

Partial least squares for functional joint models with applications to the Alzheimer's disease neuroimaging initiative study

Yue Wang¹ | Joseph G. Ibrahim² | Hongtu Zhu² ¹Department of Biostatistics, University of Washington, Seattle, Washington²Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina**Correspondence**Yue Wang, Department of Biostatistics, University of Washington, Seattle, WA 98195.
Email: taryue@gmail.com**Funding information**

NSF, Grant/Award Numbers: DMS-1407655, SES-1357666; NIH, Grant/Award Numbers: R01MH086633, R01MH116527

Abstract

Many biomedical studies have identified important imaging biomarkers that are associated with both repeated clinical measures and a survival outcome. The functional joint model (FJM) framework, proposed by Li and Luo in 2017, investigates the association between repeated clinical measures and survival data, while adjusting for both high-dimensional images and low-dimensional covariates based on the functional principal component analysis (FPCA). In this paper, we propose a novel algorithm for the estimation of FJM based on the functional partial least squares (FPLS). Our numerical studies demonstrate that, compared to FPCA, the proposed FPLS algorithm can yield more accurate and robust estimation and prediction performance in many important scenarios. We apply the proposed FPLS algorithm to a neuroimaging study. Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database.

KEYWORDS

high-dimensional data, longitudinal data, neuroimaging data, survival data

1 | INTRODUCTION

Many prospective cohort studies and clinical trials investigating neurodegenerative diseases such as Alzheimer's disease (AD) collect repeated measurements of clinical variables, event history, and biomedical imaging data. A motivating example is the Alzheimer's Disease Neuroimaging Initiative (ADNI) study. ADNI currently has four phases: ADNI1, ADNI-GO, ADNI2, and ADNI3, and the primary goal is to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), and neuropsychological assessments can be used to measure the progression of AD. Participants were assessed at multiple visits. At each visit, various clinical measures, brain images, and neuropsychological assessments were collected.

As mild cognition impairment (MCI) is known as a transitional stage between normal cognition and AD, it is of great interest to predict the progression from MCI to AD. Thus, we selected 236 patients who had been diagnosed with MCI from ADNI1 without missing data in the covariates of inter-

est. The selected patients had at least one Alzheimer's Disease Assessment Scale-Cognitive (ADAS-Cog) score after the baseline measurement. The ADAS-Cog score measures cognition functions and ranges from 0 to 70, with a higher score indicating poorer cognitive function. We consider the AD diagnosis as the survival event of interest. Among the 236 individuals, 92 individuals progressed to AD before the completion of ADNI1 and the remaining 144 individuals did not. Thus, the time of conversion from MCI to AD can be treated as right censored time-to-event data and the censoring is non-informative about the progression to AD. Demographic information of the selected 236 patients and summary statistics of the ADAS-Cog score can be found in the Supporting Information.

Figure 1 displays the average ADAS-Cog score at each follow-up time separately for the MCI and AD group. It can be seen that the ADAS-Cog score increases with time for the AD group, whereas for the MCI group, the trend is not evident. The AD group tends to have higher ADAS-Cog scores, indicating that the ADAS-Cog score may be predictive

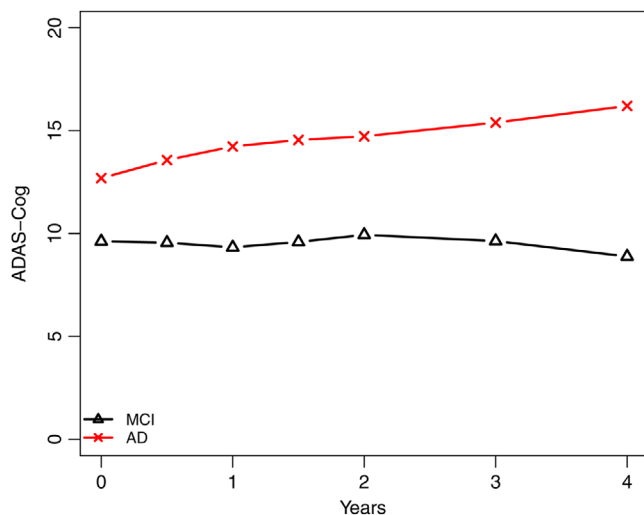


FIGURE 1 The ADAS-Cog score at year 0, 0.5, 1, 1.5, 2, 3, and 4 for both MCI and AD patients. At each time point, the score is averaged on those subjects who have not dropped out before this time point. This figure appears in color in the electronic version of this article, and any mention of color refers to that version

to the progression of AD. Moreover, a lot of existing work reported the association between brain imaging predictors with the progression of AD. For example, AD and MCI patients were shown to have 27% and 11% smaller hippocampal volumes, respectively, as compared with normal controls (Du *et al.*, 2001). Lee *et al.* (2015) and Kong *et al.* (2018) both demonstrated the predictive value of the hippocampus surface data to the progression of AD.

Thus, it is of great interest to examine the association between longitudinal ADAS-Cog scores and the progression of AD while adjusting for high-dimensional imaging predictors. Li and Luo (2017) proposed a functional joint model (FJM) for this purpose. FJM integrates functional regression models with joint models of longitudinal and time-to-event data. There is a rich literature in both areas. For instance, in the past decades, generalized functional linear models (GFLM) have been widely discussed and applied in various areas including finance, biology, and sociology. Please see Cardot *et al.* (1999, 2003), Müller and Stadtmüller (2005), and the references therein for an extensive review of GFLM. Then, Yao *et al.* (2005) extended GFLM to longitudinal data, and Lee *et al.* (2015) and Kong *et al.* (2018) both extended GFLM to the proportional hazards model. Some of the earliest work on joint models of longitudinal and time to event data is in Tsiatis *et al.* (1995) and Wulfsohn and Tsiatis (1997).

The major challenge of fitting a functional model, including FJM, is the ultrahigh dimensionality of the functional predictor. A common strategy is to find the “best” low-dimensional features that approximate the high-dimensional functional predictor, which is also known as dimension reduction. The functional principal component analysis (FPCA)

has been a popular dimension-reduction tool for functional data over decades. See extensive reviews and applications of FPCA in Besse and Ramsay (1986), Ramsay and Dalzell (1991), Boente and Fraiman (2000), James *et al.* (2000), Hall and Hosseini-Nasab (2006), Müller and Yao (2010), Li and Luo (2017), and the references therein. Based on the eigen-decomposition of the covariance kernel of the functional data, FPCA finds the low-dimensional space that preserves most of the variation in the functional data; that is, the space spanned by top eigenfunctions. Although FPCA serves as a promising tool for exploring functional data, there are two major concerns when it comes to regression. First, as the unknown slope function may not necessarily lie in the space spanned by top eigenfunctions, the estimation and prediction accuracy may suffer if the number of eigenfunctions used for regression is underestimated. Second, it is well known that consistent estimation of eigenvectors is highly challenging in ultrahigh-dimensional settings (Jung and Marron, 2009), especially for tail eigenvectors. These two concerns bring up a natural question: if (part of) the unknown slope function lies in the space spanned by some tail eigenfunctions, which cannot be accurately estimated due to the limited sample size, is there an alternative and better approach to estimate the unknown slope function?

To address this question, we consider the functional partial least squares (FPLS), first proposed by Preda and Leveder (2005) for functional linear models (FLM). Compared to FPCA, FPLS has two major advantages. First, FPLS does not depend on accurate estimates of eigenfunctions. Second, FPLS incorporates information from the outcome so that the top FPLS basis functions are always the most predictive to the outcome. However, due to its computational and theoretical complexity, FPLS has not gained enough popularity until the recent work by Delaigle and Hall (2012) that proposes a simplified FPLS algorithm for FLM, called alternative PLS (APLS). Through numerical studies, Delaigle and Hall (2012) demonstrates that APLS can capture the interaction between the functional predictor and the outcome using a fewer number of components than FPCA.

Motivated by the encouraging performance of APLS, in this article, we propose an FPLS algorithm for the estimation and prediction of FJM. Our specific contributions include: (a) We extend APLS to the complex FJM framework in a rigorously mathematical manner; (b) we show by numerical studies that the proposed FPLS algorithm can yield considerably more accurate, robust estimation and prediction results than FPCA in many important scenarios; and (c) unlike the FPCA-based algorithm in Li and Luo (2017) that only handles baseline imaging data, our FPLS algorithm can deal with longitudinal imaging data; this allows to dynamically predict disease progression. The rest of the paper is organized as follows. Section 2 introduces the FPLS algorithm for FJM. Section 3 discusses the details of the implementation of the FPLS algo-

rithm. Section 4 presents a simulation study with ultrahigh-dimensional images in various settings. Section 5 presents a thorough analysis of the selected 236 ADNI patients. Concluding remarks and discussions are presented in Section 6.

2 | METHODS

In this section, we first introduce the FJM proposed by Li and Luo (2017). Then, we review APLS for FLM proposed by Delaigle and Hall (2012). Next, we extend APLS for FJM.

2.1 | FJM

For each subject $i = 1, \dots, n$ at visit $k = 1, \dots, K_i$, we observe $\{y_{ik}, \mathbf{x}_i, \mathbf{z}_{ik}, \omega_i, T_i, \Delta_i\}$, where y_{ik} is the outcome of interest observed at time t_{ik} , \mathbf{z}_{ik} is a $p_z \times 1$ vector observed at t_{ik} , and ω_i is a $p_\omega \times 1$ vector of time-invariant covariates, which may overlap with \mathbf{z}_{ik} . The $\mathbf{x}_i = (x_i(s) : s \in S)$ is baseline imaging (functional) data observed at a set of grid points in a nondegenerate, compact space $S \subset R^K$ for the i th subject, where $K > 0$ is a positive integer. T_i is the observed survival time, defined as $T_i = \min\{T_i^*, C_i\}$, where T_i^* and C_i are, respectively, the true event time and censoring time. The event is observed if $\Delta_i = 1$, and censored otherwise. It is assumed throughout the paper that the censoring mechanism is independent of the true event process.

The FJM in Li and Luo (2017) consists of a longitudinal model and a survival model. The longitudinal model is given by

$$\begin{aligned} y_{ik} &= m_i(t_{ik}) + \epsilon_{ik} \\ &= \beta_0 + \mathbf{z}_{ik}^T \beta_1 + q_{ik}^T u_i + \int_S x_i(s) b_0(s) ds + \epsilon_{ik}, \end{aligned} \quad (1)$$

where $m_i(t_{ik})$ is the unobserved true longitudinal trajectory of the i th subject at t_{ik} , q_{ik} is a subset of \mathbf{z}_{ik} , β_0 is the intercept, β_1 is a $p_z \times 1$ vectors of regression coefficients, and $b_0(s)$ is the functional parameter that characterizes the association between $x_i(s)$ and y_{ik} . We also assume $u_i \stackrel{i.i.d.}{\sim} N(0, \Sigma_u)$ and $\epsilon_{ik} \stackrel{i.i.d.}{\sim} N(0, \sigma_\epsilon^2)$. The survival model is given by

$$\begin{aligned} \lambda_i(t|\omega_i, x_i(s), u_i) \\ = \lambda_0(t) \exp \left(\omega_i^T \gamma + \int_S x_i(s) b_1(s) ds + \alpha m_i(t) \right), \end{aligned} \quad (2)$$

where $\lambda_0(t)$ is an unknown baseline hazard function, γ is a $p_\omega \times 1$ vector of regression coefficients, and $b_1(s)$ is the functional parameter that characterizes the association between $x_i(s)$ and the survival outcome. The scalar parameter α quan-

tifies the association between the true longitudinal trajectory and the survival process at the same time.

Remark 1. We discuss our assumptions on FJM (1) and (2). First, the longitudinal marker y_{ik} in (1) is assumed to be normally distributed, which is a standard assumption in the literature. The association between the longitudinal and time-to-event processes is represented by the proportional hazards model (2), in which $m_i(t)$ is assumed to be continuous. For definiteness, the conditional hazards function $\lambda_i(t|\omega_i, x_i(s), u_i)$ is taken to depend linearly on the longitudinal marker through the current $m_i(t)$.

Remark 2. The FJM (1) and (2) can be extended to a more general FJM framework as follows:

$$\begin{aligned} y_{ik} &= m_i(t_{ik}) + \epsilon_{ik} \\ &= \beta_0 + \mathbf{z}_{ik}^T \beta_1 + q_{ik}^T u_i + \int_S x_i^{(l)}(s, t_{ik}) b_0(s) ds + \epsilon_{ik} \end{aligned} \quad (3)$$

and

$$\begin{aligned} \lambda_i(t|\omega_i, x_i(s), u_i) \\ = \lambda_0(t) \exp \left(\omega_i^T \gamma + \int_S x_i^{(s)}(s) b_1(s) ds + \alpha m_i(t) \right). \end{aligned} \quad (4)$$

Note that compared with (1), the longitudinal model (3) involves longitudinal imaging data. This extension is appealing in many applications because it allows dynamic prediction of disease progression. It is worth noting that the FPLS algorithm that will be introduced later can be seamlessly applied to this general FJM framework (3) and (4). But to simplify the notation, we stick to the FJM (1) and (2) in the following sections for illustrating our key ideas.

2.2 | The APLS algorithm

We review APLS proposed by Delaigle and Hall (2012) for the following FLM:

$$y = a_0 + \int_S x(s) b(s) ds + \epsilon, \quad (5)$$

where ϵ is a scalar random variable with $E(\epsilon|\mathbf{x}) = 0$, a_0 is an intercept, and $b(s)$ is an unknown coefficient function. Let $K(\psi)(t) = \int_S K(s, t) \psi(s) ds$ be a functional operator where $K(s, t) = \text{Cov}(x(s), x(t))$. The first p APLS basis functions can be constructed as $K(b), \dots, K^p(b)$, where $K^{j+1}(b)(s) = \int_S K^j(b)(t) K(s, t) dt$. Compared to the FPCA basis functions, that is, the eigenfunctions of $K(s, t)$, the APLS basis functions have two important features. First, the APLS basis functions involve the unknown parameter $b(s)$, indicating that the APLS basis functions incorporate information from both the functional predictor and the outcome. Second, the APLS

basis functions are not orthonormal. To see the latter feature clearly, consider an example that $x(s) = \sum_{j=1}^{\infty} \sqrt{\theta_j} \xi_j \phi_j(s)$, where $\xi_j \stackrel{i.i.d}{\sim} N(0, 1)$. The $\theta_1 \geq \theta_2 \geq \dots \geq 0$ is a sequence of nonincreasing eigenvalues and $\phi_1(\cdot), \phi_2(\cdot), \dots$ are their corresponding orthonormal eigenfunctions. It can be seen that $K(s, t) = \text{Cov}(x(s), x(t)) = \sum_{j=1}^{\infty} \theta_j \phi_j(s) \phi_j(t)$. Write $b(s) = \sum_{j=1}^{\infty} b_j \phi_j(s)$, where $b_j = \int_S b(s) \phi_j(s) ds$. Then, it can be seen that the p th APLS basis function is given by $K^p(b)(s) = \sum_{j=1}^{\infty} \theta_j^p b_j \phi_j(s)$. Therefore, it can be checked that $\int_S K^k(b)(s) K^l(b)(s) ds = \sum_{j=1}^{\infty} \theta_j^{k+l} b_j^2 > 0$ for any $k \neq l$, indicating that any two APLS basis functions are not orthogonal to each other.

Although $b(s)$ is unknown, a consistent estimator of $K(b) = \int_S K(s, t) b(s) ds$ is given by

$$\widehat{K}(b)(s) = n^{-1} \sum_{i=1}^n \{x_i(s) - \bar{x}(s)\} \{y_i - \bar{y}\}, \quad (6)$$

where $\bar{y} = n^{-1} \sum_{i=1}^n y_i$ and $\bar{x}(s) = n^{-1} \sum_{i=1}^n x_i(s)$. Then, we can sequentially estimate all APLS basis functions by using $\widehat{K}^{j+1}(b)(t) = \int_S \widehat{K}^j(b)(s) \widehat{K}(s, t) ds$ for $j \geq 1$, where $\widehat{K}(s, t) = n^{-1} \sum_{i=1}^n \{x_i(s) - \bar{x}(s)\} \{x_i(t) - \bar{x}(t)\}$. Then, an estimator of $b(s)$ in (5) using p APLS basis functions is given by $\widehat{b}_p(s) = \sum_{j=1}^p \widehat{t}_j \widehat{K}^j(b)(s)$, where

$$\begin{aligned} (\widehat{t}_1, \dots, \widehat{t}_p) = \operatorname{argmin}_{t_1, \dots, t_p} \sum_{i=1}^n \left\{ y_i - \bar{y} \right. \\ \left. - \sum_{j=1}^p t_j \int_S \widehat{K}^j(b)(s) [x_i(s) - \bar{x}(s)] ds \right\}^2. \quad (7) \end{aligned}$$

2.3 | The RAPLS algorithm

The APLS algorithm cannot be directly applied to FJM (1) and (2) due to two reasons. First, it does not adjust for additional scalar covariates. Second, the estimator $\widehat{K}(b)(s)$ in (6) is not a good estimator of $K(b)(s)$ when the relationship between $x(s)$ and y is nonlinear. To bridge the gap, we first extend the APLS algorithm to the following model:

$$y_i = \mathbf{z}_i^T \alpha + \int_S x_i(s) b(s) ds + \epsilon_i, \quad (8)$$

where α is a $p_z \times 1$ vector of regression coefficients and ϵ_i is an error term with $E(\epsilon_i | \mathbf{x}_i, \mathbf{z}_i) = 0$ and $E(\epsilon_i^2 | \mathbf{x}_i, \mathbf{z}_i) < \infty$. It is assumed that \mathbf{z}_i includes a constant 1. To estimate $b(s)$ and α , we develop a residual-based APLS (RAPLS) algorithm. Specifically, model (8) can be rewritten in the following matrix form:

$$Y = Z\alpha + \int_S X(s) b(s) ds + \epsilon, \quad (9)$$

where $\epsilon = (\epsilon_1, \dots, \epsilon_n)^T$, $Y = (y_1, \dots, y_n)^T$, $Z = [\mathbf{z}_1, \dots, \mathbf{z}_n]^T$, and $X(s) = (x_1(s), \dots, x_n(s))^T$. Let $H_Z = Z(Z^T Z)^{-1} Z^T$ be the orthogonal projection matrix onto the column space of Z and $M_Z = I_n - H_Z$, where I_n denotes an $n \times n$ identity matrix. By multiplying both sides of (9) by M_Z , we have

$$\begin{aligned} Y_Z^\perp &= M_Z Y = \int_S M_Z X(s) b(s) ds + M_Z \epsilon \\ &= \int_S X_Z^\perp(s) b(s) ds + \epsilon_Z^\perp, \quad (10) \end{aligned}$$

where $\epsilon_Z^\perp = M_Z \epsilon$. Model (10) can be regarded as a special case of model (5) with the response vector Y_Z^\perp and the functional covariate $X_Z^\perp(s)$. In this case, we need to introduce a new functional operator as follows:

$$\begin{aligned} K_Z^\perp(\psi)(t) &= \int_S K_Z^\perp(s, t) \psi(s) ds \\ &= \int_S [K(s, t) - \operatorname{tr}\{[E(\mathbf{z} \otimes \mathbf{z}^2)]^{-1} E[\mathbf{z}x(t)] E[\mathbf{z}^T x(s)]\}] \psi(s) ds. \end{aligned}$$

The operator $K_Z^\perp(\psi)(t)$ corrects the correlation between \mathbf{z} and $x(s)$. If \mathbf{z} and $x(s)$ are independent, then $K_Z^\perp(\psi)$ reduces to $K(\psi)$. Therefore, the first p RAPLS basis functions for model (10) are given by $(K_Z^\perp)^1(b), \dots, (K_Z^\perp)^p(b)$. A consistent estimator of $(K_Z^\perp)^j(b)(t)$ is given by $(\widehat{K}_Z^\perp)^j(b)(t) = n^{-1} Y_Z^{\perp T} X_Z^\perp(t)$.

Then, we can sequentially estimate all RAPLS basis functions by using

$$(\widehat{K}_Z^\perp)^{j+1}(b)(t) = \int_S (\widehat{K}_Z^\perp)^j(b)(s) \widehat{K}_Z^\perp(s, t) ds \quad \text{for } 1 \leq j \leq p-1,$$

where $\widehat{K}_Z^\perp(s, t) = n^{-1} X_Z^{\perp T}(s) X_Z^\perp(t)$. Given $(\widehat{K}_Z^\perp)^1(b)(t), \dots, (\widehat{K}_Z^\perp)^p(b)(t)$, we define

$$\begin{aligned} (\widehat{t}_{1,z}, \dots, \widehat{t}_{p,z}) \\ = \operatorname{argmin}_{(t_1, \dots, t_p)} \left\| Y_Z^\perp - \sum_{j=1}^p t_j \int_S (\widehat{K}_Z^\perp)^j(b)(t) X_Z^\perp(t) ds \right\|^2, \end{aligned}$$

where $\|\mathbf{a}\|^2 = \mathbf{a}^T \mathbf{a}$ for any column vector \mathbf{a} . Therefore, one can estimate $b(s)$ and α , respectively, by $\widehat{b}_p(s) = \sum_{j=1}^p \widehat{t}_{j,z} (\widehat{K}_Z^\perp)^j(b)(s)$ and $\widehat{\alpha}_p = (Z^T Z)^{-1} Z^T [Y - \int_S X(s) \widehat{b}_p(s) ds]$.

2.4 | The FPLS algorithm for FJM

We develop an FPLS algorithm for the FJM (1) and (2) in an iterative way by integrating RAPLS with the iterative reweighted least squares (IRLS; Green, 1984).

Before discussing the details, we introduce some additional notations. Let $y_i = (y_{i1}, \dots, y_{iK_i})^T$, $\mathbf{z}_i = (\mathbf{z}_{i1}, \dots, \mathbf{z}_{iK_i})$, and $q_i = (q_{i1}, \dots, q_{iK_i})$. We define $\delta_i = q_i^T u_i + \epsilon_i$, where $\epsilon_i = (\epsilon_{i1}, \dots, \epsilon_{iK_i})^T$. Let 1_{K_i} denote a $K_i \times 1$ vector of ones and $\Lambda_0(t) = \int_0^t \lambda_0(u) du$ denote the cumulative baseline hazard function. Denote $\Theta = \{\beta_0, \beta_1, \gamma, \alpha, \Sigma_u, \sigma_\epsilon^2\}$ and let p_0 and p_1 be the number of RAPLS basis functions used for the estimation of $b_0(s)$ and $b_1(s)$, respectively. We further denote $\Theta_{p_0, p_1}^{(m)}$, $b_{0, p_0}^{(m)}(s)$, $b_{1, p_1}^{(m)}(s)$, and $\Lambda_{0, p_0, p_1}^{(m)}(\cdot)$ as the evaluation of Θ , $b_0(s)$, $b_1(s)$, and $\Lambda_0(\cdot)$, respectively, at the m th iteration for $m \geq 1$. In the following discussion, we elaborate on how to obtain $\Theta_{p_0, p_1}^{(m)}$, $b_{0, p_0}^{(m)}(s)$, $b_{1, p_1}^{(m)}(s)$, and $\Lambda_{0, p_0, p_1}^{(m)}(\cdot)$ given their values at the previous iteration for $m \geq 1$ in three steps.

Step 1: We first rewrite model (1) as follows:

$$y_i = 1_{K_i} \beta_0 + \mathbf{z}_i^T \beta_1 + 1_{K_i} \int_S x_i(s) b_0(s) ds + \delta_i. \quad (11)$$

It can be seen that $\text{Cov}(\delta_i | \mathbf{z}_i, x_i(s), q_i) = q_i \Sigma_u q_i^T + \sigma_\epsilon^2 I_{K_i}$, referred to as V_i hereafter. Hence, given $\Theta_{p_0, p_1}^{(m)}$, V_i can be evaluated as $V_i^{(m)}$ at the m th iteration. By multiplying $\{V_i^{(m)}\}^{-1/2}$ to the left on both sides of (11), (11) can be reformulated as an ordinary least squares (OLS) problem, given by

$$\begin{aligned} \{V_i^{(m)}\}^{-1/2} y_i &= \{V_i^{(m)}\}^{-1/2} (1_{K_i} \beta_0 + \mathbf{z}_i^T \beta_1) \\ &+ \{V_i^{(m)}\}^{-1/2} 1_{K_i} \int_S x_i(s) b_0(s) ds \\ &+ \{V_i^{(m)}\}^{-1/2} \delta_i. \end{aligned} \quad (12)$$

Note that (12) has the same form as (8), and thus we can calculate the top p_0 RAPLS basis functions for (12). To simplify the notation, we denote these RAPLS basis functions by $\psi_1^{(m)}(s), \dots, \psi_p^{(m)}(s)$. Then, model (1) can be approximated by

$$y_{ik} = \beta_0 + \mathbf{z}_{ik}^T \beta_1 + q_{ik}^T u_i + \sum_{j=1}^{p_0} b_{0,j} \left(\int_S x_i(s) \psi_j^{(m)}(s) ds \right) + \epsilon_{ik}, \quad (13)$$

and $b_0(s)$ can be approximated by $\sum_{j=1}^{p_0} b_{0,j} \psi_j^{(m)}(s)$.

Step 2: We define $\mu_i = \Lambda_0(T_i) \exp(\omega_i^T \gamma + \int_S x_i(s) b_1(s) ds + \alpha m_i(T_i))$ for $i = 1, \dots, n$. Given $\Theta_{p_0, p_1}^{(m)}$, $b_{1, p_1}^{(m)}(s)$ and $\Lambda_{0, p_0, p_1}^{(m)}(\cdot)$, μ_i still cannot be evaluated because of the unobserved random effects u_i in $m_i(T_i)$. To deal with this, for any integrable function $h(u_i)$, we define

$$\begin{aligned} E_i^{(m)}\{h(u_i)\} \\ = E \left\{ h(u_i) | y_i, T_i, \Delta_i, \Theta_{p_0, p_1}^{(m)}, b_{0, p_0}^{(m)}(s), b_{1, p_1}^{(m)}(s), \Lambda_{0, p_0, p_1}^{(m)}(\cdot) \right\} \end{aligned}$$

as the posterior expectation of $h(u_i)$ at the m th iteration. Then, we can evaluate μ_i as

$$\begin{aligned} \mu_i^{(m)} &= \Lambda_{0, p_0, p_1}^{(m)}(T_i) \times \\ &E_i^{(m)} \left\{ \exp \left(\omega_i^T \gamma + \int_S x_i(s) b_1(s) ds + \alpha m_i(T_i) \right) \right\}. \end{aligned}$$

Following IRLS, for $i = 1, \dots, n$, we can define the pseudo-response $\tilde{y}_i^{(m)}$ for subject i at the m th iteration as

$$\tilde{y}_i^{(m)} = \int_S x_i(s) b_{1, p_1}^{(m)}(s) ds + \left\{ \mu_i^{(m)} \right\}^{-1} \left(\Delta_i - \mu_i^{(m)} \right). \quad (14)$$

We provide the details for deriving (14) in the Supporting Information. Next, we consider the following model:

$$\begin{aligned} \left\{ \mu_i^{(m)} \right\}^{1/2} \tilde{y}_i^{(m)} \\ = \left\{ \mu_i^{(m)} \right\}^{1/2} \int_S x_i(s) b_1(s) ds + \delta_i^{(m)}, \text{ for } i = 1, \dots, n. \end{aligned} \quad (15)$$

Note that (15) has the same form as (8), and thus we calculate the top p_1 RAPLS basis functions for (15), denoted by $\zeta_1^{(m)}(s), \dots, \zeta_p^{(m)}(s)$. Similar to Step 1, model (2) can be approximated by

$$\begin{aligned} \lambda_i(t | \omega_i, x_i(s), u_i) \\ = \lambda_0(t) \exp \left(\omega_i^T \gamma + \sum_{j=1}^{p_1} b_{1,j} \left(\int_S x_i(s) \zeta_j^{(m)}(s) ds \right) + \alpha m_i(t) \right), \end{aligned} \quad (16)$$

and $b_1(s)$ can be approximated by $\sum_{j=1}^{p_1} b_{1,j} \zeta_j^{(m)}(s)$.

Step 3: Note that FJM (1) and (2) are now, respectively, approximated by the standard joint model with low-dimensional covariates (13) and (16). Due to the unobserved random effects u_i , we propose to fit (13) and (16) by the EM algorithm (Dempster *et al.*, 1977). The details of the EM algorithm are given in the Supporting Information. Let $\Theta_{p_0, p_1}^{(m+1)}$, $\Lambda_{0, p_0, p_1}^{(m+1)}(\cdot)$, $b_{0, j}^{(m+1)}$, and $b_{1, j}^{(m+1)}$ denote the EM estimates of Θ , $\Lambda_0(\cdot)$, $b_{0, j}$, and $b_{1, j}$, respectively. Then, $b_{0, p_0}^{(m+1)}(s)$ and $b_{1, p_1}^{(m+1)}(s)$ are, respectively, given by

$$\begin{aligned} b_{0, p_0}^{(m+1)}(s) &= \sum_{j=1}^{p_0} b_{0, j}^{(m+1)} \psi_j^{(m)}(s) \text{ and } b_{1, p_1}^{(m+1)}(s) \\ &= \sum_{j=1}^{p_1} b_{1, j}^{(m+1)} \zeta_j^{(m)}(s). \end{aligned} \quad (17)$$

With appropriate initial values, the proposed FPLS algorithm for FJM is done by iterating between Steps 1-3 until $\int_S (b_{0, p_0}^{(m+1)}(s) - b_{0, p_0}^{(m)}(s))^2 ds + \int_S (b_{1, p_1}^{(m+1)}(s) - b_{1, p_1}^{(m)}(s))^2 ds$ is

less than a given threshold κ_0 , which is set to be 10^{-6} in the following numerical studies.

3 | IMPLEMENTATION

3.1 | Convergence

Based on our experience in numerical studies, we make several comments for facilitating the convergence of the proposed FPLS algorithm.

1. The stability of the proposed FPLS algorithm may suffer from the nonorthogonality of the RAPLS basis functions. As the key idea of the proposed FPLS algorithm is to find the low-dimensional space spanned by the RAPLS basis functions, we suggest orthogonalizing the RAPLS basis functions in Steps 1 and 2 before running the EM algorithm in Step 3.
2. As in most regression problems, it may be helpful to put all covariates on comparable scales. In particular, singular Hessian matrix may appear during the EM iterations (Step 3) if the two integrals $\int_S x_i(s)\psi_j^{(m)}(s)ds$ and $\int_S x_i(s)\zeta_j^{(m)}(s)ds$ are vastly different from the scalar predictors. This may happen in practice if $x_i(s)$ is not appropriately scaled because these two integrals are approximated by summing over millions of grid points. We elaborate on the scaling process as follows. As suggested in the first comment, we can orthogonalize $\psi_j^{(m)}(s)$ and $\zeta_j^{(m)}(s)$ such that $\int_S \psi_{j_1}^{(m)}(s)\psi_{j_2}^{(m)}(s)ds = \int_S \zeta_{j_1}^{(m)}(s)\zeta_{j_2}^{(m)}(s)ds = I(j_1 = j_2)$. Then, by using the Cauchy-Schwarz inequality, it is easy to show that

$$\left| \int_S \bar{x}(s)\psi_j^{(m)}(s)ds \right| \leq \|\bar{x}(s)\|_2 \text{ and } \left| \int_S \bar{x}(s)\zeta_j^{(m)}(s)ds \right| \leq \|\bar{x}(s)\|_2,$$

where $\|\bar{x}(s)\|_2 = (\int_S \bar{x}^2(s)ds)^{1/2}$. This implies that a sensible way to scale $x_i(s)$ is to multiply $x_i(s)$ by a constant such that $\|\bar{x}_i(s)\|_2$ has comparable size to the scalar predictors.

3. As for most iterative algorithms, an appropriate choice of initial values is critical for the convergence of the algorithm. We use the FPCA estimates as the initial values for the proposed FPLS algorithm. The key to finding the FPCA estimates is estimating the eigenfunctions. Unlike Li and Luo (2017), which advocates the `fpca.sc` function in the `refund` package (Goldsmith *et al.*, 2018) or `fpca.mle` and `fpca.score` functions in the `fpca` package (Peng and Paul, 2011) in R, we use the method based on the singular value decomposition (SVD) proposed by Zippunikov *et al.* (2011). The reason is that the two R

packages advocated by Li and Luo (2017) can be computationally intensive with ultrahigh-dimensional images. For example, to perform FPCA for a data set with 200 subjects, each having a 300×300 image (the simulation study in Section 4), `fpca.sc` and `fpca.mle` require at least 60 GB of RAM memory and take hours to run, whereas the SVD method only requires less than 6 GB of RAM memory and finishes in seconds. Once the eigenfunctions are estimated, one can follow the method in Li and Luo (2017) to obtain the FPCA estimators.

4. We also suggest a step of stochastic approximation to accelerate the convergence. Specifically, instead of using (17) to obtain the updates $b_{0,p_0}^{(m+1)}(s)$ and $b_{1,p_1}^{(m+1)}(s)$, we consider $b_{0,p_0}^{(m+1)}(s) = (1 - a^{(m)})b_{0,p_0}^{(m)}(s) + a^{(m)} \sum_{j=1}^{p_0} b_{0,j}^{(m+1)}\psi_j^{(m)}(s)$ and $b_{1,p_1}^{(m+1)}(s) = (1 - a^{(m)})b_{1,p_1}^{(m)}(s) + a^{(m)} \sum_{j=1}^{p_1} b_{1,j}^{(m+1)}\zeta_j^{(m)}(s)$. Here, $a^{(m)}$ is called the step size at the m th iteration. As demonstrated in Walk (1978), Yin and Zhu (1990), and the references therein, flexible step sizes can accelerate the convergence of the algorithm as well as stabilize the performance. In our numerical studies, we use $a^{(m)} = 1/m$.

To apply the proposed FPLS algorithm to 2D or 3D images, one can vectorize them in any way as the proposed algorithm is invariant with respect to arbitrary vectorization of the images. In our real data analysis where each image has more than 500 000 voxels, each iteration of the proposed algorithm takes less than 1 min to run, and the entire algorithm becomes stable in a few steps.

3.2 | Choice of p_0 and p_1

So far we have discussed the algorithm with predetermined p_0 and p_1 . As demonstrated in Li and Luo (2017), changing p_0 and p_1 can have a large effect on the estimation accuracy of the parameters in FJM. Generally, smaller p_0 and p_1 may lead to a larger bias, whereas larger p_0 and p_1 may lead to a larger variance. To balance bias and variance, we use the Bayesian information criterion (BIC) to choose p_0 and p_1 . Specifically, let $\hat{\Theta}_{p_0,p_1}$, $\hat{b}_{0,p_0}(s)$, $\hat{b}_{1,p_1}(s)$, and $\hat{\Lambda}_{0,p_0,p_1}(\cdot)$ be the estimates of Θ , $b_0(s)$, $b_1(s)$, and $\Lambda_0(\cdot)$. Then, the BIC statistic is defined as

$$\text{BIC}(p_0, p_1) = \log(n)(p_0 + p_1)$$

$$-2 \int \log \left(L_{n,\text{com}} \left(\hat{\Theta}_{p_0,p_1}, \hat{b}_{0,p_0}(s), \hat{b}_{1,p_1}(s), \hat{\Lambda}_{0,p_0,p_1}(\cdot) \right) \right) du_1 \cdots u_n,$$

where $L_{n,\text{com}}(\hat{\Theta}_{p_0,p_1}, \hat{b}_{0,p_0}(s), \hat{b}_{1,p_1}(s), \hat{\Lambda}_{0,p_0,p_1}(\cdot))$ is the complete data likelihood, of which the explicit form is given in

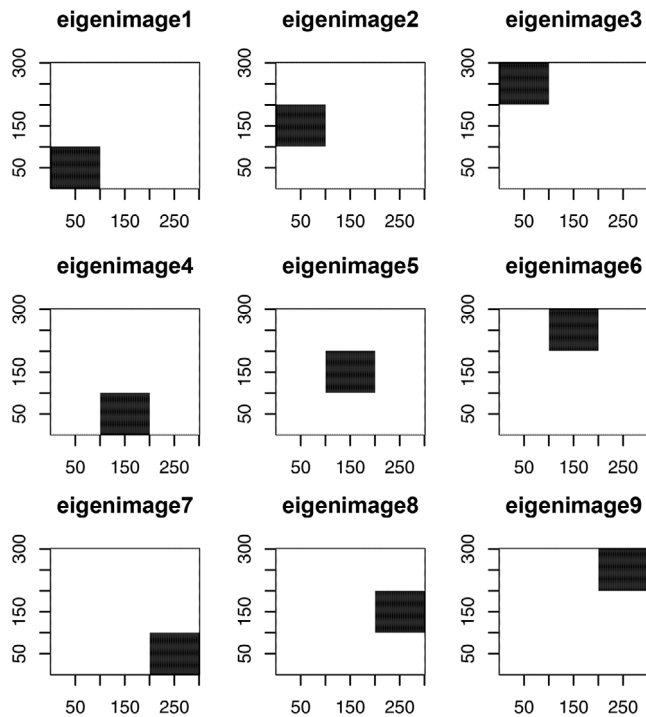


FIGURE 2 True grayscale eigenimages used in the simulation study

the Supporting Information. Numerically, we can use a grid search to find the optimal p_0 and p_1 that minimize $\text{BIC}(p_0, p_1)$.

4 | SIMULATION STUDY

In this section, we carry out a simulation study to compare the proposed FPLS algorithm with FPCA for FJM (1) and (2). For $i = 1, \dots, n$, we first independently simulated $X_i(s)$ according to the generating process

$$X_i(s) = \sum_{k=1}^9 k^{-1/4} \xi_{ik} \phi_k(s), \quad \xi_{ik} \stackrel{i.i.d.}{\sim} N(0, 1), \quad s \in \mathcal{S},$$

where the eigenimages $\phi_k(s)$ are displayed in Figure 2 and $\mathcal{S} = [1, 300] \times [1, 300]$. These eigenimages can be thought of as 2D grayscale images with pixel intensities on the $[0, 1]$ scale. The black pixels are set to 1 and the white ones are set to 0. We generated Z_i from a normal distribution with mean 0 and variance $\xi_{i2}^2/9$. This indicates that Z_i and $X_i(s)$ are correlated. We denote $m_i(t) = \beta_0 + \beta_1 t_{ij} + Z_i \beta_2 + \int_{\mathcal{S}} X_i(s) b_0(s) ds + u_i$, where $u_i \stackrel{i.i.d.}{\sim} N(0, 1)$. Then, the true event time T_i was generated based on the proportional hazards model as follows:

$$\lambda_i(t | Z_i, X_i(s), u_i) = \exp \left\{ \alpha m_i(t) + \int_{\mathcal{S}} X_i(s) b_1(s) ds + Z_i \gamma \right\}. \quad (18)$$

Next, we independently generated the censoring time C_i from a uniform distribution $U(0, c_0)$, where $c_0 > 0$ controls the censoring rate. Instead of observing both T_i and C_i , we only observed $\tilde{T}_i = \min(T_i, C_i)$ and $\Delta_i = I(T_i \leq C_i)$. Next, the longitudinal follow up time $\{t_{ij}\}_{j=1,2,3}$ was simulated from a uniform distribution in $(0, \tilde{T}_i)$, as only the measurements that are collected before the observed survival time can be used to predict the survival event. Then, the longitudinal outcome y_{ij} was generated by

$$y_i(t_{ij}) = m_i(t_{ij}) + \epsilon_{ij},$$

where $\epsilon_{ij} \stackrel{i.i.d.}{\sim} N(0, 0.4^2)$. We set $\beta_0 = 0.7, \beta_1 = 1, \beta_2 = 2, \alpha = 2$, and $\gamma = 2$, and consider two scenarios of $b_0(s)$ and $b_1(s)$: (a) $b_0(s) = b_1(s) = \sum_{k=1}^5 k^{-3/2} \phi_k(s)$ and (b) $b_0(s) = \sum_{k=5}^9 (k-4)^{-1/2} \phi_k(s), b_1(s) = \sum_{k=1}^5 k^{-1/2} \phi_k(s)$.

We consider $n = 200$ and 500 . The c_0 is selected for each scenario to yield a censoring rate of 60% that mimics our real data analysis. For each pair of n and c_0 , we generated 1000 independent data sets. For simplicity, here we consider the same number of basis functions to estimate $b_0(s)$ and $b_1(s)$, that is, $p_0 = p_1 = p$. In practice, p_0 and p_1 can be different; this will be considered in the real data analysis in Section 5. The optimal p was then selected by the BIC statistic, given in Section 3.2. For the implementation of both FPLS and FPCA, generated images $X_i(s)$'s were unfolded to obtain vectors of size $d = 300 \times 300 = 90\,000$. As mentioned in Section 3.1, our implementation of FPCA is different from that in Li and Luo (2017), and the FPLS algorithm uses the FPCA estimates as the initial values.

We examined the mean squared error (MSE) for both functional parameters according to $\text{MSE}_{b_j} = \int_{\mathcal{S}} (\hat{b}_j(s) - b_j(s))^2 ds$, where $\hat{b}_j(s)$ is the FPCA or FPLS estimator of $b_j(s)$ for $j = 0$ and 1 . Figure 3 shows results for scenario (a). In this scenario, both $b_0(s)$ and $b_1(s)$ are only informed by the top five eigenimages. The weight of the j th eigenimage, $j^{-1.5}$, decreases fast as j increases; this further favors FPCA. It can be seen that the proposed FPLS algorithm performs comparably to FPCA: FPLS yields a more accurate estimate of $b_0(s)$, whereas FPCA performs slightly better in terms of estimating $b_1(s)$. This is sensible because the IRLS procedure used for the survival model may introduce more error. In scenario (b) that does not favor FPCA, as shown in Figure 4, the proposed FPLS algorithm has considerably higher estimation accuracy than FPCA for both $b_0(s)$ and $b_1(s)$. In particular, it can be seen from Figure 4C that our FPLS algorithm yields accurate estimation of $b_0(s)$, whereas FPCA yields an MSE over 2. A straightforward calculation yields that $\int_{\mathcal{S}} b_0^2(s) ds = 2.28$, indicating that FPCA can barely capture any information from $b_0(s)$. To see the reason clearly, we first note that $b_0(s)$ lies in the span of the fifth to the ninth eigenimages. Unfortunately, these eigenimages cannot be consistently estimated in such an ultrahigh-dimensional setting ($d \gg n$). Thus, BIC tends to

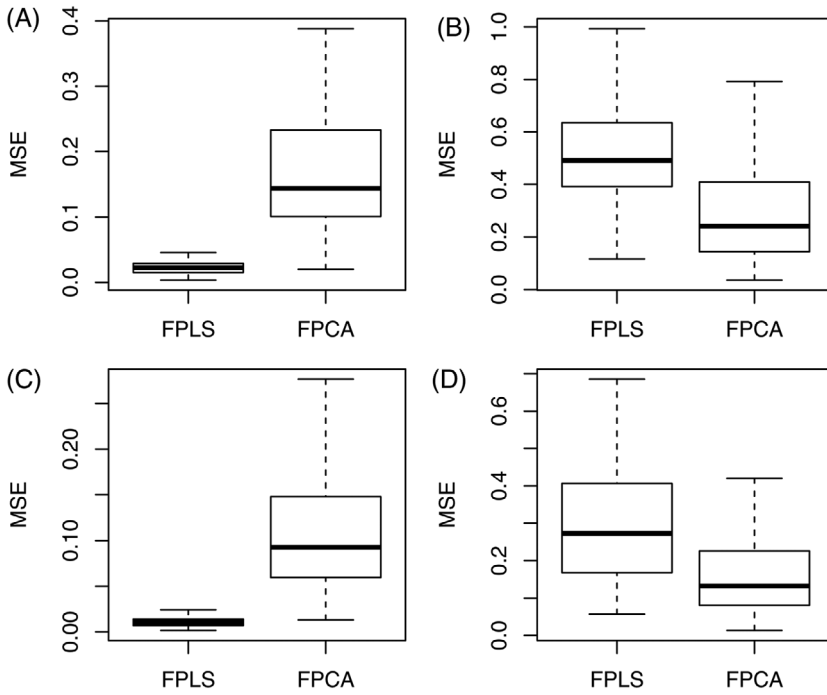


FIGURE 3 The boxplot of the MSE for scenario (a) over 1000 replications: A, $n = 200$ and MSE_{b_0} ; B, $n = 200$ and MSE_{b_1} ; C, $n = 500$ and MSE_{b_0} ; D, $n = 500$ and MSE_{b_1}

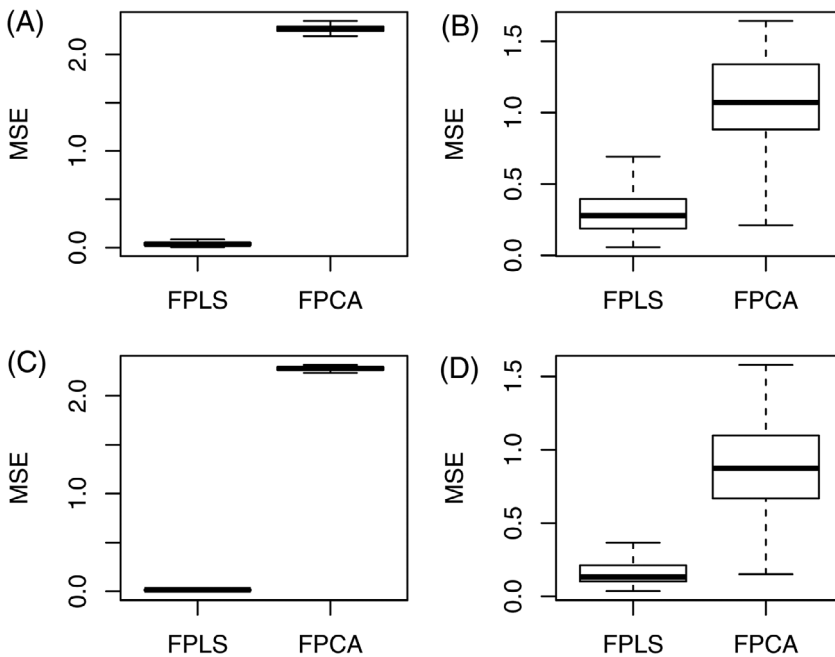


FIGURE 4 The boxplot of the MSE for scenario (b) over 1000 replications: A, $n = 200$ and MSE_{b_0} ; B, $n = 200$ and MSE_{b_1} ; C, $n = 500$ and MSE_{b_0} ; D, $n = 500$ and MSE_{b_1}

select small p ($p < 5$) for FPCA, indicating that the resulting FPCA estimator is very close to 0. In contrast, as the FPLS basis functions incorporate information from the outcome, which contains information on $b_0(s)$, the FPLS algorithm can still yield an accurate estimate of $b_0(s)$ regardless of inconsistent estimation of the eigenimages. In practice, as we never know how the functional predictor informs the outcome, the proposed FPLS algorithm may be a more robust and accurate prediction tool, especially in neuroimaging studies with high-dimensional images and limited sample size.

5 | ANALYSIS OF ADNI DATA

In this section, we jointly model the longitudinal trajectory of the ADAS-Cog score and the time of conversion from MCI to AD using the selected 236 subjects. Specifically, we consider

$$y_{ij} = m_i(t_{ij}) + \epsilon_{ij} \text{ with}$$

$$m_i(t_{ij}) = \beta_0 + \mathbf{z}_i^T \beta_1 + \beta_2 t_{ij} + \int_S x_i(s) b_0(s) ds + u_i$$

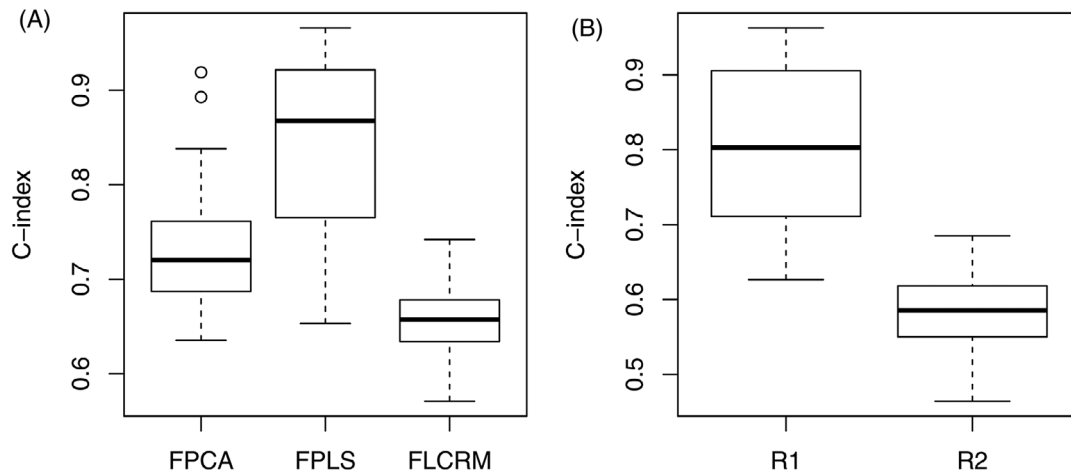


FIGURE 5 Boxplots of the C-index over 100 replications for the FPLS, FPCA, FLCRM, R1, and R2

and

$$\lambda_i(t|\mathbf{z}_i, x_i(s)) = \lambda_0(t) \exp\left(\mathbf{z}_i^T \gamma + \int_S x_i(s) b_1(s) ds + \alpha m_i(t)\right).$$

Here, t_{ij} is the follow-up time for the i th subject at the j th visit and y_{ij} is the ADAS-Cog score of the i th subject at the j th visit. The scalar covariate vector \mathbf{z}_i includes gender (1 = male; 0 = female), handedness (1 = right; 0 = left), and age at the first MCI diagnosis. The functional predictor $x_i(s)$ is the PET imaging data measured on $160 \times 160 \times 96$ voxels. PET directly measures the regional use of glucose, which indirectly reflects the brain activity of different brain regions. The PET images we used here underwent four preprocessing steps, which are introduced in detail in the Supporting Information. We also removed the background regions outside the skull, so around 500 000 voxels remained. The parameter α links the two models. If α is nonzero, then there may be an unobserved association between the longitudinal and survival outcome. The random intercept u_i is assumed to be normally distributed with mean 0 and variance σ_u^2 . Given u_i , ϵ_{ij} follows a normal distribution with mean 0 and variance σ_ϵ^2 . The implementation of FPCA and FPLS is the same as that in Section 4.

We first used cross validation to compare the proposed FPLS algorithm with FPCA in terms of prediction accuracy of the survival time. To examine the predictive value of the longitudinal ADAS-Cog scores, we also predicted the survival time based on the following functional linear Cox regression model (FLCRM; Kong *et al.*, 2018):

$$\lambda_i(t|\mathbf{z}_i, x_i(s)) = \lambda_0(t) \exp\left(\mathbf{z}_i^T \gamma + \int_S x_i(s) b_1(s) ds\right).$$

The prediction accuracy was examined according to the concordance index (C-index; Harrell *et al.*, 1996), which can be calculated using the function "concordance.index()" in the R package "survcomp" (Schroeder *et al.*, 2011). More

specifically, we randomly selected 118 subjects as the training set and the remaining 118 subjects form the test set. We repeated this procedure 100 times. For each method in each replication, we fitted the FJM using the training set with the optimal number of basis functions, that is, p_0 and p_1 , selected by BIC, and then computed the C-index using the test set. Inspecting Figure 5A shows that FPLS yields higher prediction accuracy than FPCA. Moreover, both FPLS and FPCA substantially outperform FLCRM, demonstrating that the ADAS-Cog score may be an important predictor of the progression from MCI to AD.

To examine the predictive value of PET imaging data, two reduced models are considered. The first reduced model excludes the imaging predictor from the longitudinal model, and the second one excludes the imaging predictor from the survival model, denoted by R1 and R2, respectively. Using the same training and testing data sets, we fitted these two reduced models using our FPLS algorithm with the optimal p_0 or p_1 selected by BIC and calculated the C-index. Figure 5B shows that two reduced models yield lower prediction accuracy than the FPLS in Figure 5A, demonstrating the predictive value of the PET imaging data in terms of jointly predicting ADAS-Cog scores and the progression to AD. Moreover, as we aim at predicting the survival outcome, the prediction accuracy suffers more when the imaging predictor is removed from the survival model.

Finally, we considered the complete cohort of 236 subjects to estimate the unknown parameters using our FPLS algorithm. The optimal p_0 and p_1 selected by BIC are 18 and 10, respectively. The estimated α is 2.8, indicating that patients with higher ADAS-Cog score may be more likely to progress to AD. Figure 6 displays the positive regions of the estimates of $b_0(s)$ and $b_1(s)$. It can be seen that several functional regions over the brain, such as the hippocampus, frontal lobe, and temporal horn of the lateral ventricle are identified to be positively associated with the progression from MCI to AD.

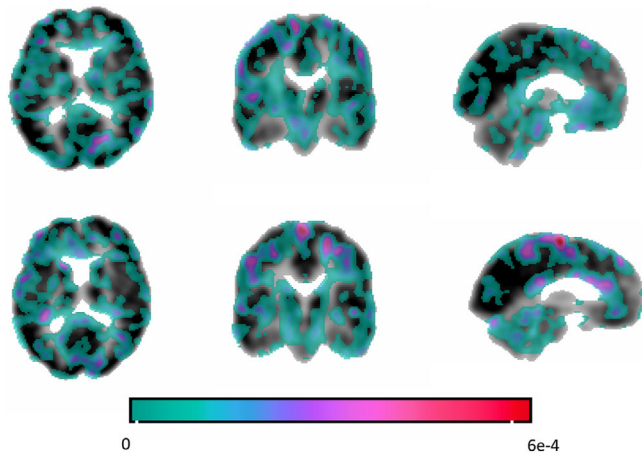


FIGURE 6 Positive regions of the estimated coefficient images obtained from the FPLS algorithm with $p_0 = 18$ and $p_1 = 10$: The top and bottom rows display the positive regions of the estimated $b_0(s)$ and $b_1(s)$, respectively. From left to right, each coefficient is displayed in the views of transverse, coronal and sagittal planes. The slices are located at (48,80,80). This figure appears in color in the electronic version of this article, and any mention of color refers to that version

6 | DISCUSSIONS

In this article, we developed a novel FPLS algorithm for the estimation and prediction of FJM. We examined its performance in simulation studies and an application to ADNI. As shown in the numerical studies, FPLS can yield comparable results to FPCA in the setting that strongly favors FPCA, whereas it yields accurate estimates in the setting where FPCA completely fails. Hence, the proposed FPLS algorithm may be a more robust and powerful prediction tool for FJM than FPCA when massive neuroimaging data are involved.

It should be noted, however, that it is very challenging to theoretically derive (asymptotic) confidence intervals of any FPLS-based estimators. The reason may be inherent to the key feature of FPLS; that is, the FPLS basis functions depend on the outcome. Consequently, the design matrix in FPLS regressions also involves the error term. Thus, bootstrapping methods have been suggested for constructing confidence intervals of PLS type of estimators (Wold *et al.*, 2001). Further studies of the inferential problems associated with the proposed FPLS algorithm may be a fruitful area of future research.

ACKNOWLEDGMENTS

Dr. Zhu's work was partially supported by NIH grants R01MH086633 and R01MH116527. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (<http://adni.loni.usc.edu>). As such, the investigators within the ADNI con-

tributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

ORCID

Yue Wang  <https://orcid.org/0000-0002-4847-8826>

Hongtu Zhu  <https://orcid.org/0000-0002-6781-2690>

REFERENCES

- Besse, P. and Ramsay, J.O. (1986) Principal components analysis of sampled functions. *Psychometrika*, 51, 285–311.
- Boente, G. and Fraiman, R. (2000) Kernel-based functional principal components. *Statistics Probability Letters*, 48, 335–345.
- Cardot, H., Ferraty, F., Mas, A. and Sarda, P. (2003) Testing hypotheses in the functional linear model. *Scandinavian Journal of Statistics*, 30, 241–255.
- Cardot, H., Ferraty, F. and Sarda, P. (1999) Functional linear model. *Statistics Probability Letters*, 45, 11–22.
- Delaigle, A. and Hall, P. (2012) Methodology and theory for partial least squares applied to functional data. *The Annals of Statistics*, 34, 2159–2179.
- Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39, 1–38.
- Du, A.T., Schuff, N., Amend, D., Laakso, M.P., Hsu, Y.Y., Jagust, W.J., Yaffe, K., Kramer, J.H., Reed, B., Norman, D., Chui, H.C. and Weiner, M.W. (2001) Magnetic resonance imaging of the entorhinal cortex and hippocampus in mild cognitive impairment and Alzheimer's disease. *The Journal of Neurology, Neurosurgery, and Psychiatry*, 71, 441–447.
- Goldsmith, J., Scheipl, F., Huang, L., Wrobel, J., Gellar, J., Harezlak, J., McLean, M.W., Swihart, B., Xiao, L., Crainiceanu, C. and Reiss, P.T. (2018) *refund: Regression with Functional Data*. R package version 0.1-17.
- Green, P.J. (1984) Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives. *Journal of the Royal Statistical Society. Series B*, 46, 149–192.
- Hall, P. and Hosseini-Nasab, M. (2006) On properties of functional principal components analysis. *Journal of the Royal Statistical Society, Series B*, 68, 109–126.
- Harrell, F.E., Lee, K.L. and Mark, D.B. (1996) Tutorial in biostatistics multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine*, 15, 361–387.
- James, G., Hastie, T. and Sugar, C. (2000) Principal component models for sparse functional data. *Biometrika*, 87, 587.
- Jung, S. and Marron, J.S. (2009) PCA consistency in high dimension, low sample size context. *The Annals of Statistics*, 37, 4104–4130.
- Kong, D., Ibrahim, J.G., Lee, E. and Zhu, H. (2018) FLCRM: functional linear cox regression model. *Biometrics*, 74, 109–117.
- Lee, E., Zhu, H., Kong, D., Wang, Y., Giovanello, K.S. and Ibrahim, J.G. (2015) Bflcrm: A Bayesian functional linear cox regression model for predicting time to conversion to Alzheimer's disease. *The Annals of Applied Statistics*, 9, 2153–2178.

- Li, K. and Luo, S. (2017) Functional joint model for longitudinal and time-to-event data: an application to Alzheimer's disease. *Statistics in Medicine*, 36, 3560–3572.
- Müller, H.-G. and Stadtmüller, U. (2005) Generalized functional linear models. *The Annals of Statistics*, 33, 774–805.
- Müller, H.-G. and Yao, F. (2010) Methodology and theory for partial least squares applied to functional data. *Biometrika*, 97, 791–805.
- Peng, J. and Paul, D. (2011) *fpca: Restricted MLE for Functional Principal Components Analysis*. R package version 0.2-1.
- Preda, C., Saporta, G. and Leveder, C. (2005) PLS classification of functional data. *Computational Statistics and Data Analysis*, 49, 223–235.
- Ramsay, J.O. and Dalzell, C.J. (1991) Some tools for functional data analysis (with discussion). *Journal of the Royal Statistical Society*, 53, 539–572.
- Schroeder, M.S., Culhan, A.C., Quackenbush, J. and Haibe-Kains, B. (2011) survcomp: an R/Bioconductor package for performance assessment and comparison of survival models. *Bioinformatics*, 27(22), 3206–3208.
- Tsiatis, A.A., DeGruttola, V. and Wulfsohn, M.S. (1995) Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with aids. *Journal of the American Statistical Association*, 90, 27–37.
- Walk, H. (1978) Martingales and the Robbins-Monro procedure in $D[0, 1]$. *Journal of Multivariate Analysis*, 8, 430–452.
- Wold, S., Sjöström, M. and Eriksson, L. (2001) PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58, 109–130.
- Wulfsohn, M.S. and Tsiatis, A.A. (1997) A joint model for survival and longitudinal data measured with error. *Biometrics*, 53, 330–339.
- Yao, F., Muller, H.G. and Wang, J.L. (2005) Functional linear regression analysis for longitudinal data. *The Annals of Statistics*, 33, 2873–2903.
- Yin, G. and Zhu, Y. (1990) On H-valued Robbins-Monro processes. *Journal of Multivariate Analysis*, 34, 116–140.
- Zipunnikov, V., Caffo, B., Yousem, D., Davatzikos, C., Schwartz, B. and Crainiceanu, C. (2011) Functional principal components model for high-dimensional brain imaging. *Neuroimage*, 58, 772–784.

SUPPORTING INFORMATION

Web Appendices, Tables and Figures referenced in Sections 2.4 are available with this paper at the Biometrics website on Wiley Online Library. Our code and example data for implementing the FPLS algorithm are also available at the Biometrics website on Wiley Online Library.

How to cite this article: Wang Y, Ibrahim JG, Zhu H. Partial least squares for functional joint models with applications to the Alzheimer's disease neuroimaging initiative study. *Biometrics*. 2020;1–11. <https://doi.org/10.1111/biom.13219>