently demonstrated by the contrast in color between the first and the last row in Figure 5. As the second FPC score is negatively associated with the hazard in the Cox model, the resultant predicted hazard ratio to the baseline increases over time as a result of the changes in the FPC scores, as shown in Figure S.1 in the supplementary material. Moreover, the Cox model allows us to plot the predicted event-free probabilities given the survival up to time t^*, which are shown in Figure S.2 in the supplementary material. Compare to $t^* = 6$, the predicted event-free probability function is slightly more tilted to the downside, indicating a deterioration in the disease status.

5. Simulation Studies

5.1. Simulation: FPCA

We conduct simulation studies to evaluate the empirical performance of the proposed FPCA method in terms of recovering the underlying modes of variation from longitudinal series of images. The goal is to establish the accuracy of extracting and recovering the FPCs based on the proposed alternating regression approach.

We first create the true underlying multi-dimensional stochastic process, assuming that the FPCs are known. The mean function estimation in our method is expected to perform well as the estimator is consistent, and thus our focus lies on the estimation of FPCs. We generate the underlying true multi-dimensional surfaces $X_i(u,v,t)$ as the sum product of two sets of FPCs and scores, with a mean function that satisfies $\mu(u,v,t) = 0$,

$$X_i(u,v,t) = \xi_{i1}\psi_1(u,v,t) + \xi_{i2}\psi_2(u,v,t), \qquad i = 1,\ldots,n.$$

We use the first two FPCs from the analysis of the ADNI data as the underlying true FPCs $\psi_1(u,v,t)$ and $\psi_2(u,v,t)$ and they naturally satisfy $\|\psi_j\|^2 = 1$, and $\langle\psi_j,\psi_k\rangle = 1$ if $j = k$, and 0 otherwise. We assume that the FPC scores follow a Gaussian distribution. The FPC scores ξ_{i1} and ξ_{i2} are independently drawn from normal distributions with mean 0 and decreasing standard deviations of 4000 and 1000, respectively: $\xi_{i1} \sim \text{N}(0, 4000^2)$, and $\xi_{i2} \sim \text{N}(0, 1000^2)$. Note that at any (u,v,t), the actual magnitudes of $\psi_1(u,v,t)$ and $\psi_2(u,v,t)$ are in the range of 10^{-5}, and thus we use large values of standard deviations. The decreasing standard deviations is designed to reflect the fact that the higher ranked FPC very often explains a larger proportion of variation.

Having created the true underlying surfaces, the observed data points are simulated from the multi-dimensional surfaces in addition to a random term of error. We consider both dense and sparse sampling schemes. Under the dense sampling scheme, all the subjects share the same set of time points when the images are observed; and under sparse sampling, the time points of the images are oftentimes not the same across subject. The observed images are drawn over a 30×30 grid, and at a series of timepoints $y_{ij} = X_i(u_{ij}, v_{ij}, t_{ij}) + \epsilon_{ij}$, where the random

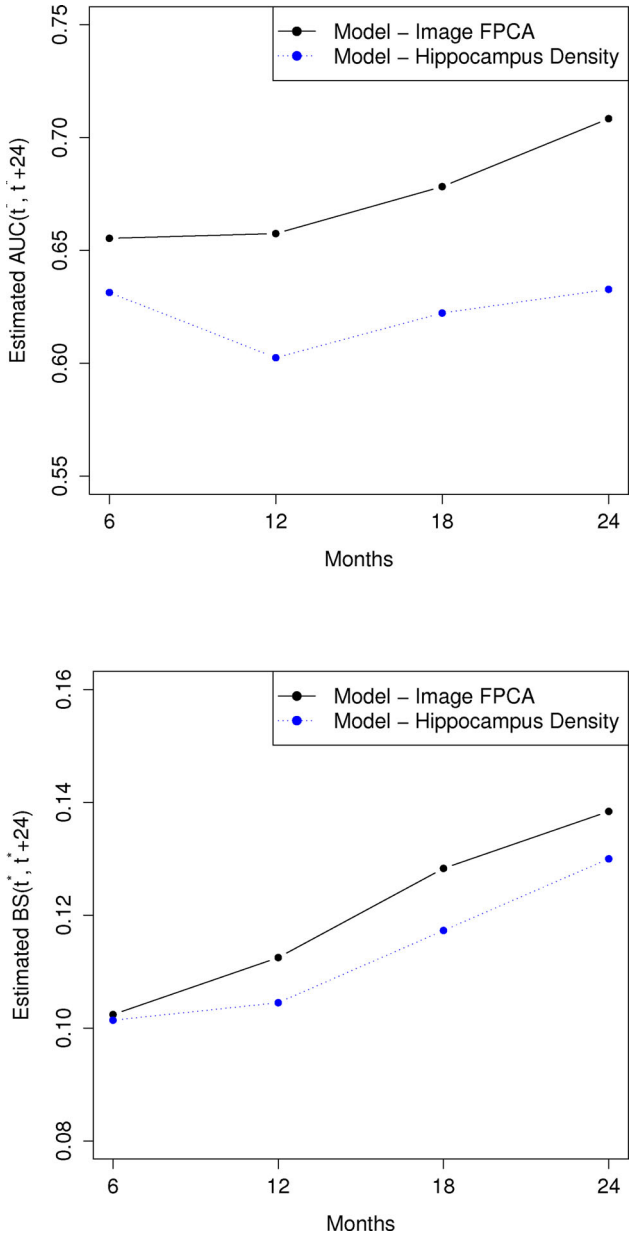

Figure 3. Comparison of AUC and Brier score in time window $(t^*, t^* + 24)$ of the proposed FPCA model based on varying images, versus the model based on hippocampus pixel density given data observed up to $t^* = 6, 12, 18, 24$ under a sliding window framework in the ADNI study.

errors ϵ_{ij} are independently drawn from a normal distribution with a mean of zero and a standard deviation of 0.01. Under dense sampling, the time points are fixed at $\{0, 5, 10, \ldots, 30\}$. Under sparse sampling, for each subject, we randomly draw n_t values from $\{0, 2.5, 5, \ldots, 30\}$ as the time points; we experiment with $n_t = 3$ and 6.

We apply the proposed multi-dimensional FPCA on the simulated data and measure the difference between the extracted FPCs and the true ones. To evaluate how well the proposed method approximates the true FPCs, we use the integrated mean squared error (IMSE), which is defined as

$$\text{IMSE}(\widehat{\psi}_k) = \int\int\int \left\{\widehat{\psi}_k(u,v,t) - \psi_k(u,v,t)\right\}^2 du\, dv\, dt, \quad k = 1, 2.$$

The IMSE is interpreted as the integrated squared differences between the estimated functional principal component functions $\widehat{\psi}_k$ and the true ones ψ_k. Note that as both $\widehat{\psi}_k$ and ψ_k have a norm of 1, that is, the integration of $\widehat{\psi}_k^2$ and ψ_k^2 is equal to 1, the IMSE would fall in the range of $[0, 4]$. A poorly estimated FPC typically has an IMSE greater than 1, and the IMSE of a precisely estimated FPC is very close to 0. Furthermore, to evaluate how accurate the method estimates the FPC scores, we compute the mean squared standardized error (MSE) between the estimated FPC scores $\widehat{\xi}_{ik}$ and the true ones ξ_{ik} averaged across all the subjects $i = 1, \ldots, n$. We compute the IMSEs and MSEs based on 100 data replications. For one data replication with 200 subjects and 6 timepoints per subject, it typically takes less than 5 iterations (in 5 min) to extract the first FPC and 10-15 iterations (in 15 minutes) for the second FPC.

Figures S.3 to S.4 in the supplementary material shows the boxplots of the IMSE of the estimated FPCs $\widehat{\psi}_1(\cdot)$ and $\widehat{\psi}_2(\cdot)$, and the MSE of the estimated FPC scores $\widehat{\xi}_1$ and $\widehat{\xi}_2$ under the dense and sparse sampling scheme with 6 timepoints, with the number

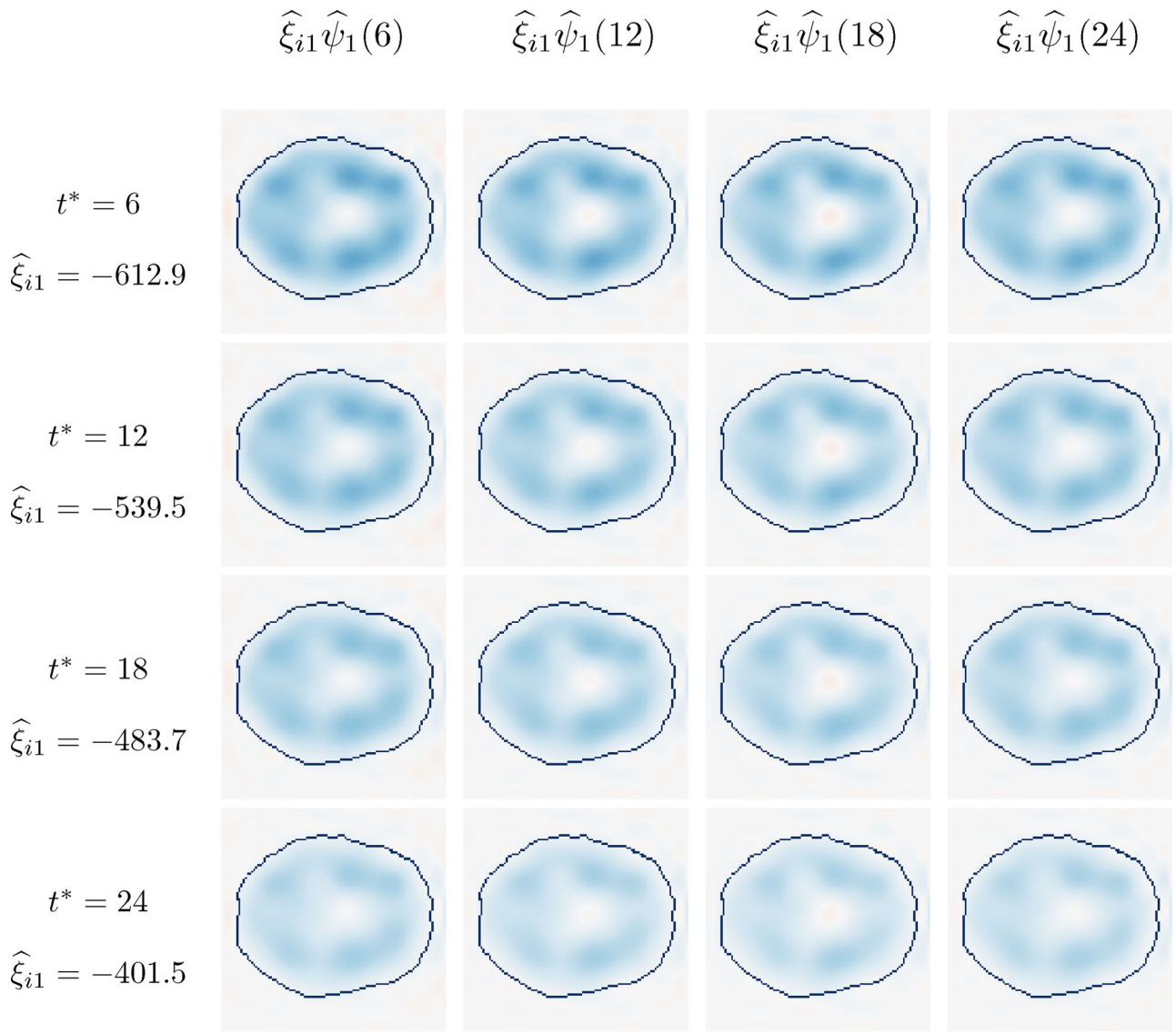

Figure 4. Illustration of the dynamic updating of the FPC scores for the first FPC. Each row plots the product of FPC functions and the corresponding FPC scores for a single subject in the ADNI study, evaluated at month 6, 12, 18, and 24. The top to the bottom rows correspond respectively to the updating of the FPC scores based on the observed longitudinal images up to month 6, 12 ,18, and 24. The magnitudes of positive and negative values are shaded by a gradual color scheme in red and blue, respectively.

of subjects varying from $n = 50$, 200 to 500. It is evident that the IMSE and MSE both converge towards zero as the sample size increases. Furthermore, Figure S.5 in the supplementary material shows the boxplots of the IMSE and MSE when there are as few as only 3 sparsely sampled timepoints for each subject, with varying number of subjects $n = 100$, 400, and 1000, indicating that our method still works well under a very sparse sampling scheme given a large enough sample size. Figure S.6 shows the boxplots of the IMSE and MSE as the number of sparsely sampled timepoints increases from 3 to 5 under a fixed number of subjects $n = 200$, which depicts a decreasing trend. These results establish the empirical validity and consistency of the proposed FPCA method.

5.2. Simulation: Dynamic Survival Prediction

Next, we conduct simulations studies to evaluate the finite-sample performance of the dynamic prediction model. The simulation study is designed to assess the predictive performance of the proposed dynamic prediction framework given the longitudinal images drawn from a prespecified nonlinear stochastic process. The total number of data replications is 100 and in each replication, the sample size is 300 and each subject has longitudinal observations at multiple time points.

To generate the longitudinal images, we consider a complex nonlinear longitudinal submodels as follows. For the jth pixel of the ith subject, rather than assuming the FPCs are known, the pixel value Y_{ij} is simulated from a more complex model

$$Y_{ij} = X_i(u_{ij}, v_{ij}, t_{ij}) + \epsilon_{ij},$$
$$X_i(u_{ij}, v_{ij}, t_{ij}) = \beta_0 + \beta_1 w_i + \sin(6\pi u_{ij}) + 9(v_{ij} - 0.5)^2 + \log(t_{ij} + 1) + b_i,$$

where ϵ_{ij} is a random measurement error term drawn from a normal distribution with a mean of 0, b_i is the subject-specific random effect generated from a standard normal distribution $N(0, 1)$, w_i is a scalar predictor generated from $N(3, 1)$. We set $\beta_0 = 1.5$, $\beta_1 = 2$. We assume that the u_{ij} and v_{ij} are

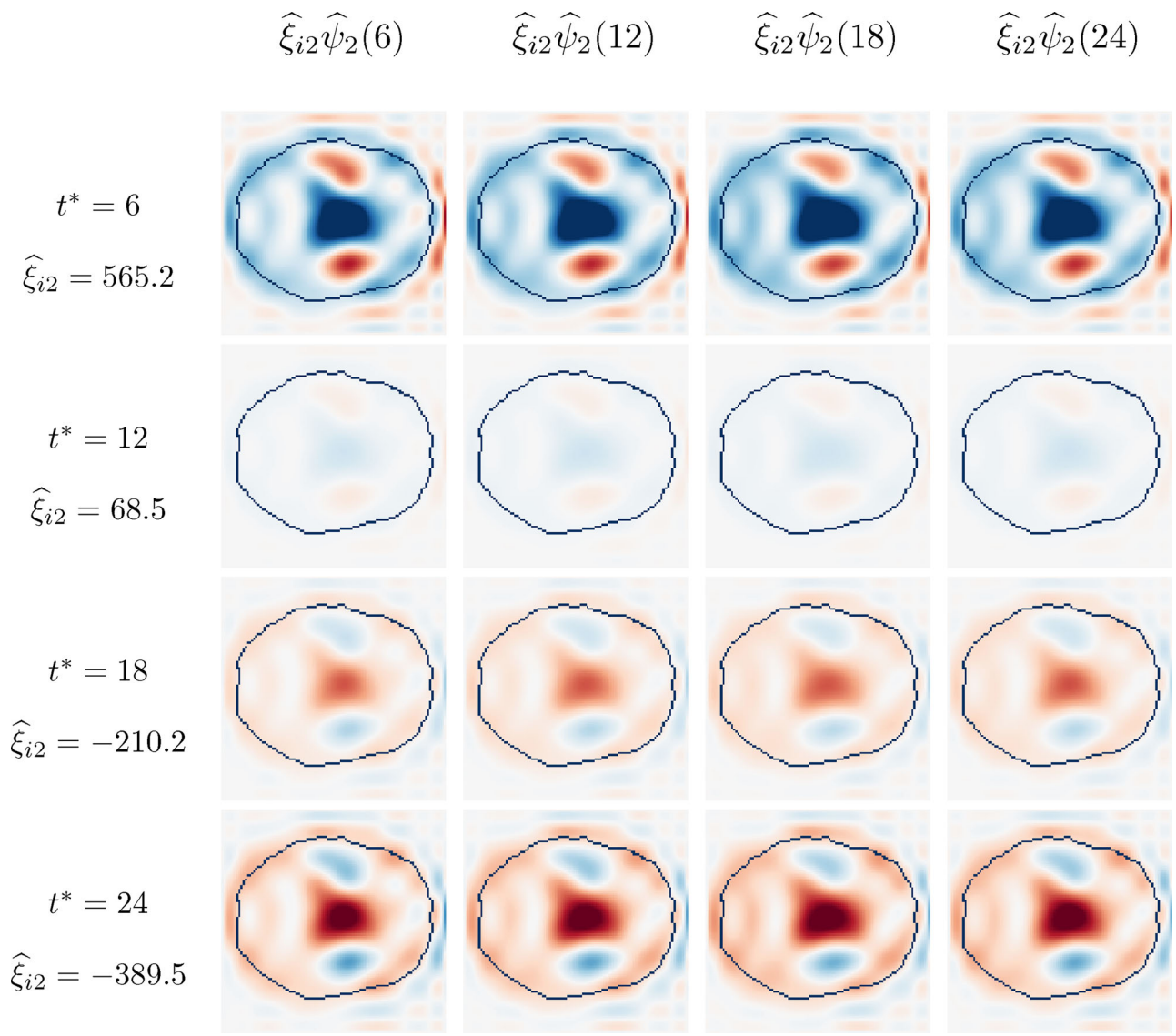

Figure 5. Illustration of the dynamic updating of the FPC scores for the second FPC. Each row plots the product of FPC functions and the corresponding FPC scores for a single subject in the ADNI study, evaluated at month 6, 12, 18, and 24. The top to the bottom rows correspond respectively to the updating of the FPC scores based on the observed longitudinal images up to month 6, 12 ,18, and 24. The magnitudes of positive and negative values are shaded by a gradual color scheme in red and blue, respectively.

sampled from a two-dimensional equidistant grid between 0 and 1 with a distance of 0.1, and the observation time $t_{ij} \in \{0, 3, 6, 9, 12, 15, 18, 21\}$.

To generate the survival time, we first assume a constant baseline hazard function $h_0(t) = \exp(-7)$ and the survival submodel is

$$h_i(t) = h_0(t) \exp\left\{z_i\gamma + \alpha \int_0^t \int_0^1 \int_0^1 X_i(u, v, s) du dv ds\right\}$$

where z_i is sampled from the Bernoulli distribution with a probability of 0.5. We set $\gamma = -2.5$ and $\alpha = 0.1$. To simulate the survival time t_{ij}, we use inverse sampling to first draw $u \sim \text{Unif}(0, 1)$, and then set the survival time to be $H^{-1}(u)$, the inverse of the cumulative hazard function. Moreover, we assume that the censoring time is independent from survival time, and generate the censoring time from a uniform distribution to achieve a censoring rate of about 30%.

After laying out the simulation settings for the longitudinal submodel and the survival submodel, we now describe how the dynamic predictions are constructed and evaluated. In each data replication, out of the 300 simulated samples, we randomly select 200 as the training data and use the remaining 100 for testing the accuracy of survival prediction. We first fit the tensor product B-spline on all the longitudinal images in the training set. The number of basis functions for the two image dimensions and the time dimensions are 12, 12, and 4, respectively. We have experimented with other choices of the number of basis functions (e.g., 9/9/6) and the results are very similar. We then apply the proposed FPCA method to extract the first two FPCs. Then, we fit the Cox proportional hazards model on the FPC scores and baseline covariates. After fitting the model, individualized dynamic predictions are made for each of the 100 subjects in the testing set. To be specific, for subject i, we use the longitudinal images up to time point t^* to compute the FPC score and use the model fitted from the training data to predict the survival probability $\widehat{S}_{n+1}(t^* + \Delta t \mid t^*)$ at some future time point $t^* + \Delta t$. To assess prediction performance of the model, we adopt several measures. The AUCs are computed based on the

predicted survival probabilities of all subjects in the testing set. We also compute the true AUC based on the true conditional survival probabilities $S_{n+1}(t^* + \Delta t \mid t^*)$, which are calculated with known parameter values in the survival model. Moreover, we use the Brier score as a measure of discrimination to assess the discrepancies between the predicted and true risks. For comparison, we also construct the true Brier score which uses the true event probability in place of the estimated probability in the calculation. These dynamic prediction measures are calculated for each time window $(t^*, t^* + \Delta t)$ given the data observed up to $t^* = 9, 12, 15$, for $\Delta t = 3, 6$.

Table S.1 in the supplementary material presents the prediction measures recorded dynamically for different landmark time points and time windows by averaging the results from 100 testing data replications. It is evident that the AUCs of the proposed model are in good alignment with the true AUC, which indicates that the proposed model has good prediction performance in terms of discrimination. The Brier score is a measure of the accuracy of survival predictions in terms of how well the model's predicted probabilities of the event match the actual outcomes, and a value close to 0 would be ideal. We observe that the Brier scores is relatively small and close to the true Brier score, indicating good agreement between the predicted risk and the true risk. Due to randomness in the simulation, in some cases, the Brier scores are lower than the true ones. We also experimented with a varying σ_ϵ, the variances of the random measurement error term ϵ_{ij}, to examine whether model can still maintain the good predictive accuracy under a larger degree of noise in the longitudinal images. It can be seen that the model has a robust performance even when the noise variance increases.

6. Discussion

In this article, we have extended the dynamic prediction framework for univariate trajectories to sparse multi-dimensional functional data. Our proposed FPCA method uses alternating regression on trivariate tensor product splines to extract major mode of variation of a three-dimensional random process. For general arbitrary dimensional random process, a more principled methodological development on FPCA on large-scale or high-dimensional data is needed. As direction of further research, it is worthwhile to explore alternative basis function that are more versatile than trivariate tensor product spline and can handle functional data with larger dimensions (Consagra 2022), or use more computationally efficient dimension reduction technique in the estimation of FPCA (Chen and Jiang 2017). For example, Consagra (2022) proposed the use of data-driven marginal product basis for general arbitrary dimensional functional data, and Chen and Jiang (2017) discussed the use of local smoother and random-projection-based eigendecomposition of the covariance function to accommodate functional data of large scale and dimension.

Furthermore, it is worth noting that in the case of the ADNI data, a subject's MRI brain scans are still obtainable after the occurrence of Alzheimer's disease, and are used in our application; and thus our method would not suffer from the common pitfalls of entanglement between the partially observed longitudinal data and survival time: typically, the separation of estimating the FPC and estimating the survival model may cause bias because the longitudinal data are no longer observable after the event occurs, and conducting FPCA on the longitudinal data that is partially observed up until the event time may miss out on capturing the effect of survival on the longitudinal process. Fortunately, this is not the case for our application on the ADNI data because the MRI scans are still regularly collected even after the subject is diagnosed with AD.

In the ADNI data, the longitudinal functional images are essentially noiseless and observable after the events; this renders our method applicable despite its two-stage nature. However, when there is substantial noise in the functional data, which may lead to biases in the estimated FPC scores under our two-stage approach, the joint modelling approach serves as a particularly useful alternative. Under the joint model, one should no longer separate the estimation of the FPC and the fitting of the survival model, otherwise it may incur biases in the FPC scores. Instead, a one-step joint model approach that can simultaneously model the likelihood of the multidimensional random process, as well as the survival outcome, would render a more statistically principled prediction of the survival probabilities. As avenues of future research, we may develop a one-step multi-dimensional functional joint integrative model that builds upon the work on the functional joint model such that the estimation of the multi-dimensional FPC and the survival prediction can be achieved in a single model.

Supplementary Materials

Supplementary Document: The supplementary document includes the estimation details for the mean function and the additional results of simulation studies.

Acknowledgments

The authors thank the editor, the associate editor and two anonymous referees for many insightful comments. These comments are very helpful for us to improve our work.

Disclosure Statement

No potential competing interest was reported by the authors.

Funding

Dr. Cao's research is supported by the Discovery grant (RGPIN-2023-04057) from the Natural Sciences and Engineering Research Council of Canada (NSERC). Dr. Shi's research is supported by the Discovery grant from NSERC (RGPIN-2021-02963).

ORCID

Jiguo Cao http://orcid.org/0000-0001-7417-6330

References

Chiou, J. M., Chen, Y. T., and Yang, Y. F. (2014), "Multivariate Functional Principal Component Analysis: A Normalization Approach," *Statistica Sinica*, 24, 1571–1596. [1,2]

Chen, L. H., and Jiang, C. R. (2017), "Multi-Dimensional Functional Principal Component Analysis," *Statistics and Computing*, 27, 1181–1192. [1,11]

Consagra, W. (2022), "Methods for Multidimensional Functional Data Analysis in Modern Neuroimaging," Doctoral Dissertation, University of Rochester, Rochester, NY. [11]

Happ, C., and Greven, S. (2018), "Multivariate Functional Principal Component Analysis for Data Observed on Different (Dimensional) Domains," *Journal of the American Statistical Association*, 113, 649–659. [2]

Hasenstab, K., Scheffler, A., Telesca, D., Sugar, C. A., Jeste, S., DiStefano, C., and Senturk, D. (2017), "A Multi-Dimensional Functional Principal Components Analysis of EEG Data," *Biometrics*, 73, 999–1009. [1,2]

Jiang, S., Xie, Y., and Colditz, G. (2020), "Functional Ensemble Survival Tree: Dynamic Prediction of Alzheimer's Disease Progression Accommodating Multiple Time-Varying Covariates," *Journal of the Royal Statistical Society*, Series C, 70, 66–79. [2]

Kong, D., Ibrahim, J. G., Lee, E., and Zhu, H. (2018), "FLCRM: Functional Linear Cox Regression Model," *Biometrics*, 74, 109–117. [2,6]

Lawson, C. L., and Hanson, R. J. (1974), *Solving Least Squares Problems*, Englewood Cliffs: Prentice Hall. [5]

Li, K., and Luo, S. (2019a), "Dynamic Prediction of Alzheimer's Disease Progression Using Features of Multiple Longitudinal Outcomes and Time-to-Event Data," *Statistics in Medicine*, 38, 4804–4818. [2]

Li, K., and Luo, S. (2019b), "Dynamic Predictions in Bayesian Functional Joint Models for Longitudinal and Time-to-Event Data: An Application to Alzheimer's Disease," *Statistical Methods in Medical Research*, 28, 327–342. [2]

Li, K., and Luo, S. (2019c), "Bayesian Functional Joint Models for Multivariate Longitudinal and Time-to-Event Data," *Computatoinal Statistics and Data Analysis*, 129, 14–29. [2]

Li, C., Xiao, L., and Luo, S. (2020), "Fast Covariance Estimation for Multivariate Sparse Functional Data," *Stat*, 9, e245. [1,2]

Lin, J., Li, K., and Luo, S. (2021), "Functional Survival Forests for Multivariate Longitudinal Outcomes: Dynamic Prediction of Alzheimer's Disease Progression," *Statistical Methods in Medical Research*, 30, 99–111. [1,2]

Lin, Z., Wang, L., and Cao, J. (2016), "Interpretable Functional Principal Component Analysis," *Biometrics*, 72, 846–854. [1]

Ma, D., Popuri, K., Bhalla, M., Sangha, O., Lu, D., Cao, J., Jacova, C., Wang, L., and Beg, M. F. (2018), "Quantitative Assessment of Field Strength, Total Intracranial Volume, Sex, and Age Effects on the Goodness of Harmonization for Volumetric Analysis on the ADNI Database," *Human Brain Mapping*, 40, 125–136. [6]

Nie, Y., Wang, L., Liu, B., and Cao, J. (2018), "Supervised Functional Principal Component Analysis," *Statistics and Computing*, 28, 713–723. [2]

Nie, Y., and Cao, J. (2020), "Sparse Functional Principal Component Analysis in a New Regression Framework," *Computational Statistics & Data Analysis*, 152, 107016. [2]

Nie, Y., Wang, L., and Cao, J. (2022), "Recovering the Underlying Trajectory from Sparse and Irregular Longitudinal Data," *Canadian Journal of Statistics*, 50, 122–141. [2]

Popuri, K., Ma, D., Wang, L., and Beg, M. F. (2020), "Use Machine Learning to Quantify Structural MRI Neurodegeneration Patterns of Alzheimer's Disease into Dementia Score: Independent Validation on 8834 Images from ADNI, AIBL, OASIS and MIRIAD Databases," *Human Brain Mapping*, 41, 4127–4147. [6]

Ramsay, J. O., and Dalzell, C. (1991), "Some Tools for Functional Data Analysis," *Journal of the Royal Statistical Society*, Series B, 53, 539–572. [1]

Sang, P., Wang, L., and Cao, J. (2017), "Parametric Functional Principal Component Analysis," *Biometrics*, 73, 802–810. [1]

Shi, H., Dong, J., Wang, L., and Cao, J. (2021), "Functional Principal Component Analysis for Longitudinal Data with Informative Dropout," *Statistics in Medicine*, 40, 712–724. [2]

Shi, H., Yang, Y., Wang, L., Ma, D., Beg, M. F., Pei, J., and Cao, J. (2022), "Two-Dimensional Functional Orthogonal Approximation Method for Image Feature Extraction," *Journal of Computational and Graphical Statistics*, 31, 1127–1140. [2]

Silverman, B. W. (1996), "Smoothed Functional Principal Components Analysis by Choice of Norm," *The Annals of Statistics*, 24, 1–24. [1]

Welsh, A. H., Lin, X., and Carroll, R. J. (2002), "Marginal Longitudinal Nonparametric Regression," *Journal of the American Statistical Association*, 97, 482–493. [1]

Yao, F., Müller, H. G., and Wang, J. L. (2005), "Functional Data Analysis for Sparse Longitudinal Data," *Journal of the American Statistical Association*, 100, 577–590. [3]

Zou, H., Li, K., Zeng, D., and Luo, S. (2021), "Bayesian Inference and Dynamic Prediction of Multivariate Joint Model with Functional Data: An Application to Alzheimer's Disease," *Statisics in Medicine*, 40, 6855–6872. [2]

Zou, H., Xiao, L., Zeng, D., Luo, S. (2023), "Multivariate Functional Mixed Model with MRI Data: An Application to Alzheimer's Disease," *Statistics in Medicine*, 42, 1492–1511. [2]