Prediction of Cognitive Function via Brain Region Volumes with Applications to Alzheimer's Disease Based on Space-Factor-Guided Functional Principal Component Analysis

Shoudao Wen[1], Yi Li[2], Dehan Kong[3] and Huazhen Lin[1*]
[1]Center of Statistical Research, School of Statistics,
and New Cornerstone Science Laboratory,
Southwestern University of Finance and Economics, Chengdu, China
[2] Department of Biostatistics, University of Michigan, Ann Arbor, USA
[3] Department of Statistical Sciences, University of Toronto, Toronto, Canada

March 16, 2025

Abstract

Alzheimer's disease (AD) is a prevalent and irreversible brain disorder and early prediction of cognitive function is vital for detecting the onset. The volumes of brain regions can serve as features for predicting cognitive decline, facilitating early detection and intervention. In order to offer a comprehensive representation of brain tissue changes in AD, we employ volume density curves to investigate the relationship between brain regions and cognitive function. However, analyzing these volume curves is complex due to their highly spatial and intrinsic dependence and piecewise structure. To address these challenges, we propose Space-Factor-Guided Functional Principal Component Analysis (SF-FPCA). This method utilizes factor processes to extract low-dimensional features for intrinsic correlations among regions of interest (ROIs) and applies Functional Principal Component Analysis (FPCA) to these processes to address temporal dependence. Furthermore, by decomposing the loadings into smooth functions of spatial coordinates and a piecewise constant matrix, we identify regions exhibiting smoothness within each region while discontinuities between these regions. We apply SF-FPCA to analyze data from the Alzheimer's Disease Neuroimaging Initiative (ADNI). Our results demonstrate that SF-FPCA provides the best fit compared to other methods. In addition, features extracted from volume

*Corresponding author. Email: linhz@swufe.edu.cn.

curves using SF-FPCA enable more accurate prediction of cognitive function compared to scalar volumes alone. Leveraging these extracted features, we identify 36 important ROIs influencing cognitive decline. Our investigation into brain atrophy also reveals distinct mechanisms between the left and right hemispheres, shedding light on the nuanced effects of brain region changes on cognitive decline in AD.

Key words and phrases: Volume density curve; cognitive function; high-dimensional functional data; spatial correlation; factor analysis

1 Introduction

Alzheimer's disease (AD) is a prevalent and irreversible brain disorder among the elderly population, characterized by a range of behavioral deficits including loss of memory, thinking, language skills, and reasoning abilities. Given the irreversible nature of AD and its inability to be completely cured, early prediction of cognitive decline is crucial for detecting the disease onset in its initial stages. Changes in brain volume, characterized by progressive atrophy in specific structures, serve as a prominent neural correlate of AD (Talwar et al., 2021). As AD progresses, neuronal damage and death lead to alterations in brain volume, particularly affecting regions associated with cognitive function and memory (Fjell and Walhovd, 2010; Albert et al., 2011).

The volumes of regions of interest (ROIs) can directly serve as features for predicting cognitive decline in AD patients (de Jong et al., 2008; Ferrarini et al., 2008; Henneman et al., 2009; Evans et al., 2010), aiding in the early detection of the disease and the implementation of appropriate intervention measures. Furthermore, studying volume changes in specific ROIs provides valuable insights into the pathophysiology of AD, shedding light on the underlying mechanisms involved in disease progression and facilitating the identification of potential biomarkers for monitoring disease progression.

Numerous studies have demonstrated the association between changes in ROI volumes and AD. The atrophy of the hippocampus, in particular, has been extensively investigated and is widely recognized as one of the most established biomarkers for AD (Likeman et al., 2005; Leung et al., 2010; Antoniano-Villalobos et al., 2014). Other brain regions, such as the temporal lobe, parietal lobe, and cingulate gyrus, have also been implicated in AD pathology (Chincarini et al., 2014; Jacobs et al., 2011; Barnes et al., 2007). In this study, our primary objective is to identify ROIs that are associated with cognitive function by analyzing the volume of these regions, aiming to enhance the prediction of cognitive de-

cline. We utilize the most recently updated data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) study, which includes magnetic resonance imaging (MRI) data and mini-mental state examination (MMSE) scores. Following appropriate preprocessing steps, detailed in Section S4, we convert the MRI data into volume density curves. These curves, representing the volumes of ROIs, can be treated as high-dimensional functional variables in our analysis.

While using scalar volume measurements to assess the impact of ROI volume changes in AD offers a direct approach, leveraging density curves instead provides several advantages. Firstly, density curves offer a more comprehensive representation of changes in brain tissue associated with AD. Secondly, density curves offer a greater number of observation points for analysis, thereby increasing sensitivity to AD-related changes. Density curves capture variations in tissue density across different brain regions, and several studies (Zhong et al., 2021; Li et al., 2022, 2023) have utilized density curve analysis to investigate AD. In our study, we compare the predictive performance of regression models for cognitive function using high-dimensional volume curves and scalar volumes. The results illustrated in Figure 2 demonstrate that the regression model constructed with features extracted from volume curves as covariates outperforms the model based on scalar volumes.

Analyzing the volume density data presents several challenges. Firstly, we need to characterize the intrinsic and spatial correlations among a large number of ROIs. A reasonable assumption is that ROIs located closer to each other tend to exhibit greater similarity. However, the brain comprises spatially contiguous regions, each consisting of various ROIs. These regions may differ in cortical thickness, neuron density, connectivity patterns, and functional expressions. The second challenge is to retain the distinctions between different regions while accounting for spatial correlations. This requires our analysis to capture both the similarities within regions and the differences between them. Similar challenges have been discussed in previous neuroimaging studies, such as those highlighted by Zhu

et al. (2014). Recently, methods for handling high-dimensional functional data have been categorized into three main approaches: univariate functional principal component analysis (uFPCA-HD), sparse functional principal component analysis (sFPCA-HD), and factor models (FM-HD). uFPCA-HD (Gao et al., 2019) involves applying uFPCA separately to each functional variable, largely overlooking the correlation among the functional variables. Hu and Yao (2022) proposed sFPCA, which assumes that only a small number of curves are important. However, this assumption may not hold true when variables are correlated, leading to biased results. On the other hand, FM-HD methods do account for inter-variable correlations but often ignore temporal dependence. For instance, Hu and Yao (2024) proposed dynamic principal component analysis (DPCA), which is based on standard factor analysis at each time point. Similarly, Hallin et al. (2022) and Guo et al. (2022) extended the factor model by treating curves as elements in Hilbert space.

When dealing with high-dimensional functions that exhibit spatial correlation, several methods have been developed to incorporate spatial information: Zhang et al. (2016) introduced a functional conditional autoregressive (CAR) model tailored for spatially correlated functional data, considering spatial correlation within the covariance structure; Marinescu et al. (2019) proposed Data-driven Inference of Vertexwise Evolution (DIVE), assuming the existence of latent variables for each ROI to cluster the curves. However, these methods rely solely on the adjacency matrix of spatial locations, neglecting precise spatial coordinates and thus not fully leveraging spatial information. Another common approach is spatio-temporal FPCA. For example, Hyun et al. (2016) developed FPCA for spatio-temporal Gaussian processes (FPCA-ST) by extending one-dimensional eigenfunctions over time to multi-dimensional eigenfunctions over space and time. Similarly, Chen et al. (2017) proposed product FPCA (pFPCA), utilizing the product of eigenfunctions in the space and time domains. Both FPCA-ST and pFPCA assume the smoothness of all functional variables across spatial coordinates, disregarding differences between different

brain regions.

We propose a Space-Factor-Guided Functional Principal Component Analysis (SF-FPCA) approach to effectively capture the intrinsic and spatial correlation and piecewise smoothness of the ROI volume curves. The fundamental concept behind SF-FPCA is to employ factor processes to organize the intrinsic correlations of ROI measures. Subsequently, we apply univariate Functional Principal Component Analysis (uFPCA) to each component of these factor processes to address temporal dependence and further extract features. Our approach offers several novel aspects compared to existing methods. In uFPCA-HD, applying FPCA to each functional variable can lead to a large number of duplicate features due to inter-variable correlation, while some features related to this correlation are easily ignored. In contrast, SF-FPCA efficiently utilizes correlation information among ROIs by employing factor processes, thereby avoiding overlapping features. By applying FPCA to the factor processes, SF-FPCA also captures temporal dependence more effectively than FM-HD and can further extract features suitable for subsequent regression analysis. Furthermore, by decomposing the loadings into smooth functions of spatial coordinates and a piecewise constant matrix, SF-FPCA incorporates spatial coordinates to describe the relationship between ROI correlations and spatial information, distinguishing it from other methods. Compared to DIVE and the functional CAR model, SF-FPCA fully leverages available spatial information and provides a more detailed modeling of spatial correlation. Additionally, compared to pFPCA and FPCA-ST, SF-FPCA considers both the smooth and discrete information of spatial locations, accounting for differences between different brain regions and enabling a more comprehensive analysis.

Our analysis results show that predicting cognitive function based on volume curves outperforms predictions based on scalar volume (see Figure 2), underscoring the advantages of functional data analysis. Moreover, SF-FPCA achieves higher predictive accuracy for fitting high-dimensional functional curves and MMSE compared to state-of-the-art compet-

ing methods (see Tables 1 and 2). This suggests that the features extracted by SF-FPCA most effectively represent the volume of ROI. Based on the extracted features from SF-FPCA, we identify 36 ROIs associated with cognitive decline. We validate some critical regions affecting AD, such as the left hippocampus and certain ROIs in the temporal and frontal lobes. Additionally, we investigate the impact of cerebellar atrophy on cognitive decline. Specifically, we uncover distinct impact mechanisms of different locations within cerebellar vermal lobules on cognitive decline. Furthermore, we find that the mechanisms differ between the left and right hemispheres and identify several ROIs that have received limited or no research attention previously.

In Section 2, we introduce the SF-FPCA model, while the estimating method is discussed in Section 3. Section 4 applies the proposed method to analyze the ADNI data and presents the scientific findings and interpretations. Section 5 further evaluates the utility of the method via extensive simulations, and Section 6 concludes the paper with a brief remark. The details on the data source and preprocessing steps, along with some simulation and real data analysis results, are included in the supplementary materials.

2 Model

Suppose that the observations $\mathbf{X}_i(\cdot)$ $(i = 1, \cdots, n)$ are independent and identically distributed with $\mathbf{X}(t)$, where $\mathbf{X}(t) = \{X_1(t), \cdots, X_p(t)\}'$ being the high-dimensional functional variables. Particularly, in the ADNI data, $X_j(t)$ is the brain local volumetric measure at quantile $t \in [0,\ 1]$ of ROI $j = 1, \ldots, p$ with $p = 95$. The high-dimensional feature is reflected in the number of volume curves, with a detailed introduction to ADNI provided in Supplementary Material C. For simplicity, we assume that $\mathbb{E}\{X_j(t)\} = 0$ for any j. Figure 1 in Section 4 plots the heat map and the first 15 eigenvalues of the numerical approximation version of $\int \mathbb{E}\{\mathbf{X}(t)\mathbf{X}'(t)\}dt$, revealing high correlations among different ROIs and

indicating an obvious low-rank structure in the data. Following the idea of factor analysis (Bai and Ng, 2013), we assume that $\mathbf{X}(t)$ are correlated due to a shared vector of factor processes $\mathbf{h}(t)$, given by:

$$\mathbf{X}(t) = \mathbf{B}\mathbf{h}(t) + \mathbf{u}(t), \tag{1}$$

where $\mathbf{h}(t) = \{h_1(t), \cdots, h_q(t)\}'$ is a q-dimensional vector of factor processes with $q \ll p$, $\mathbf{B} = (\mathbf{b}_1, \cdots, \mathbf{b}_p)' = (b_{jk})_{p\times q}$ is a deterministic loading matrix, and $\mathbf{u}(t) = \{u_1(t), \cdots, u_p(t)\}'$ denotes the measurement error, which is independent of $\mathbf{h}(t)$ with $\mathbb{E}\{\mathbf{u}(t)\} = \mathbf{0}$ and $\text{var}\{\mathbf{u}(t)\} = \sigma^2\mathbf{I}_p$.

However, Model (1) only captures correlations among ROIs and does not incorporate spatial coordinates, thus neglecting potential spatial information. To address this limitation, we propose a modification. Considering the 3-dimensional (3D) coordinates $\mathbf{s}_j \in [0,1]^3$ of the center of ROI j, we extend Model (1) to:

$$\mathbf{X}(t) = (\mathbf{B} + \mathbf{f})\mathbf{h}(t) + \mathbf{u}(t), \tag{2}$$

where $\mathbf{f} = \{\mathbf{f}(\mathbf{s}_1), \cdots, \mathbf{f}(\mathbf{s}_p)\}'$ with $\mathbf{f}(\cdot) = \{f_1(\cdot), \cdots, f_q(\cdot)\}'$ denotes smooth functions of the spatial coordinates, and $\mathbf{B}$ is a matrix accounting for the extra non-smooth information. Model (2) explains inter-variable and spatial correlations between ROIs through low-rank and smooth structures. Furthermore, to capture differences between different regions, we impose a piecewise constant structure on $\mathbf{B}$. Particularly, we assume $\mathbf{b}_j = \mathbf{b}_{j'}$ if ROIs j and j' belong to the same piece, while the number and composition of the pieces are allowed to be unknown. In fact, model (2) is highly flexible. Particularly, when $\mathbf{B} = \mathbf{0}$, the model captures cases where loadings vary smoothly over space; when $\mathbf{f} = \mathbf{0}$, it reduces to a factor model, ignoring spatial coordinates and spatial information. Since $\mathbf{B}$ is constant within the same piece due to the piecewise structure, the loadings are smooth within each piece. In addition, the distinct piecewise constants reflect the differences among the different pieces.

Thus, $\mathbf{B}$ and $\mathbf{f}(\cdot)$ jointly consider the piecewise smoothness, which simultaneously reflect spatial correlation and differences between different regions. Additionally, we allow each nonzero $\mathbf{b}_j$ to be different, indicating that the piecewise constant is not necessary. It serves only for interpretability, not as a constraint.

In order to further extract low-dimensional features, we apply the Karhunen-Loève expansion (Ash and Gardner, 1975) to the factor processes $h_j(t)$ $(j = 1, \cdots, q)$, yielding:

$$h_j(t) = \sum_{k=1}^{\infty} \xi_{jk}\phi_{jk}(t),$$

where $\phi_{jk}(\cdot)$ is the k-th orthogonal eigenfunction of the covariance function $\mathrm{cov}\{h_j(s), h_j(t)\}$ satisfying $\int \phi_{jk}(t)\phi_{jk'}(t)dt = 1_{\{k=k'\}}$; ξ_{jk} represents the functional principal component score with $\mathbb{E}(\xi_{jk}) = 0$ and $\mathrm{cov}(\xi_{jk}, \xi_{jk'}) = \rho_{jk}1_{\{k=k'\}}$; and ρ_{jk} denotes the k-th eigenvalue corresponding to $\mathrm{cov}\{h_j(s), h_j(t)\}$.

Under the constraints $\rho_{j1} \geqslant \rho_{j2} \geqslant \cdots > 0$ and $\sum_{k=1}^{\infty} \rho_{jk} < \infty$, which imply that $\sup_{t\in[0,1]} \mathbb{E}\{\sum_{k=1}^{K} \xi_{jk}\phi_{jk}(t) - \sum_{k=1}^{\infty} \xi_{jk}\phi_{jk}(t)\}^2 \to 0$ as K is large enough, we hence assume that $h_j(t)$ has the truncated form as,

$$h_j(t) = \sum_{k=1}^{K} \xi_{jk}\phi_{jk}(t). \tag{3}$$

Denote $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_1', \cdots, \boldsymbol{\zeta}_q')'$ with $\boldsymbol{\zeta}_j = (\xi_{j1}, \cdots, \xi_{jK})'$ and $\boldsymbol{\Phi}(t) = \mathrm{diag}\{\boldsymbol{\Phi}_1(t), \cdots, \boldsymbol{\Phi}_q(t)\}$ with block j being $\boldsymbol{\Phi}_j(t) = \{\phi_{j1}(t), \cdots, \phi_{jK}(t)\}'$. Coupled with (3), model (2) can be rewritten as

$$\mathbf{X}(t) = (\mathbf{B} + \mathbf{f})\boldsymbol{\Phi}'(t)\boldsymbol{\zeta} + \mathbf{u}(t). \tag{4}$$

Model (4) is termed a *space-factor-guided functional principal component analysis (SF-FPCA)* model. The proposed SF-FPCA model (4) naturally incorporates the intrinsic and spatial correlations among ROI volume curves across quantiles t, facilitating a simple and

sufficient representation of spatially correlated high-dimensional functional data. Particularly, by decomposing the loadings into smooth functions of spatial coordinates $\mathbf{f}(\mathbf{s})$ and the piecewise constant matrix $\mathbf{B}$, we account for the existence of multiple piecewise smooth regions with discontinuities and significant spatial correlations. The score $\boldsymbol{\zeta}$, representing low-dimensional features derived from high-dimensional volume curves, can be directly utilized for regression modeling of MMSE. By applying the transformation in Section 4, we can identify the ROIs that are relevant to cognitive decline based on the estimated scores. In this context, SF-FPCA specifies the inter-variable and spatial correlations among functional variables and the temporal dependence. Compared to state-of-the-art methods, SF-FPCA makes full use of spatial coordinate information and accounts for the piecewise smoothness in volume curves with respect to spatial coordinates within ROIs.

However, obtaining identifiability for model (4) is challenging and may not be easily achievable. To ensure identifiability, we introduce assumptions regarding the smoothness of $\mathbf{f}(\cdot)$ by assuming it follows a truncated basis function expansion approximation:

$$\mathbf{f}(\cdot) = \left\{ \sum_{k=1}^{\tau} \eta_{1k} M_k(\cdot), \cdots, \sum_{k=1}^{\tau} \eta_{qk} M_k(\cdot) \right\}' = \left\{ \boldsymbol{\eta}_1' \mathbf{M}(\cdot), \cdots, \boldsymbol{\eta}_q' \mathbf{M}(\cdot) \right\}' = \boldsymbol{\eta}' \mathbf{M}(\cdot),$$

where $\boldsymbol{\eta} = (\boldsymbol{\eta}_1, \cdots, \boldsymbol{\eta}_q)$ represents the coefficient matrix, and $\mathbf{M}(\cdot) = \{M_1(\cdot), \cdots, M_\tau(\cdot)\}'$ are pre-defined smooth basis functions. Denoting $\mathbf{M} = \{\mathbf{M}(\mathbf{s}_1), \cdots, \mathbf{M}(\mathbf{s}_p)\}'$, model (4) is modified to

$$\mathbf{X}(t) = (\mathbf{B} + \mathbf{M}\boldsymbol{\eta})\boldsymbol{\Phi}'(t)\boldsymbol{\zeta} + \mathbf{u}(t). \tag{5}$$

We will demonstrate that $\boldsymbol{\eta}$ is exactly identifiable, thereby ensuring the identifiability of $\mathbf{f}(\cdot)$ up to basis approximation. Let $\lfloor x \rfloor$ denote the largest integer not exceeding x for any real number $x \geqslant 0$. For an index set $\mathcal{C}$ and matrix $\mathbf{A}$, $|\mathcal{C}|$ represents the cardinality of $\mathcal{C}$, $\mathbf{A}_{\cdot j}$ denotes the j-th column of $\mathbf{A}$, and $\mathbf{A}_{\mathcal{C}}$ refers to the corresponding rows of $\mathbf{A}$. To guarantee the identifiability of model (5), we impose the following constraints:

(I1) $\max_{j\in\{1,\cdots,q\}} \|\mathbf{B}_{\cdot j}\|_0 \leqslant \lfloor (p-m)/2q \rfloor$, $\text{rank}(\mathbf{M}_{\mathcal{C}}) = \tau$, $\forall \mathcal{C} \subseteq \{1,\cdots,p\}$ with $|\mathcal{C}| = m$ for some integer m satisfying $q \leqslant \tau \leqslant m \leqslant p$.

(I2) $\boldsymbol{\eta}'\boldsymbol{\eta} = \mathbf{I}_q$ and $\eta_{j1} > 0$ $(j = 1,\cdots,q)$.

(I3) $\mathbb{E}(\boldsymbol{\zeta}\boldsymbol{\zeta}')$ is diagonal with $\sum_{k=1}^{K} \text{var}(\xi_{1k}) \geqslant \cdots \geqslant \sum_{k=1}^{K} \text{var}(\xi_{qk}) > 0$ and $\text{var}(\xi_{j1}) \geqslant \cdots \geqslant \text{var}(\xi_{jK}) > 0$ $(j = 1,\cdots,q)$.

(I4) $\int \boldsymbol{\Phi}(t)\boldsymbol{\Phi}'(t)dt = \mathbf{I}_{Kq}$ and $\phi_{jk}(0) > 0$ $(j = 1,\cdots,q; k = 1,\cdots,K)$.

Condition (I1) requires the rows of $\mathbf{B}$ to be sparse, aligning with similar requirements in Wang et al. (2017) and Bing et al. (2024). Conditions (I2)–(I4) impose restrictions on the loadings, scores, and eigenfunctions, which are common in the literature on factor analysis (Bai and Ng, 2013; Fan et al., 2013; Jiang et al., 2019; Liu et al., 2023) and FPCA (James et al., 2000; Zhou et al., 2018). It is important to note that if Conditions (I2)-(I4) are not imposed, then model (5) remains valid when $\mathbf{B}$, $\boldsymbol{\eta}$, $\boldsymbol{\Phi}'(t)$, and $\boldsymbol{\zeta}$ are replaced by $\mathbf{BA}$, $\boldsymbol{\eta}\mathbf{A}$, $\mathbf{A}'\boldsymbol{\Phi}'(t)\mathbf{D}$, and $\mathbf{D}'\boldsymbol{\zeta}$, respectively, for any orthogonal matrices $\mathbf{A}$ and $\mathbf{D}$. Hence, the orthogonal operations on $\boldsymbol{\eta}$, $\boldsymbol{\zeta}$, and $\boldsymbol{\Phi}(t)$ serve only to ensure the identifiability of $\boldsymbol{\eta}$, $\boldsymbol{\Phi}(t)$, and $\boldsymbol{\zeta}$, rather than as a restriction. The identifiability of the model (5) is stated below, with the proof provided in Supplementary Material A.

Theorem 1. *Under the conditions (I1)-(I4),* $\mathbf{B}, \boldsymbol{\eta}, \boldsymbol{\Phi}(t)$ *and* $\boldsymbol{\zeta}$ *in model (5) are identifiable.*

3 Estimation Procedure

We have the observations $\mathbf{X}_i(t)$ $(i = 1,\cdots,n)$. For possibly irregular and subject-specific time points, we assume that $\mathbf{X}_i(\cdot)$ are measured at $\mathbf{t}_i = (t_{i1},\cdots,t_{in_i})'$. In ADNI data analysis, t represents quantiles. Let $\{\mathbf{h}_i(t), \boldsymbol{\zeta}_i\}$ $(i = 1,\cdots,n)$ be iid copies of $\{\mathbf{h}(t), \boldsymbol{\zeta}\}$. We propose an iterative algorithm to estimate $\mathbf{B}, \boldsymbol{\eta}, \boldsymbol{\Phi}(t)$, and $\boldsymbol{\zeta}_i$ in model (5). Denote

the estimate from the rth step by superscript (r). To initiate the iteration, we start by computing initial values.

3.1 Initial Value

Denote $\mathbf{C} = \mathbf{B} + \mathbf{f} = \{b_{jk} + f_k(\mathbf{s}_j)\}_{p\times q}$, and the covariance matrices of $\mathbf{X}_i(t)$ and $\mathbf{u}_i(t)$ by $\mathbf{\Sigma}_{\mathbf{X}}(t)$ and $\mathbf{\Sigma}_{\mathbf{u}}(t)$, respectively. Let $\widetilde{\mathbf{\Sigma}}_{\mathbf{X}} = \int \mathbf{\Sigma}_{\mathbf{X}}(t)dt$, $\widetilde{\mathbf{\Sigma}}_{\mathbf{u}} = \int \mathbf{\Sigma}_{\mathbf{u}}(t)dt$ and $\mathbf{\Lambda}_{\boldsymbol{\zeta}} = \text{diag}(\sum_{k=1}^{K}\rho_{1k}, \cdots, \sum_{k=1}^{K}\rho_{qk})$. By calculating the covariance matrix and integrating over t on both sides of (2), we have

$$\widetilde{\mathbf{\Sigma}}_{\mathbf{X}} = \mathbf{C}\mathbf{\Lambda}_{\boldsymbol{\zeta}}\mathbf{C}' + \widetilde{\mathbf{\Sigma}}_{\mathbf{u}}, \tag{6}$$

which implies that $p^{-1/2}\mathbf{C}$ is the eigenvector of $\widetilde{\mathbf{\Sigma}}_{\mathbf{X}}$. Following Bai and Ng (2013), we then estimate the initial values of $p^{-1/2}\mathbf{C}$ by the orthogonal eigenvectors of the numerical approximation version of $\widetilde{\mathbf{\Sigma}}_{\mathbf{X}}$; i.e.,

$$\mathbf{C}^{(0)} = p^{1/2}\mathbf{E}_{eigen}\left\{n^{-1}\sum_{i=1}^{n} n_i^{-1}\sum_{l=1}^{n_i}\mathbf{X}_i(t_{il})\mathbf{X}_i'(t_{il}); q\right\}, \tag{7}$$

where $\mathbf{E}_{eigen}(\mathbf{A}; q)$ is a matrix composed of the orthogonal eigenvectors corresponding to the q largest eigenvalues of matrix $\mathbf{A}$. Obviously, $p^{-1}\mathbf{C}^{(0)\prime}\mathbf{C}^{(0)} = \mathbf{I}_q$.

Notice that $\mathbf{C} = \mathbf{B} + \mathbf{f} = \mathbf{B} + \mathbf{M}\boldsymbol{\eta}$, we then separate $\mathbf{f} = \mathbf{M}\boldsymbol{\eta}$ from $\mathbf{C}$ by performing linear regression of $\mathbf{C}_{\cdot k}$ $(k = 1, \cdots, q)$ on $\mathbf{M}$, respectively. This yields the initial values of $\boldsymbol{\eta}$, $\mathbf{f}(\mathbf{s})$ and $\mathbf{B}$ as

$$\boldsymbol{\eta}^{(0)} = (\mathbf{M}'\mathbf{M})^{-1}\mathbf{M}'\mathbf{C}^{(0)},\ \mathbf{f}^{(0)}(\mathbf{s}) = \boldsymbol{\eta}^{(0)\prime}\mathbf{M}(\mathbf{s}) \text{ and } \mathbf{B}^{(0)} = \mathbf{C}^{(0)} - \mathbf{M}\boldsymbol{\eta}^{(0)}. \tag{8}$$

Finally, by multiplying $(\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}'$ on both sides of (2), we have

$$\mathbf{h}_i(t) \approx (\mathbf{C}'\mathbf{C})^{-1}\mathbf{C}'\mathbf{X}_i(t). \tag{9}$$

Since $p^{-1}\mathbf{C}^{(0)\prime}\mathbf{C}^{(0)} = \mathbf{I}_q$, we obtain the initial values of $\mathbf{h}_i(t)$ $(i = 1, \cdots, n)$ by

$$\mathbf{h}_i^{(0)}(t) = p^{-1}\mathbf{C}^{(0)\prime}\mathbf{X}_i(t). \tag{10}$$

3.2 Iterative Step

In the $(r+1)$-th step, we firstly update $\mathbf{B}$ by the following penalized least square loss:

$$\mathbf{B}^{(r+1)} = \text{argmin}_{\mathbf{B}} \sum_{j=1}^{p}\sum_{i=1}^{n}\sum_{l=1}^{n_i}\left\{X_{ij}(t_{il}) - \mathbf{f}^{(r)\prime}(\mathbf{s}_j)\mathbf{h}_i^{(r)}(t_{il}) - \mathbf{b}_j'\mathbf{h}_i^{(r)}(t_{il})\right\}^2 + \lambda_1\sum_{j=1}^{p}\|\mathbf{b}_j\|_2 + \lambda_2\sum_{j'<j=1}^{p} w(\|\mathbf{s}_j - \mathbf{s}_{j'}\|_2)\cdot\|\mathbf{b}_j - \mathbf{b}_{j'}\|_2. \quad (11)$$

The penalty $w(\|\mathbf{s}_j - \mathbf{s}_{j'}\|_2)\cdot\|\mathbf{b}_j - \mathbf{b}_{j'}\|_2$ is used to identify different pieces. The weight $w(\|\mathbf{s}_j - \mathbf{s}_{j'}\|_2)$ decreases as $\|\mathbf{s}_j - \mathbf{s}_{j'}\|_2$ increases, reflecting that ROIs closer to each other are more likely to belong to the same piece. A similar idea can be found in Li and Sang (2019). A simple choice for $w(\|\mathbf{s}_j - \mathbf{s}_{j'}\|_2)$ is $1/\|\mathbf{s}_j - \mathbf{s}_{j'}\|_2$. We apply the ADMM algorithm to solve (11) and optimize the parameters λ_1 and λ_2 by minimizing normalized prediction error (PE), defined in Section 4, through a 5-fold cross-validation. The calculation process of ADMM and the cross-validation procedure are elaborated in Supplementary Material B.

By the equation (5), we have $\mathbf{X}_i(t) - \mathbf{B}\mathbf{h}_i(t) \approx \mathbf{M}\boldsymbol{\eta}\mathbf{h}_i(t)$. Then, by left-multiplyling $\mathbf{M}'$ and right-multiplyling $\mathbf{h}_i'(t_{il})$ on both sides and making the summation over i, we update $\boldsymbol{\eta}$ by

$$\boldsymbol{\eta}^{(r+1)} = (\mathbf{M}'\mathbf{M})^{-1}\mathbf{M}'\left[\sum_{i=1}^{n}\sum_{l=1}^{n_i}\left\{\mathbf{X}_i(t_{il}) - \mathbf{B}^{(r+1)}\mathbf{h}_i^{(r)}(t_{il})\right\}\mathbf{h}_i^{(r)\prime}(t_{il})\right]\times\left\{\sum_{i=1}^{n}\sum_{l=1}^{n_i}\mathbf{h}_i^{(r)}(t_{il})\mathbf{h}_i^{(r)\prime}(t_{il})\right\}^{-1}. \quad (12)$$

After obtaining the updated estimators of $\boldsymbol{\eta}$ for each step, we perform an orthogonal transformation to essure $\boldsymbol{\eta}^{(r+1)\prime}\boldsymbol{\eta}^{(r+1)} = \mathbf{I}_q$. We then obtain $\mathbf{f}^{(r+1)}(\mathbf{s}) = \boldsymbol{\eta}^{(r+1)\prime}\mathbf{M}(\mathbf{s})$.

Given $\mathbf{f}^{(r+1)}(\mathbf{s})$ and $\mathbf{B}^{(r+1)}$, following (9), we update $\mathbf{h}_i(t)$ by

$$\mathbf{h}_i^{(r+1)}(t) = (\mathbf{C}^{(r+1)\prime}\mathbf{C}^{(r+1)})^{-1}\mathbf{C}^{(r+1)\prime}\mathbf{X}_i(t)$$

with $\mathbf{C}^{(r+1)} = \mathbf{B}^{(r+1)} + \mathbf{f}^{(r+1)}$. Now, we consider updating the scores $\boldsymbol{\zeta}$. Let $\mathbf{Z}(\cdot) = \{(Z_1(\cdot), \cdots, Z_\omega(\cdot)\}'$ denote a vector of B-spline basis functions on $[0,\ 1]$, then we have $\phi_{jk}(\cdot) \approx \boldsymbol{\Theta}_{jk}'\mathbf{Z}(\cdot)$. Denote $\boldsymbol{\Theta}_j = (\boldsymbol{\Theta}_{j1}, \cdots, \boldsymbol{\Theta}_{jK})'$. With the B-spline approximation, identification condition (I4) on $\boldsymbol{\Phi}(t)$ can be written as $\omega^{-1}\boldsymbol{\Theta}_j\boldsymbol{\Theta}_j' = \mathbf{I}_K$ $(j = 1, \cdots, q)$

and $\omega \int \mathbf{Z}(t)\mathbf{Z}'(t)dt = \mathbf{I}_\omega$. In practice, we firstly generate B-spline basis functions at fine grid points $t_1, \cdots, t_{n_{grid}}$ as $\mathbf{Z}_{grid} = \{\mathbf{Z}(t_1), \cdots, \mathbf{Z}(t_{n_{grid}})\}'$ using the *bsplineS* function from the *fda* package in R. Then we apply singular value decomposition (SVD) on $\mathbf{Z}_{grid}$ as $\mathbf{Z}_{grid} = \mathbf{SVD}$ and obtain $\sqrt{\frac{n_{grid}}{\omega}}\mathbf{S}$ as the final $\mathbf{Z}_{grid}$ to ensure $\frac{\omega}{n_{grid}}\mathbf{Z}'_{grid}\mathbf{Z}_{grid} = \mathbf{I}_\omega$. Moreover, by the equation (3), for each $j = 1, \cdots, q$, we obtain

$$h_{ij}(t_{il}) \approx \mathbf{\Phi}'_j(t_{il})\boldsymbol{\zeta}_{ij} \approx \mathbf{Z}'(t_{il})\mathbf{\Theta}'_j\boldsymbol{\zeta}_{ij}.$$

Multiplyling $\mathbf{Z}(t_{il})$ on both sides and making the summation over the observations of individual i, we have

$$\left\{\sum_{l=1}^{n_i}\mathbf{Z}(t_{il})\mathbf{Z}'(t_{il})\right\}^{-1}\left\{\sum_{l=1}^{n_i}\mathbf{Z}(t_{il})h_{ij}(t_{il})\right\} \approx \mathbf{\Theta}'_j\boldsymbol{\zeta}_{ij},$$

which is a factor model with response $\mathbf{w}_{ij} = \{\sum_{l=1}^{n_i}\mathbf{Z}(t_{il})\mathbf{Z}'(t_{il})\}^{-1}\{\sum_{l=1}^{n_i}\mathbf{Z}(t_{il})h_{ij}(t_{il})\}$, factor $\boldsymbol{\zeta}_{ij}$ and loading $\mathbf{\Theta}_j$. Denote $\mathbf{W}_j = (\mathbf{w}_{1j}, \cdots, \mathbf{w}_{nj})' \in \mathbb{R}^{n\times\omega}$ and $\boldsymbol{\zeta}_{[j]} = (\boldsymbol{\zeta}_{1j}, \cdots, \boldsymbol{\zeta}_{nj})' \in \mathbb{R}^{n\times K}$, we hence update $(\boldsymbol{\zeta}_{[j]}, \mathbf{\Theta}_j)$ by

$$(\boldsymbol{\zeta}^{(r+1)}_{[j]}, \mathbf{\Theta}^{(r+1)}_j) = \underset{\boldsymbol{\zeta}_{[j]},\mathbf{\Theta}_j}{\arg\min}\left\|\mathbf{W}^{(r+1)}_j - \boldsymbol{\zeta}_{[j]}\mathbf{\Theta}_j\right\|^2_F,$$

with $\|\cdot\|_F$ being the Frobenius-norm of a matrix, where $\mathbf{W}^{(r+1)}_j$ is $\mathbf{W}_j$ with $\mathbf{h}_i(t)$ replaced by $\mathbf{h}^{(r+1)}_i(t)$. Following Bai and Ng (2013), which states that the factors can be estimated by principal component analysis (PCA), then we update $(\boldsymbol{\zeta}_{[j]}, \mathbf{\Theta}_j)$ by

$$\mathbf{\Theta}^{(r+1)\prime}_j = \omega^{1/2}\mathbf{E}_{eigen}(\mathbf{W}^{(r+1)\prime}_j\mathbf{W}^{(r+1)}_j; K), \tag{13}$$

$$\boldsymbol{\zeta}^{(r+1)}_{[j]} = \omega^{-1/2}\mathbf{W}^{(r+1)}_j \times \mathbf{E}_{eigen}(\mathbf{W}^{(r+1)\prime}_j\mathbf{W}^{(r+1)}_j; K). \tag{14}$$

Finally, we update $\phi_{jk}(t)$ by $\phi^{(r+1)}_{jk}(t) = \mathbf{\Theta}^{(r+1)\prime}_{jk}\mathbf{Z}(t)$. Repeat the above steps for $r = 1, 2, \cdots$ till $\mathbf{B}, \boldsymbol{\eta}$ and $(\mathbf{\Theta}_j, \boldsymbol{\zeta}_{[j]})$ $(j = 1, \cdots, q)$ converge.

3.3 Determining the number of factors and eigenfunctions

Under factor model for high-dimensional data, the criteria based on the parallel analysis (Hayton et al., 2004), penalized loss (Bai and Ng, 2002), eigenvalue ratio test (Lam and Yao, 2012) and adjusted eigenvalues thresholding (Fan et al., 2022) are proposed to estimate the number of factors. Here, using the idea of parallel analysis (Hayton et al., 2004), we select q by comparing the eigenvalues of the observed data with those of random noise data. We firstly calculate the eigenvalues of the observed data as

$$\lambda_{\mathbf{X},j} = \lambda_j \left\{ n^{-1} \sum_{i=1}^{n} n_i^{-1} \sum_{l=1}^{n_i} \mathbf{X}_i(t_{il}) \mathbf{X}_i'(t_{il}) \right\},$$

where $\lambda_j(A)$ is the j-th eigenvalue of the matrix (or covariance function) A. Then, we calculate the variance as $\text{var}(\mathbf{X}) = (np)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{p} n_i^{-1} \sum_{l=1}^{n_i} X_{ij}^2(t_{il})$, and randomly generate $\widetilde{X}_{ij}(t_{il}) \sim N(0, \text{var}(\mathbf{X}))$ $(i = 1, \cdots, n; j = 1, \cdots, p; l = 1, \cdots, n_i)$. We calculate the eigenvalues of the random nosie data as

$$\lambda_{\widetilde{\mathbf{X}},j} = \lambda_j \left\{ n^{-1} \sum_{i=1}^{n} n_i^{-1} \sum_{l=1}^{n_i} \widetilde{\mathbf{X}}_i(t_{il}) \widetilde{\mathbf{X}}_i'(t_{il}) \right\}.$$

We repeat this process 50 times and take the average of $\lambda_{\widetilde{\mathbf{X}},j}$ across all repetitions as $\bar{\lambda}_{\widetilde{\mathbf{X}},j}$ for each $j = 1, \cdots, p$. Finally, we determine the number of factors q as

$$q = \max\{j : \lambda_{\mathbf{X},j} > \bar{\lambda}_{\widetilde{\mathbf{X}},j}, j = 1, \cdots, p\}. \tag{15}$$

We use the similar method to select K. For each $j = 1, \cdots, q$, we firstly calculate the k-th eigenvalue of the covariance function of $h_{ij}^{(0)}(t)$ as $\lambda_{h_j,k} = \lambda_k\{\Sigma_{h_j}(s,t)\}$, where $\Sigma_{h_j}(s,t)$ is calculated by Yao et al. (2005) and $h_{ij}^{(0)}(t)$ is calculated by (10). Then we calculate the variance as $\text{var}(h_j) = (nq)^{-1} \sum_{i=1}^{n} n_i^{-1} \sum_{l=1}^{n_i} h_{ij}^{(0)2}(t_{il})$ and randomly generate $\widetilde{h}_{ij}(t_{il}) \sim N(0, \text{var}(h_j))$ $(i = 1, \cdots, n; l = 1, \cdots, n_i)$. We calculate the k-th eigenvalue of noise data as $\lambda_{\widetilde{h}_j,k} = \lambda_k\{\Sigma_{\widetilde{h}_j}(s,t)\}$, repeat this process 50 times and take the average of $\lambda_{\widetilde{h}_j,k}$ across all

repetitions as $\bar{\lambda}_{\tilde{h}_j,k}$ for each k. Finally, we determine the number of eigenfunctions K as

$$K = \max_{j=1,\cdots,q} K_j, \tag{16}$$

where $K_j = \max\{k : \lambda_{h_j,k} > \bar{\lambda}_{\tilde{h}_j,k}\}$.

4 Analysis of the atrophy of ROIs on MMSE outcomes with the ADNI data

In this section, we identify and evaluate the ROIs that are associated with cognitive decline by analyzing the ADNI data. There are two challenges in modeling the considered ADNI data. The first is the curse of dimensionality in analyzing ROI atrophy's impact on MMSE outcomes. We collect 95 ROI curves, resulting in high-dimensional functional data. Figure 1 (left) displays heat map of the covariance matrix calculated as $n^{-1}\sum_{i=1}^{n} n_i^{-1}\sum_{l=1}^{n_i}\mathbf{X}_i(t_{il})\mathbf{X}_i'(t_{il})$, where $t_{il} \in [0,1]$ is the quantile. The heat map reveals high correlations among different ROIs, highlighting the second challenge of incorporating this correlation when modeling the volume curves. Motivated by the two problems, we plot the first 15 eigenvalues of $n^{-1}\sum_{i=1}^{n} n_i^{-1}\sum_{l=1}^{n_i}\mathbf{X}_i(t_{il})\mathbf{X}_i'(t_{il})$ in Figure 1 (right). It indicates that the considered data show an obvious low-rank structure, suggesting that SF-FPCA can be used for the analysis.

To evaluate the fitting performance of volume curves between SF-FPCA and other existing methods, we perform unsupervised learning by applying SF-FPCA, FM-HD, DPCA, sFPCA-HD, uFPCA-HD, FPCA-ST, pFPCA, and DIVE. All methods are applied to model volume curves without considering age, gender, and MMSE score. We randomly divide the data into a 50% training set and a 50% test set. This random splitting process is repeated 200 times to ensure robustness. Then we consider the normalized prediction error PE= $\sum_{i,j} n_i^{-1}\sum_{l=1}^{n_i}\{\widehat{X}_{ij}(t_{il}) - X_{ij}(t_{il})\}^2 / \sum_{i,j} n_i^{-1}\sum_{l=1}^{n_i} X_{ij}^2(t_{il})$ for the observations $X_{ij}(t_{il})$

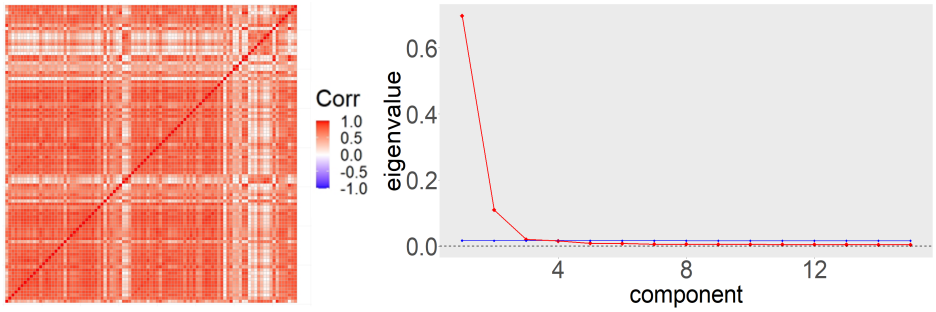


Figure 1: (a) The heat map of the matrix $n^{-1}\sum_{i=1}^{n} n_i^{-1}\sum_{l=1}^{n_i}\mathbf{X}_i(t_{il})\mathbf{X}_i'(t_{il})$; (b) Red solid line: The first 15 eigenvalues of the matrix $n^{-1}\sum_{i=1}^{n} n_i^{-1}\sum_{l=1}^{n_i}\mathbf{X}_i(t_{il})\mathbf{X}_i'(t_{il})$ in the ADNI analysis; Blue solid line: The average of the first 15 eigenvalues of the matrix $n^{-1}\sum_{i=1}^{n} n_i^{-1}\sum_{l=1}^{n_i}\widetilde{\mathbf{X}}_i(t_{il})\widetilde{\mathbf{X}}_i'(t_{il})$; Black dashed line: x-axis.

in the testing set, which is independent of the training data but has the same distribution as the training data. By Figure 1, based on the criteria in Section 3.3, we select $q = 3$ and $K = 2$ for SF-FPCA. For other competitive methods, we use grid search to select parameters and display their corresponding optimal results. In Table 1, we present the PEs of volume curves of SF-FPCA and other methods. The results show that SF-FPCA achieves the highest prediction accuracy.

Furthermore, since the proposed SF-FPCA is used to address three challenges: intrinsic dependence, spatial dependence, and spatial distinctions, which are expressed by factor model with simple loading $\mathbf{B}$, smoothing loading $\mathbf{f}$ and piecewise structure on loading, respectively. To demonstrate the necessity of the new framework, we compare the prediction accuracy for volume curves of four methods: factor model with simple loading $\mathbf{B}$ and $\mathbf{f} = \mathbf{0}$, factor model with smoothing loading $\mathbf{f}$ and $\mathbf{B} = \mathbf{0}$, factor model with both simple and smoothing loading $\mathbf{B} + \mathbf{f}$ and then the proposed factor model with piecewise smoothness for $\mathbf{B} + \mathbf{f}$, termed by SF-FPCA$_{\mathbf{B}}$, SF-FPCA$_{\mathbf{f}}$, SF-FPCA$_{\mathbf{B}+\mathbf{f}}$ and SF-FPCA, respectively. Here, SF-FPCA$_{\mathbf{B}}$, SF-FPCA$_{\mathbf{f}}$, and SF-FPCA$_{\mathbf{B}+\mathbf{f}}$ are derived by applying the proposed algorithm for SF-FPCA with either $\mathbf{B}$ or $\mathbf{f}$ set to zero or by removing the piecewise constant regularization on $\mathbf{B}$. The prediction accuracy for volume curves of these four methods is shown in Table 1. The results show that our proposed SF-FPCA, simul-

taneously considering intrinsic dependence, spatial dependence, and spatial distinctions, demonstrates the highest prediction accuracy for volume curves.

Table 1: PEs and the corresponding standard deviations (reported in parentheses) of the volume curves in ADNI data.

	SF-FPCA	SF-FPCA$_{\mathbf{B}}$	SF-FPCA$_{\mathbf{f}}$	SF-FPCA$_{\mathbf{B+f}}$
PE	0.0662(0.0011)	0.0707(0.0012)	0.0698(0.0010)	0.0697(0.0011)
	FPCA-ST	pFPCA	DIVE	FM-HD
PE	0.0958(0.0017)	0.0772(0.0008)	0.2914(0.0021)	0.0716(0.0007)
	uFPCA-HD	sFPCA-HD	DPCA	
PE	0.1074(0.0009)	0.2184(0.0030)	0.1093(0.0015)	

Next, we assess the impact of ROI volumes on MMSE scores. Initially, we compare the fitting performances of volume curves and scalar volumes on MMSE predictions. Each MMSE regression model includes age and gender (male=1, female=0) as additional covariates. The estimated scores $\widehat{\boldsymbol{\zeta}}$ from SF-FPCA serve as covariates for linear regression (Fun-LR), while scalar volumes are directly utilized for ridge regression (Vol-RR). Furthermore, we explore dimensionality reduction of scalar volumes using a classical factor model, employing the extracted factors as covariates for linear and ridge regressions. For the scalar volume factor model, we consider scenarios with 2, 10, and 20 factors, denoted by Vol-F2-LR, Vol-F10-LR, and Vol-F20-LR, respectively. Additionally, for 20 factors, we also incorporate ridge regression, indicated by Vol-F20-RR. We split the data into 50% training and 50% test sets, and tune the parameters of all ridge regressions using 10-fold cross-validation. The results depicted in Figure 2 reveal that using scores extracted from SF-FPCA as covariates yields superior predictive accuracy compared to models utilizing scalar volumes, whether directly or via factorization. This finding underscores the advantage of employing volume curves over scalar volumes.

Additionally, we evaluate the MMSE prediction accuracy for SF-FPCA and other competing methods, including commonly used functional regression models such as 3D func-

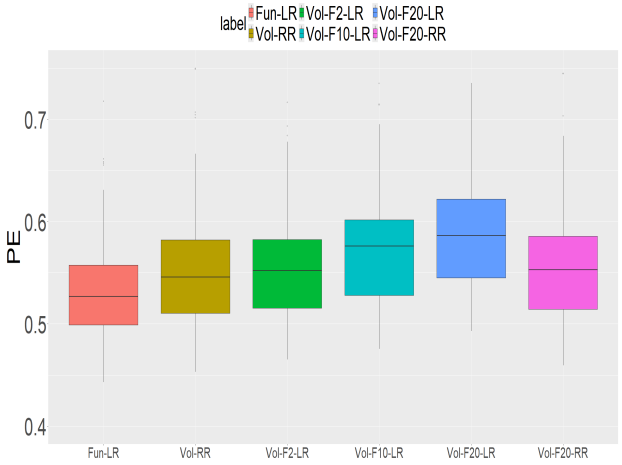


Figure 2: PEs of the regression models based on scores from SF-FPCA, scalar volumes, and the factors from scalar volumes, respectively. Fun-LR is the linear regression using scores from SF-FPCA as covariates; Vol-RR is the linear regression directly using scalar volumes; Vol-F2-LR, Vol-F10-LR, and Vol-F20-LR are the linear regressions using latent factors from scalar volumes with the number of factors being 2, 10, and 20; Vol-F20-RR is the ridge regression using latent factors from scalar volumes with the number of factors being 20.

tional regression (FLR-3D, Wang et al. 2014), functional linear regression and functional additive regression (FLR and FAR, Fan et al. 2015), and the functional single-index model (FSIM, Jiang et al. 2020). We also compare with linear regression models for MMSE using features extracted from uFPCA-HD, FM-HD, DPCA, sFPCA, FPCA-ST, pFPCA and DIVE, respectively, incorporating age and gender as covariates. For the methods FLR-3D, FLR, FAR and FSIM, we also include age and gender as linear terms in the model. As indicated in Table 2, our proposed SF-FPCA, utilizing scores extracted from the volume curves, demonstrates the highest prediction accuracy for MMSE. Combining the results from Table 1 on the prediction accuracy of the volume curves, we conclude that SF-FPCA effectively captures the complex relationships within the volume curves across different ROIs and retains sufficient information, thereby enhancing the predictive accuracy for MMSE.

SF-FPCA is designed to address several important aspects of high-dimensional functional data. Firstly, by employing the FPCA framework, SF-FPCA takes into account the dependence over time by capturing the temporal patterns and variations in the functional

Table 2: PEs and the corresponding standard deviations (reported in parentheses) of MMSE in ADNI data.

	SF-FPCA	FPCA-ST	pFPCA	DIVE
PE	0.5340(0.0521)	0.5798(0.0448)	0.5850(0.0458)	0.6083(0.0417)
	FM-HD	uFPCA-HD	sFPCA-HD	DPCA
PE	0.8470(0.1279)	0.5844(0.0468)	0.6346(0.0509)	1.3893(0.2357)
	FLR-3D	FLR	FAR	FSIM
PE	0.5639(0.0562)	0.6369(0.1045)	0.6364(0.1042)	0.5629(0.0578)

data. Secondly, SF-FPCA considers the inter-variable correlation among functional variables by a factor model structure. Additionally, SF-FPCA recognizes that the variation of measured values with respect to spatial coordinates, which may simultaneously capture spatial correlation and differences between different pieces.

We provide a detailed explanation of the analysis steps based on SF-FPCA and present our findings. In the analysis, if $\mathbf{b}_j = \mathbf{b}_{j'}$, then we let ROIs j and j' belong to the same piece. Based on this approach, 89 ROIs are divided into three pieces of sizes 26, 29, and 34. In addition, 2 ROIs are grouped into one set, while the remaining 4 ROIs are each assigned to a unique piece. The detailed results are presented in Table S5 in Supplementary Material F. To visualize the locations of these different pieces, we also plot Figure 3. We use blue to mark the ROIs in the right hemisphere and red for the ROIs in the left hemisphere. This figure provides a spatial representation of the ROIs, allowing for a better understanding of their anatomical positions within the brain. Here are our findings:

- In piece 1, the main ROIs include cerebellar vermal lobules (ROI_{630}, ROI_{631}, ROI_{632}), ventricles (ROI_4, ROI_5, ROI_{14}, ROI_{15}, ROI_{43}), caudate (ROI_{11}, ROI_{50}), lateral occipital (ROI_{1011}, ROI_{2011}), and anterior cingulate cortex (ROI_{1002}, ROI_{2002}, ROI_{2026}). These parts are predominantly located in the middle, deep, and posterior regions of the brain, as depicted in Figure 3 (a).

- Piece 2 mainly consists of lateral orbitofrontal (ROI_{1012}, ROI_{2012}), par opercularis

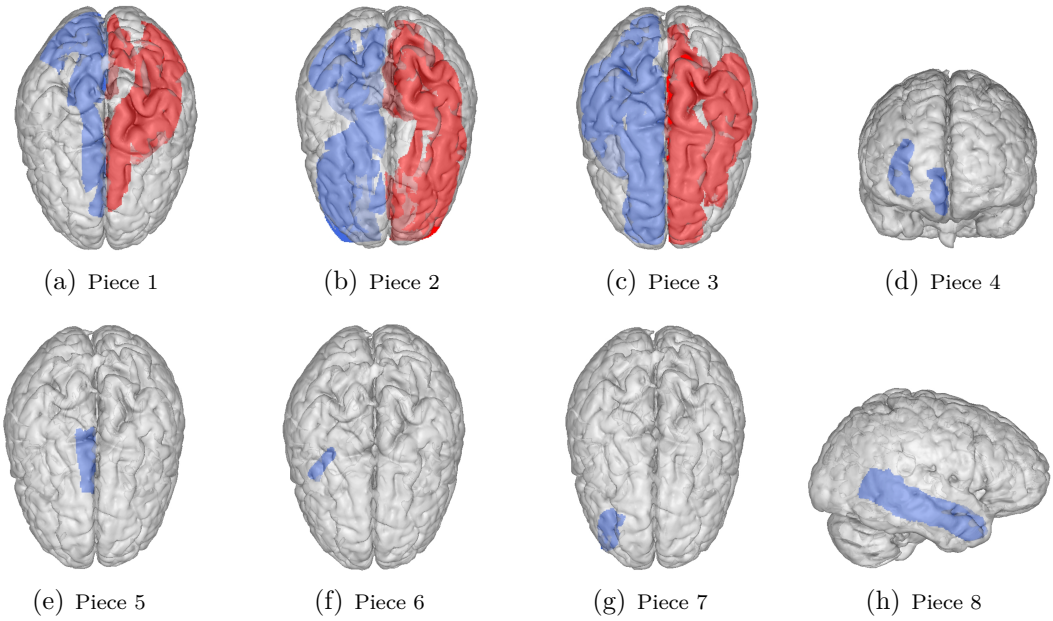
precentral (ROI_{1018}, ROI_{2018}), pars triangularis (ROI_{1020}, ROI_{2020}), and precentral gyrus (ROI_{1024}, ROI_{2024}). These regions are primarily located in the anterior part of the cerebral cortex. According to Figure 3 (b), the remaining ROIs in piece 2 are situated at the bottom of the cerebral cortex.

- Piece 3 covers the remaining majority of ROIs, primarily distributed in the upper and lateral regions of the cerebral cortex, such as medial orbitofrontal (ROI_{1014}), right superior parietal (ROI_{2029}), superior frontal (ROI_{1028}, ROI_{2028}), and some areas of the right temporal lobe (ROI_{2009}, ROI_{2030}, ROI_{2034}). In addition, piece 3 also includes most parts of the hippocampus (ROI_{53}, ROI_{1016}, ROI_{2016}). See Figure 3 (c) for a spatial visualization of these ROIs.

- We observe that certain ROIs in the right hemisphere are outside these three groups, as depicted in Figure 3 (d)-(h), whereas this pattern is not observed in the left hemisphere. This discrepancy emphasizes the distinctions between the two hemispheres. We plan to delve deeper into these differences in our subsequent research on cognitive function.

- As illustrated in Figure 3 (d), the right medial orbitofrontal (ROI_{2014}) and insula (ROI_{2035}) appear to be separate, yet they may share some similarities. Research by Levens and Phelps (2010) suggests that the right insula and right orbitalfrontal cortex are implicated in emotional interference resolution in working memory tasks. Moreover, we observe that the right middle temporal (ROI_{2015}), right pars orbitalis (ROI_{2019}), right posterior cingulate (ROI_{2023}), and right transverse temporal (ROI_{2034}) exhibit distinct characteristics compared to other regions in Figure 3 (e)-(h), respectively. These differences warrant further investigation through detailed medical analysis.

Our analysis primarily investigates the impact of ROI volumes on cognitive func-



(a) Piece 1 (b) Piece 2 (c) Piece 3 (d) Piece 4

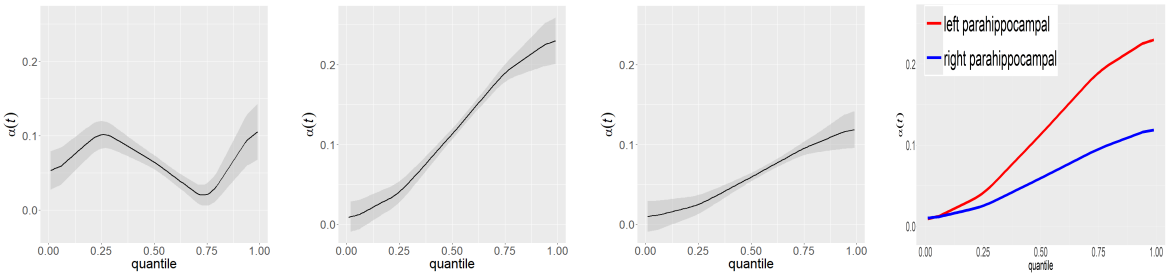
(e) Piece 5 (f) Piece 6 (g) Piece 7 (h) Piece 8

Figure 3: The locations of the ROIs in 8 pieces in the brain (ROIs in the left hemisphere are marked in red and in the right are marked in blue). (a): Piece 1: ROIs primarily located in the middle, deep, and posterior regions of the brain; (b): Piece 2: ROIs primarily located in the anterior part and bottom of the cerebral cortex; (c): Piece 3: ROIs primarily located in the upper and lateral regions of the cerebral cortex; (d): Piece 4: right medial orbitofrontal and insula; (e): Piece 5: right posterior cingulate; (f): Piece 6: right transverse temporal; (g): Piece 7: right pars orbitalis; (h): Piece 8: right middle temporal.

tions. Denote the regression coefficients corresponding to $\boldsymbol{\zeta}$ by $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_6)'$, and $\boldsymbol{\zeta}'\boldsymbol{\beta}$ represents the measurement of the effect of ROIs on MMSE. Following the transformation outlined in Supplementary Material E, we obtain $\boldsymbol{\zeta}'\boldsymbol{\beta} = \int \boldsymbol{\alpha}(t)'\mathbf{X}(t)dt$, where $\boldsymbol{\alpha}(t) = \{\alpha_1(t), \cdots, \alpha_p(t)\}' = (\mathbf{B} + \mathbf{f})\{(\mathbf{B} + \mathbf{f})'(\mathbf{B} + \mathbf{f})\}^{-1}\boldsymbol{\Phi}'(t)\boldsymbol{\beta}$ is the regression coefficient function of $\mathbf{X}(t)$ and can be used to elucidate the impact of the estimated eigen components on MMSE. The varying sign of $\alpha_j(t)$ over $t \in [0, 1]$ suggests that $X_j(\cdot)$ may have a positive effect in some regions and a negative effect in others, making it challenging to assess the overall impact of $X_j(\cdot)$. Therefore, in this paper,we consider the atrophy of ROI j to be a influential risk factor for AD if the pointwise confidence interval of $\alpha_j(\cdot)$ does not envelop the x-axis, which indicates $\alpha_j(\cdot) > 0$ or $\alpha_j(\cdot) < 0$. Based on the estimates of $\boldsymbol{\alpha}(\cdot)$, we identify 36 ROIs associated with cognitive functions and all the effects are positive, which strongly implies the positive correlation between ROI atrophy and MMSE decline. Among

the 36 ROIs, 21 are situated in the left hemisphere, 12 in the right hemisphere, and the remaining 3 ROIs are the brain stem, cerebellar vermal lobules I-V, and VI-VII.

For each ROI that has an impact on AD, we plot its mean curve and pointwise confidence bands based on 200 boostrap sampling, leading to the following observations. Figure 4 (a) illustrate that left hippocampal atrophy (ROI_{17}) correlates with cognitive decline. As discussed in Section 1, hippocampal volume is one of the most established biomarkers for AD (Foundas et al., 1997; Antoniano-Villalobos et al., 2014). While the right hippocampus is not significant. The results consistent with recent findings (Scheltens et al., 2021) suggesting more severe atrophy of the left hippocampus in AD patients compared to the right. Additionally, we find that atrophy of the parahippocampal gyrus (ROI_{1016}, ROI_{2016}) is associated with cognitive decline, as shown in Figures 4(b) and 4(c), aligning with previous studies (van Hoesen et al., 2010). The left and right parahippocampal gyrus exhibit similar mechanisms, with the left having a greater impact than the right, as depicted in Figure 4(d).

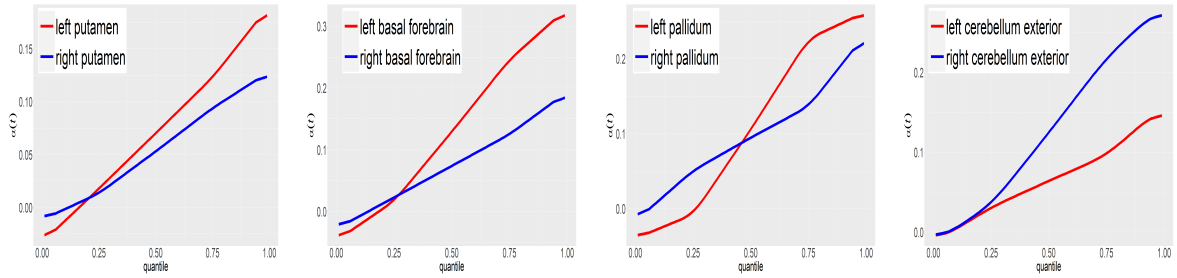


(a) left hippocampus (b) left parahippocampal (c) right parahippocampal (d) left and right parahippocampal

Figure 4: (a) Functional regression coefficient estimates and confidences bands of left hippocampus; (b) and (c): Functional regression coefficient estimates and confidences bands of left and right parahippocampal gyrus; (d): Comparison of the estimates of left and right parahippocampal gyrus (left in red and right in blue) .

It appears that Figures 4(b) - 4(d) suggest a similar impact of left and right brain atrophy on MMSE. However, among the 36 ROIs, aside from the parahippocampal gyrus, we only identify four additional brain regions where the impact of the left and right parts

on MMSE is similar, as depicted in Figures 5. These regions include the putamen (ROI_{12}, ROI_{51}), basal forebrain (ROI_{91}, ROI_{92}), pallidum (ROI_{13}, ROI_{52}), and cerebellum exterior (ROI_6, ROI_{45}). The effect of atrophy in the basal forebrain on cognitive function has been studied in Hall et al. (2008), while de Jong et al. (2008) investigated the influence of putamen and pallidum atrophy on cognitive decline. The regression results in de Jong et al. (2008) suggested that the effect of left putamen is larger than that of the right. However, there are few relevant studies on the cerebellum exterior. Even among regions with similar mechanisms, there is no clear distinction in the impact between the left and right hemispheres. The results in Figures 4(d) and 5 show that three pairs of ROIs exhibit a greater impact in the left hemisphere, one pair shows a higher impact in the left hemisphere only at high quantiles, and one pair shows a greater impact in the right hemisphere. Apart from these five pairs of ROIs, the trends of the coefficient functions at the same locations in the left and right hemispheres are even different. These observations indicate differences in the impact of the left and right hemispheres on cognitive function.



(a) putamen (b) basal forebrain (c) pallidum (d) cerebellum exterior

Figure 5: Comparison of the estimates of the pairwise ROI in the left (red) and right (blue): (a): Putamen; (b): Basal forebrain; (c): Pallidum; (d): Cerebellum exterior.

Furthermore, Figure S3(a) in Supplementary Material F suggests that left middle temporal (ROI_{1015}) atrophy induces cognitive decline, consistent with findings by Fruijtier et al. (2019) who pointed out that medial temporal lobe atrophy is one of the three best-validated neuroimaging biomarkers for AD. We further identify more detailed ROIs that affect cogni-

tive functions in the temporal lobe, such as left transverse temporal (ROI_{1034}), right inferior temporal (ROI_{2009}), and right fusiform gyrus (ROI_{2007}) atrophy, all shown to induce cognitive decline in Figures S3(b)-S3(d), consistent with findings by Yang et al. (2019). We also find differences in the impacts of cognitive decline between the left and right temporal lobes. Additionally, we identify numerous ROIs in the frontal lobe that have an effect on MMSE, such as left lateral orbitofrontal (ROI_{1012}), which has been well-studied in Rizvi et al. (2021). There is relatively little research on left pars opercularis (ROI_{1018}) and left pars triangularis (ROI_{1020}) atrophy for AD. The related results are shown in Figures S4 in Supplementary Material F.

We investigate the effects of cerebellar atrophy on cognitive function, including cerebellum exterior (ROI_6, ROI_{45}), putamen (ROI_{12}, ROI_{51}), left ventral diencephalon (DC, ROI_{28}), and cerebellar vermal lobules I-VII (ROI_{630}, ROI_{631}) in Figures S5 in Supplementary Material F. There has been increasing attention on the association between cerebellar atrophy and AD, such as de Jong et al. (2008) for putamen. Unlike the temporal and frontal lobes, our findings indicate that left and right cerebellum may exert similar influences on cognitive function, particularly in cerebellum exterior and putamen. Furthermore, we find that the atrophy in different locations of the cerebellar vermal lobules has varying impacts on cognitive decline. Toniolo et al. (2018) found that the mechanisms of damage in different locations of cerebellar vermal lobules vary from the early to late clinical stages of AD. Our results show that the impact mechanisms on MMSE are completely different between lobules I-V and VI-VII, and lobules VIII-X may not have any impact at all.

We show another 13 ROIs in Figures S6 in Supplementary Material F, some of which are located in the parietal lobe (ROI_{1008}, ROI_{1025}, ROI_{2029}) and cingulate gyrus (ROI_{1010}, ROI_{1023}, ROI_{2023}). Among these 13 ROIs, 7 have been well studied in existing literature, such as Ji et al. (2021) for brain stem (ROI_{16}), and Jacobs et al. (2011) for the parietal lobe. We find in Figures S6(j) and S6(k) that the exacerbation of cognitive decline is associated

with atrophy in pairs of posterior cingulate (ROI_{1023}, ROI_{2023}). We also observe significant differences in the impact mechanisms between the left and right hemisphere at the pairwise ROIs. Furthermore, we identify other 6 ROIs that are relevant to cognitive decline and are rarely or never studied, which are left thalamus proper (ROI_{10}), left accumbens area (ROI_{26}), left precentral (ROI_{1024}), left insula (ROI_{1035}), right cuneus (ROI_{2005}) and right paracentral (ROI_{2017}). Most of them are located in the left hemisphere.

5 Numerical Studies

In this section, we conduct simulations to assess the prediction accuracy for $\mathbf{X}(\cdot)$ of the proposed method (SF-FPCA) by comparing it with state-of-the-art methods, including uFPCA-HD, FM-HD, DPCA, and sFPCA-HD for high-dimensional functional data, and FPCA-ST, pFPCA, and DIVE for spatially correlated high-dimensional functional data. We also evaluate the prediction accuracy for Y by comparing these methods with functional regression approaches, such as FLR-3D, FLR, FAR, and FSIM. Additionally, we provide supplementary simulation results related to the performance of SF-FPCA in Supplementary Materials D, which includes estimation consistency, selection of the number of factors and eigenfunctions, and computational costs.

The performance of the estimator is evaluated via the PE of the high-dimensional funticonal data $\mathbf{X}(\cdot)$ and response Y for the observations in the testing set, which is independent of the training data but has the same distribution and size as the training data. The PE is calculated based on 200 repetitions. To compare the prediction accuracy of high-dimensional functional data, we consider three scenarios of $\mathbf{X}(\cdot)$.

Scenario X1: To be fair for comparison, we generate data not adhering to any model considered in the paper. For each $j = 1, \cdots, p$, we generate each dimension of the 3D coordinates $\mathbf{s}_j = (s_{j1}, s_{j2}, s_{j3})'$ from $U(0, 1)$. Following similar settings in Hyun et al.

(2016) and Chen et al. (2017), we generate data from the following model:

$$X_{ij}(t) = \sum_{k=1}^{3} \chi_{1,k}\psi_k(\mathbf{s}_j, t) + \sum_{l=1}^{4}\sum_{k=1}^{4} \chi_{2,lk}\phi_{1,l}(\mathbf{s}_j)\phi_{2,k}(t) + \varepsilon_{ij}(t),$$

where $\chi_{1,k} \sim N(0, k^2+3)$; $\psi_1(\mathbf{s}_j, t) = \sqrt{2}/45 \sin\{\pi(\mathbf{s}_{j1}+s_{j2}+s_{j3})/15\}\cos(\pi t/18), \psi_2(\mathbf{s}_j, t) = \sqrt{2}/45\cos\{\pi(\mathbf{s}_{j1}+s_{j2}+s_{j3})/15\}\cos(\pi t/18), \psi_3(\mathbf{s}_j, t) = 1/(30\sqrt{3})\cos(\pi t/12)$; $\chi_{2,lk} \sim N(0, \sqrt{(17-lk)/4})$; $\phi_{1,l}(\mathbf{s}) = \sin(2l\pi\|s\|_2/3)/10$ if l is odd and $\cos(2l\pi\|s\|_2/3)/10$ if l is even; $\phi_{2,k}(t) = \sqrt{2}\sin(2k\pi t)$ if k is odd and $\sqrt{2}\cos(2k\pi t)$ if k is even; and $\varepsilon_{ij}(t) \sim N(0, 0.2)$. We set $(n, p) = (200, 100)$. For each individual, 20 observation time points are randomly sampled from $U(0, 1)$.

Scenario X2: We generate data from the estimated SF-FPCA model for analyzing the ADNI data, where $n_i = 101$, $n = 803$ and $p = 95$. The $\mathbf{s}_j$ represents the 3D coordinates of the center for ROI j. We take $q = 3$ and $K = 2$ that selected in Section 4. Denote $\widehat{\mathbf{B}}, \widehat{\boldsymbol{\eta}}, \widehat{\boldsymbol{\Theta}}, \widehat{\boldsymbol{\zeta}}_i$ as the resulting estimators of SF-FPCA based on the ADNI analysis, then we generate data as $\mathbf{X}_i(t) = (\mathbf{B}+\mathbf{f})\boldsymbol{\Phi}'(t)\boldsymbol{\zeta}_i + \mathbf{u}_i(t)$, where $\mathbf{B}$ is generated based on the grouping structure of ROIs, $\mathbf{f}$ and $\boldsymbol{\Phi}(t)$ are generated by adding random noise to $\widehat{\mathbf{f}}$ and $\widehat{\boldsymbol{\Phi}}(t)$ respectively, $\boldsymbol{\zeta}_i = \widehat{\boldsymbol{\zeta}}_i$ and $u_{ij}(t) \sim N(0, 0.01)$. Particularly, $\mathbf{b}_j = (1, -1, 1)'$ if ROI j belongs to pieces 1 or 4-8, $\mathbf{b}_j = (0, 0, 0)'$ if ROI j belongs to piece 2, and $\mathbf{b}_j = (5, 5, -5)'$ if it belongs to piece 3; $f_j(\mathbf{s}) = \boldsymbol{\eta}_j'\mathbf{M}(\mathbf{s})$ with $\boldsymbol{\eta}_j = \widehat{\boldsymbol{\eta}}_j + N(\mathbf{0}, 0.01\mathbf{I}_\tau)$; and $\phi_{jk} = \boldsymbol{\Theta}_{jk}'\mathbf{Z}(t)$ with $\boldsymbol{\Theta}_{jk} = \widehat{\boldsymbol{\Theta}}_{jk}' + N(\mathbf{0}, 0.03\mathbf{I}_m)$.

Scenario X3: We compare the proposed SF-FPCA method with other competing methods using the original density curves only in the ADNI3 study without undergoing the log quantile density transformation. Particularly, $X_{ij}(t)$ represents the density of volume t at ROI j for individual i, $n = 380$, $p = 95$, and $\mathbf{s}_j$ still denotes the 3D coordinates for the center of ROI j.

Table 3 presents the PEs of SF-FPCA, FM-HD, DPCA, sFPCA-HD, uPFCA-HD, FPCA-ST, pFPCA, and DIVE with their corresponding optimal tuning parameters in Scenarios X1-X3. By leveraging both factor structure and spatial information, our pro-

posed SF-FPCA consistently outperforms across all scenarios. In addition, we also note that FM-HD typically performs well due to its utilization of the factor structure. Incorporating spatial information, methods such as FPCA-ST, pFPCA, and DIVE demonstrate superior performance compared to sFPCA-HD and DPCA across most scenarios.

Combining the findings from Scenario X3 with those illustrated in Table 1 from Section 4, SF-FPCA achieves the highest prediction accuracy utilizing both original density curves and its transformed curves. This indicates that the features extracted by SF-FPCA can comprehensively represent the ROI information.

Table 3: The PEs of $\mathbf{X}(\cdot)$ of all the mentioned methods in Scenario X1-X3.

	Scenario X1	Scenario X2	Scenario X3
SF-FPCA	0.4131(0.0068)	0.1669(0.0037)	0.0865(0.0037)
FPCA-ST	0.5471(0.0092)	0.3057(0.0009)	0.1487(0.0032)
pFPCA	0.4845(0.0088)	0.2001(0.0011)	0.1372(0.0026)
DIVE	0.6220(0.0034)	0.3072(0.0023)	0.6116(0.0024)
FM-HD	0.4271(0.0073)	0.2285(0.0010)	0.1188(0.0020)
uFPCA-HD	0.4838(0.0070)	0.2085(0.0012)	0.1781(0.0034)
sFPCA-HD	0.7561(0.0048)	0.6582(0.0016)	0.5023(0.0048)
DPCA	0.7652(0.0038)	0.5646(0.0009)	0.1848(0.0035)

Furthermore, to facilitate comparison with all competing methods, we further add the response variable Y for each setting. Particularly,

Scenario Y1: The predictor functions $\mathbf{X}_i(\cdot)$ are generated under Scenario X1, and the response Y_i is modeled by the following nonlinear functional regression:

$$Y_i = \sum_{j=1}^{p-1} \int \{\sin(t) + \cos(t)\} X_{ij}(t) X_{i,j+1}(t)\ dt + \sum_{j=1}^{p} \int \sin(t) X_{ij}(t)\ dt$$
$$+ \sum_{j=1}^{p} \int \cos(t) sign\{X_{ij}^2(t)\} X_{ij}^2(t)\ dt + \sum_{j=1}^{p} \int \cos(t) X_{ij}^3(t)\ dt + \varepsilon_i,$$

where $sign(a)$ is the sign of a and $\varepsilon_i \sim N(0, 2)$.

Scenario Y2: The generation of $\mathbf{X}_i(\cdot)$ follows Scenario X2, while Y_i is modeled using

the functional linear regression model:

$$Y_i = \sum_{j=1}^{p} \int \sin(2\pi t) X_{ij}(t)\ dt + \varepsilon_i,$$

where $\varepsilon_i \sim N(0, 2)$.

Scenario Y3: $\mathbf{X}_i(\cdot)$ is identical to Scenario X3, which consists of the original density curves only from the ADNI3 study without the log quantile density transformation, and Y_i represents the corresponding real MMSE.

As shown in Table 4, the linear regression model based on the extracted scores using our proposed SF-FPCA achieves the highest prediction accuracy for Y across all scenarios. Combining the results of predicting $\mathbf{X}(\cdot)$ in Table 3, SF-FPCA captures the complex dependencies among the high-dimensional functional data and retains sufficient information, thereby enhancing the predictive accuracy for Y.

Table 4: The PEs of Y of all the mentioned methods in Scenario Y1-Y3.

	Scenario Y1	Scenario Y2	Scenario Y3
SF-FPCA	0.4934(0.0987)	0.2539(0.0198)	0.8999(0.0811)
FLR-3D	0.5810(0.1559)	0.3803(0.0264)	0.9681(0.0738)
FLR	0.6043(0.0947)	0.4189(0.0192)	0.9747(0.0230)
FAR	0.6379(0.1107)	0.6316(0.0232)	1.0071(0.0227)
FSIM	0.5696(0.1937)	0.4330(0.0624)	1.0383(0.1980)
FPCA-ST	0.6177(0.1126)	0.2801(0.0217)	0.9419(0.0589)
pFPCA	0.5596(0.1075)	0.3076(0.0230)	0.9906(0.0323)
DIVE	0.5926(0.1911)	0.7667(0.0211)	1.0165(0.0278)
FM-HD	0.5854(0.1211)	0.3477(0.0343)	1.0877(0.1835)
uFPCA-HD	0.6036(0.1175)	0.3151(0.0234)	1.0085(0.0415)
sFPCA-HD	0.6642(0.1405)	0.7816(0.0261)	1.0869(0.1122)
DPCA	0.6195(0.1423)	0.5256(0.0623)	1.5102(0.3801)

6 Discussion

We have utilized ROI volume changes to explore the intricate relationships between brain atrophy and cognitive decline, aiming to provide a comprehensive understanding of brain tissue alterations in AD. Instead of relying on scalar volumes, we employ volume density curves of ROIs, which offer enhanced sensitivity to AD-related changes and lead to improved prediction of cognitive function.

Given the complex interplay among high-dimensional ROI volume curves and their spatial relationships, we develop a computationally efficient model called SF-FPCA. SF-FPCA effectively captures correlations among functional variables using factor processes, while applying FPCA to these factor processes enables us to address temporal dependence and extract low-dimensional features. Importantly, SF-FPCA fully incorporates the high spatial correlation and piecewise smooth characteristics among various ROIs, resulting in superior fitting accuracy compared to existing competing methods.

Using the features extracted from SF-FPCA, we identify 36 ROIs associated with cognitive decline, shedding light on several ROIs that have received limited or no research attention in previous studies. Furthermore, we delve into the impact of cerebellar atrophy on cognitive decline and uncover differences in mechanisms between the left and right hemispheres. This comprehensive analysis contributes to a deeper understanding of the intricate relationship between brain structure alterations and cognitive decline in AD. ADNI patients are primarily from North America, rather than being representative of a global population. As a result, some of our new findings, particularly those that are less studied, may apply only to this dataset. Evaluating the transferability of our findings by applying the proposed method to other datasets with different populations, such as the UK Biobank, would be valuable.

In our study, we utilize SF-FPCA to extract information from ROI volume curves and

integrate age and gender as predictors within a linear regression framework to forecast cognitive function. However, there are several avenues for further research that could build upon our work.

Firstly, future studies could extend the linear SF-FPCA approach to a generalized framework that accommodates nonlinear relationships. By capturing more complex relationships between ROI volume curves and latent processes, we may enhance feature extraction and improve predictive performance. Nonetheless, exploring the computational and theoretical feasibility of such a nonlinear framework would require further investigation.

Secondly, there is potential to enhance AD analysis by incorporating additional information and employing more sophisticated modeling techniques. For instance, integrating neural networks to fuse MRI data, biomarkers, and volumetric measurement curves could provide a comprehensive approach to extract information for predicting cognitive function more accurately. However, explaining the impact of different types of data on cognitive function within complex models may pose challenges.

Data Availability Statement

The MRI scans utilized in this study were sourced from the ADNI database (http://adni.loni.usc.edu/). The data used in this article were obtained from the ADNI database. While investigators within the ADNI contributed to the design and implementation of ADNI and provided data, they did not participate in the 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.

Disclosure Statement

The authors report there are no competing interests to declare.

Funding

The research was supported by National Key R&D Program of China (No.2022YFA1003702), National Natural Science Foundation of China (Nos. 11931014), and New Cornerstone Science Foundation.

References

Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., Gamst, A., Holtzman, D. M., Jagust, W. J., Petersen, R. C., Snyder, P. J., Carrillo, M. C., Thies, B., and Phelps, C. H. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.*, 7(3):270–279.

Antoniano-Villalobos, I., Wade, S., and Walker, S. G. (2014). A bayesian nonparametric regression model with normalized weights: A study of hippocampal atrophy in Alzheimer's disease. *J. Am. Stat. Assoc.*, 109(506):477–490.

Ash, R. and Gardner, M. (1975). *Topics in stochastic processes*. Academic Press.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Bai, J. and Ng, S. (2013). Principal components estimation and identification of static factors. *J. Econometrics*, 176(1):18–29.

Barnes, J., Godbolt, A. K., Frost, C., Boyes, R. G., Jones, B. F., Scahill, R. I., Rossor, M. N., and Fox, N. C. (2007). Atrophy rates of the cingulate gyrus and hippocampus in AD and FTLD. *Neurobiol. Aging*, 28(1):20–28.

Bing, X., Cheng, W., Feng, H., and Ning, Y. (2024). Inference in high-dimensional multivariate response regression with hidden variables. *J. Am. Stat. Assoc.*, 119(547):2066–2077.

Chen, K., Delicado, P., and Müller, H.-G. (2017). Modelling function-valued stochastic processes, with applications to fertility dynamics. *J. R. Stat. Soc. B*, 79(1):177–196.

Chincarini, A., Bosco, P., Gemme, G., Esposito, M., Rei, L., Squarcia, S., Bellotti, R., Minthon, L., Frisoni, G., Scheltens, P., Frölich, L., Soininen, H., Visser, P. J., and Nobili, F. (2014). Automatic temporal lobe atrophy assessment in prodromal AD: Data from the DESCRIPA study. *Alzheimers. Dement.*, 10(4):456–467.

de Jong, L. W., van der Hiele, K., Veer, I. M., Houwing, J. J., Westendorp, R. G. J., Bollen, E. L. E. M., de Bruin, P. W., Middelkoop, H. A. M., van Buchem, M. A., and van der Grond, J. (2008). Strongly reduced volumes of putamen and thalamus in Alzheimer's disease: an MRI study. *Brain*, 131(12):3277–3285.

Evans, M. C., Barnes, J., Nielsen, C., Kim, L. G., Clegg, S. L., Blair, M., Leung, K. K.,

Douiri, A., Boyes, R. G., Ourselin, S., and Initiative, A. D. N. (2010). Volume changes in Alzheimer's disease and mild cognitive impairment: cognitive associations. *Eur. Radiol.*, 20(3):674–682.

Fan, J., Guo, J., and Zheng, S. (2022). Estimating number of factors by adjusted eigenvalues thresholding. *J. Am. Stat. Assoc.*, 117(538):852–861.

Fan, J., Liao, Y., and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *J. R. Stat. Soc. B*, 75(4):603–680.

Fan, Y., James, G. M., and Radchenko, P. (2015). Functional additive regression. *Ann. Stat.*, 43(5):2296–2325.

Ferrarini, L., Palm, W. M., Olofsen, H., van der Landen, R., Blauw, G. J., Westendorp, R. G. J., Bollen, E. L. E. M., Middelkoop, H. A. M., Reiber, J. H. C., van Buchem, M. A., and Admiraal-Behloul, F. (2008). MMSE scores correlate with local ventricular enlargement in the spectrum from cognitively normal to Alzheimer disease. *NeuroImage*, 39(4):1832–1838.

Fjell, A. M. and Walhovd, K. B. (2010). Structural brain changes in aging: courses, causes and cognitive consequences. *Rev. Neurosci.*, 21(3):187–222.

Foundas, A. L., Leonard, C. M., Mahoney, S. M., Agee, O. F., and Heilman, K. M. (1997). Atrophy of the hippocampus, parietal cortex, and insula in Alzheimer's disease: A volumetric magnetic resonance imaging study. *Cogn. Behav. Neurol.*, 10(2):81–89.

Fruijtier, A. D., Visser, L., Maurik, I., Zwan, M., and Smets, E. (2019). Abide delphi study: topics to discuss in diagnostic consultations in memory clinics. *Alzheimers. Res. Ther.*, 11(1):77.

Gao, Y., Shang, H. L., and Yang, Y. (2019). High-dimensional functional time series forecasting: an application to age-specific mortality rates. *J. Multivariate. Anal.*, 170:232–243.

Guo, S., Qiao, X., and Wang, Q. (2022). Factor modelling for high-dimensional functional time series. *arXiv:2112.13651*.

Hall, A., Moore, R. Y., Lopez, O. L., Kuller, L., and Becker, J. T. (2008). Basal forebrain atrophy is a presymptomatic marker for Alzheimer's disease. *Alzheimers Dement.*, 4(4):271–279.

Hallin, M., Nisol, G., and Tavakoli, S. (2022). High-dimensional functional factor models I: Representation results. *J. Time Ser. Anal.*, 44(5-6):578–600.

Hayton, J. C., Allen, D. G., and Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: a tutorial on parallel analysis. *Organ. Res. Methods*, 7(2):191–205.

Henneman, W. J., Sluimer, J. D., Barnes, J., van der Flier, W. M., Sluimer, I. C., Fox, N. C., Scheltens, P., Vrenken, H., and Barkhof, F. (2009). Hippocampal atrophy rates in Alzheimer disease. *Neurology*, 72(11):999–1007.

Hu, X. and Yao, F. (2022). Sparse functional principal component analysis in high dimensions. *Stat. Sinica*, 38(32):1939–1960.

Hu, X. and Yao, F. (2024). Dynamic principal subspaces in high dimensions. *J. Am. Stat.*

Assoc., 119(545):308–319.

Hyun, J. W., Li, Y., Huang, C., Styner, M., Lin, W., Zhu, H., and for the Alzheimers Disease Neuroimaging Initiative (2016). STGP: Spatio-temporal gaussian process models for longitudinal neuroimaging data. *NeuroImage*, 134(1):550–562.

Jacobs, H. I. L., van Boxtel, M. P. J., Uylings, H. B. M., Gronenschild, E. H. B. M., Verhey, F. R., and Jolles, J. (2011). Atrophy of the parietal lobe in preclinical dementia. *Brain Cognition*, 75(2):154–163.

James, G. M., Hastie, T. J., and Sugar, C. A. (2000). Principal component models for sparse functional data. *Biometrika*, 87(3):587–602.

Ji, X., Wang, H., Zhu, M., He, Y., and Fu, Y. (2021). Brainstem atrophy in the early stage of Alzheimer's disease: a voxel-based morphometry study. *Brain Imaging Behav.*, 15:49–53.

Jiang, F., Baek, S., Cao, J., and Ma, Y. (2020). A functional single index model. *Stat. Sinica*, 30(1):303–324.

Jiang, F., Ma, Y., and Wei, Y. (2019). Sufficient direction factor model and its application to gene expression quantitative trait loci discovery. *Biometrika*, 106(2):417–432.

Lam, C. and Yao, Q. (2012). Factor modeling for high-dimensional time series: Inference for the number of factors. *Ann. Stat.*, 40(2):694–726.

Leung, K. K., Barnes, J., Ridgway, G. R., Bartlett, J. W., Clarkson, M. J., Macdonald, K., Schuff, N., Fox, N. C., Ourselin, S., and Initiative, A. D. N. (2010). Automated cross-sectional and longitudinal hippocampal volume measurement in mild cognitive impairment and Alzheimer's disease. *NeuroImage*, 51(4):1345–1359.

Levens, S. M. and Phelps, E. A. (2010). Insula and orbital frontal cortex activity underlying emotion interference resolution in working memory. *J. Cognitive. Neurosci.*, 22(12):2790–2803.

Li, F. and Sang, H. (2019). Spatial homogeneity pursuit of regression coefficients for large datasets. *J. Am. Stat. Assoc.*, 114(527):1050–1062.

Li, T., Li, T., Zhu, Z., and Zhu, H. (2022). Regression analysis of asynchronous longitudinal functional and scalar data. *J. Am. Stat. Assoc.*, 117(539):1228–1242.

Li, T., Zhu, H., Li, T., and Zhu, H. (2023). Asynchronous functional linear regression models for longitudinal data in reproducing kernel Hilbert space. *Biometrics*, 79(3):1880–1895.

Likeman, M., Anderson, V. M., Stevens, J. M., Waldman, A. D., Godbolt, A. K., Frost, C., Rossor, M. N., and Fox, N. C. (2005). Visual assessment of atrophy on magnetic resonanceimaging in the diagnosis of pathologically confirmed young-onset dementias. *Arch. Neurol.*, 62(9):1410–1415.

Liu, W., Lin, H., Zheng, S., and Liu, J. (2023). Generalized factor model for ultra-high dimensional correlated variables with mixed types. *J. Am. Stat. Assoc.*, 118(542):1385–1401.

Marinescu, R. V., Eshaghi, A., Lorenzi, M., Young, A. L., and Oxtoby, N. P. (2019). Dive: A spatiotemporal progression model of brain pathology in neurodegenerative disorders.

NeuroImage, 192(15):166–177.

Rizvi, B., Lao, P. J., Chesebro, A. G., Dworkin, J. D., Amarante, E., Beato, J. M., Gutierrez, J., Zahodne, L. B., Schupf, N., Manly, J. J., Mayeux, R., and Brickman, A. M. (2021). Association of regional white matter hyperintensities with longitudinal Alzheimer-like pattern of neurodegeneration in older adults. *JAMA Netw. Open.*, 4(10):e2125166.

Scheltens, P., De Strooper, B., Kivipelto, M., Holstege, H., Chételat, G., Teunissen, C. E., Cummings, J., and van der Flier, W. M. (2021). Alzheimer's disease. *Lancet*, 397(10284):1577–1590.

Talwar, P., Kushwaha, S., Chaturvedi, M., and Mahajan, V. (2021). Systematic review of different neuroimaging correlates in mild cognitive impairment and Alzheimer's disease. *Clin. Neuroradiol.*, 31:953–967.

Toniolo, S., Serra, L., Olivito, G., Marra, C., Bozzali, M., and Cercignani, M. (2018). Patterns of cerebellar gray matter atrophy across Alzheimer's disease progression. *Front. Cell. Neurosci.*, 12(430):1–8.

van Hoesen, G. W., Augustinack, J. C., Dierking, J., Redman, S. J., and Thangavel, R. (2010). The parahippocampal gyrus in Alzheimer's disease: clinical and preclinical neuroanatomical correlates. *Ann. NY. Acad. Sci.*, 911(1):254–274.

Wang, J., Zhao, Q., Hastie, T., and Owen, A. B. (2017). Confounder adjustment in multiple hypothesis testing. *Ann. Stat.*, 45(5):1863–1894.

Wang, X., Nan, B., Zhu, J., and Koeppe, R. (2014). Regularized 3d functional regression for brain image data via haar wavelets. *Ann. Appl. Stat.*, 8(2):1045–1064.

Yang, H., Xu, H., Li, Q., Jin, Y., Jiang, W., Wang, J., Wu, Y., Li, W., Yang, C., Li, X., Xiao, S., Shi, F., and Wang, T. (2019). Study of brain morphology change in Alzheimer's disease and amnestic mild cognitive impairment compared with normal controls. *Gen. Psychiatr.*, 32(2):e100005.

Yao, F., Müller, H.-G., and Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data. *J. Am. Stat. Assoc.*, 100(470):577–590.

Zhang, L., Baladandayuthapani, V., Zhu, H., Baggerly, K. A., Majewski, T., Czerniak, B. A., and Morris, J. S. (2016). Functional CAR models for large spatially correlated functional datasets. *J. Am. Stat. Assoc.*, 111(514):772–786.

Zhong, Q., Lin, H., and Li, Y. (2021). Cluster non-gaussian functional data. *Biometrics*, 77(3):852–865.

Zhou, L., Lin, H., and Liang, H. (2018). Efficient estimation of the nonparametric mean and covariance functions for longitudinal and sparse functional data. *J. Am. Stat. Assoc.*, 113(524):1550–1564.

Zhu, H., Fan, J., and Kong, L. (2014). Spatially varying coefficient model for neuroimaging data with jump discontinuities. *J. Am. Stat. Assoc.*, 109(507):1084–1098.