Statistical image analysis of longitudinal RAVENS images

Seonjoo Lee[1,2], Vadim Zipunnikov[3], Daniel S. Reich[3,4] and Dzung L. Pham[5]*

[1] *Department of Psychiatry and Biostatistics, Columbia University, New York, NY, USA,* [2] *New York State Psychiatric Institute, New York, NY, USA,* [3] *Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA,* [4] *Division of Neuroimmunology and Neurovirology, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, USA,* [5] *Center for Neuroscience and Regenerative Medicine, The Henry M. Jackson Foundation for the Advancement of Military Medicine, Bethesda, MD, USA*

Regional analysis of volumes examined in normalized space (RAVENS) are transformation images used in the study of brain morphometry. In this paper, RAVENS images are analyzed using a longitudinal variant of voxel-based morphometry (VBM) and longitudinal functional principal component analysis (LFPCA) for high-dimensional images. We demonstrate that the latter overcomes the limitations of standard longitudinal VBM analyses, which does not separate registration errors from other longitudinal changes and baseline patterns. This is especially important in contexts where longitudinal changes are only a small fraction of the overall observed variability, which is typical in normal aging and many chronic diseases. Our simulation study shows that LFPCA effectively separates registration error from baseline and longitudinal signals of interest by decomposing RAVENS images measured at multiple visits into three components: a subject-specific imaging random intercept that quantifies the cross-sectional variability, a subject-specific imaging slope that quantifies the irreversible changes over multiple visits, and a subject-visit specific imaging deviation. We describe strategies to identify baseline/longitudinal variation and registration errors combined with covariates of interest. Our analysis suggests that specific regional brain atrophy and ventricular enlargement are associated with multiple sclerosis (MS) disease progression.

Keywords: longitudinal functional principal component analysis, regional analysis of volumes examined in normalized space, voxel-based morphometry, multiple sclerosis, brain volume measurement

1. INTRODUCTION

Magnetic resonance imaging (MRI) is commonly used in the study of brain structure. Many studies are based on measurements of tissue volumes within a number of predefined regions of interest (ROIs); for example, see Bartzokis et al. (2001) and Bermel et al. (2003). Although ROI analysis can directly quantify the volume of structures and reduce the dimensionality of images, the ROIs have to be defined before the analysis is conducted. In disease studies, this can be difficult without sufficient prior knowledge about what and how various regions will be affected. Moreover, ROI based measurements can be time-consuming and labor-intensive. The results of the analysis will depend on the quality of the ROI delineation and thus depend upon the experience of the operator and accuracy of segmentation algorithms.

Voxel-based morphometry (VBM) is a complementary technique that measures local brain volumes in a normalized space and thus does not suffer from these limitations (Ashburner and Friston, 2000, 2001). In this work, we consider Regional Analysis of Volumes Examined in Normalized Space (RAVENS), which registers each subject brain to a template of anatomy so that the intensities of the RAVENS image represent regional volumes relative to those of template (Shen and Davatzikos, 2002). In voxel-based morphometry methods such as RAVENS, segmentations of structures such as the ventricles, are mapped to a template brain. If a subject's ventricles are larger than the template brain's ventricles, each voxel in the ventricles need to be shrunken to be mapped to the template. This in turn increases the intensity of the RAVENS map at each voxel, implying a larger volume was present in the subject at each voxel. **Figure 1** displays examples of ventricular RAVENS images in the template space. The first subject has much larger ventricles than the second subject (and template). Its RAVENS image of ventricles is displayed underneath the associated T1 image with red and blue colors representing higher and lower intensities, respectively. Subject 1 has larger ventricles, depicted by red in the RAVENS image. Similarly, the second brain, having a smaller ventricle than that of the first subject, has lower intensities in its RAVENS image, depicted by yellow and cyan in RAVENS image. By applying statistical VBM analysis of RAVENS images (RAVENS-VBM) to the resulting spatial distributions of gray matter (GM), white matter (WM), and ventricular cerebrospinal fluid (CSF), local atrophy or enlargement can be detected if the intensities significantly change across subjects.

In many disease studies, *longitudinal patterns* of brain structure between and within control and patient groups are of interest. Such studies are often based on ROI volume measurements followed by statistical analysis, such as a linear mixed model. Several neuroimaging software platforms, including: FSL (Smith et al., 2004), the SPM-VBM toolbox (available at http://dbm.neuro.uni-jena.de/vbm) and SurfStat (Worsley et al., 2009), support flexible longitudinal models. Statistical inference of the contrast between two different time points is the most commonly used approach (Bendfeldt et al., 2011). Numerous other approaches for longitudinal imaging data have been proposed for prediction. The methods include support vector machine classifiers (Chen and Bowman, 2011) and Bayesian spatial models (Derado et al., 2013).

In practice, there are frequently cases that VBM does not find significant longitudinal trend. Possible causes are (1) the chosen statistical method is not sophisticated enough to extract longitudinal information; (2) a substantial amount of visit-to-visit variation to longitudinal signals exists; (3) heterogeneous longitudinal patterns exist within the diseases population.

The obvious solution to overcome such limitations is to combine the VBM analysis with more sophisticated statistical methods such as linear mixed models. However, for the first two cases, hypothesis driven VBM analyses cannot further exploit the data. In that case, figuring out the underlying structures of variation in the longitudinal data would be of interest. Further, we want to quantify the longitudinal and cross-sectional variability, and the association between each subject and their spatial patterns.

Thus, our main goal is to introduce a new statistical framework for longitudinal VBM analysis. To achieve the goal, we consider a data-driven analysis to provide a more complete statistical framework to analyze high-dimensional longitudinal brain images. A framework to allow for this conceptual partition of variability is longitudinal functional principal component analysis (LFPCA; Greven et al., 2011). This method was originally proposed for low to moderate dimensional functional data and

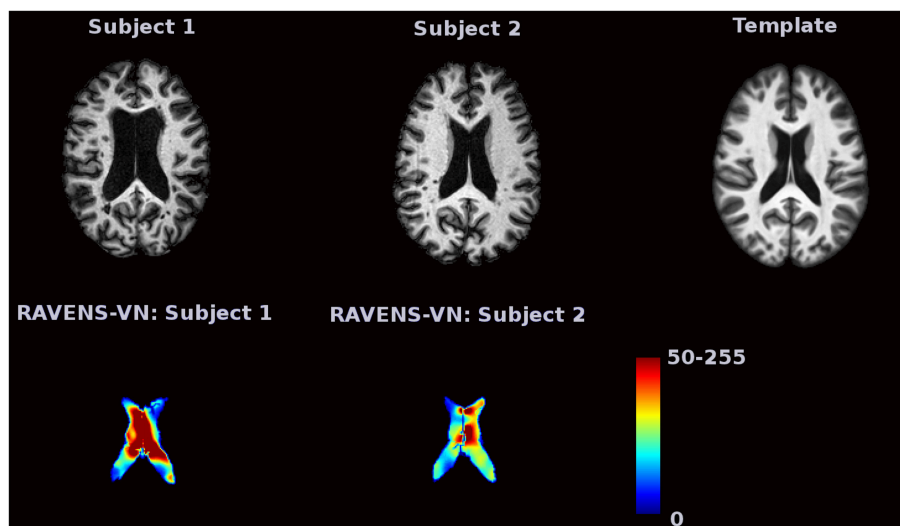**FIGURE 1 | The image intensities of the RAVENS image represent regional volumes relative to those of the template.** Red color represents high intensity and blue color represents lower intensity. The first brain, having a larger ventricle than the template brain, has brighter intensities in the RAVENS image. The second brain, which has smaller ventricles, has lower intensities in the associated RAVENS image.

was extended to high dimensional data by Zipunnikov et al. (2014). The main idea of high-dimensional inference is based on projecting onto the intrinsic low dimensional space spanned by high-dimensional vectors (Di et al., 2009; Zipunnikov et al., 2011b). More precisely, we start by modeling the observed data with high-dimensional longitudinal functional principal component analysis (HD-LFPCA). Each RAVENS ventricular image is unfolded into a $p \times 1$ dimensional vector, where $p \approx 80,000$ is the number of voxels in the RAVENS ventricular image. These vectors are decomposed in their baseline, longitudinal and visit-to-visit components; each component is then estimated from the data. The method takes only a few minutes on a standard PC.

In this paper we focus on LFPCA as a useful tool for longitudinal voxel-based analyses, particularly to quantify cross-sectional and longitudinal variability in the data. The simulation study illustrates the application of LFPCA to a simplified imaging setting. It demonstrates that LFPCA effectively separates longitudinal, cross sectional, and other variations. Notably, the simulation study shows that LFPCA can separate registration errors from baseline and longitudinal components of interest.

2. MATERIALS AND METHODS

2.1. Participants

Forty eight MS patients (aged 42 ± 12 years at baseline) were enrolled in a longitudinal study of brain volume change. The study population included 33 female and 16 male patients; 28 patients with relapsing-remitting MS (RRMS), 13 patients with secondary progressive MS (SPMS), 5 patients with primary progressive MS (PPMS) and 2 patients with clinically isolated syndrome (CIS). One hundred forty eight T1 images have been acquired, with three images per subject for 44 subjects and 4 images per subject for 3 subjects. The average time interval between scans was 368 days (± 27). All images were spatially normalized via registration of T1 maps into the mean template, generated using Advanced Normalization Tools (Avants et al., 2010, 2011) from 30 randomly chosen MS patients among those with more than three visits. Ethical approval for the study was granted by IRB-2 and Johns Hopkins Medicine Institutional Review Board. All participants signed their fully informed consent.

2.2. MRI Protocol and Image Analysis

High resolution 3D magnetization-prepared rapid acquisition of gradient echoes (MPRAGE; acquired resolution: $1.1 \times 1.1 \times 1.1$ mm; TR:~10 ms; TE: 6 ms; TI = 835 ms; flip angle: 8°; SENSE factor:2; averages:1) were acquired on a 3.0 T MRI scanner (Intera, Philips Medical Systems).

In the processing, the follow-up images are affinely registered to their baselines via FMRIB's Linear Image Registration Tool (Jenkinson et al., 2002). All T1 images were segmented into GM, WM, VN, and lesions with Lesion-TOADS (Shiee et al., 2010) that was specifically designed for tissue and MS lesion segmentation. In general, as MS progresses, multifocal lesions in the white matter develop, and newly developed legions can cause inaccuracies in the registration and RAVENS map computation.

Thus, we masked those lesions in the registration using the Lesions-TOADS software. After segmentation, the final tissue maps of GM, WM, and VN were normalized using HAMMER-SUITE (Shen and Davatzikos, 2002) to generate RAVENS images. Finally, the RAVENS maps were separately smoothed with 4 mm FWHM using SPM8.

2.3. Longitudinal Functional Principal Component Analysis

In this section, we provide a description of the original LFPCA approach developed by Greven et al. (2011) and its extension for high-dimensional data analysis (Zipunnikov et al., 2014). Throughout this section, we refer to both as LFPCA.

2.3.1. Random Intercept and Random Slope Model

Consider a longitudinal brain imaging study with subjects labeled by index i with each visit indexed by j and scan time by variable t_{ij} for $j = 1, \ldots, J_i$. Each image is unfolded into a p-dimensional column vector $\mathbf{y}_{ij}(v)$; the index v of each entry corresponds to a particular location in the brain for each subject and visit in normalized space. A random slope and random intercept model is commonly used to analyze longitudinal data, and it has been extended to functional (Greven et al., 2011) and imaging (Zipunnikov et al., 2014) studies as follows:

$$y_{ij}(v) = \eta(v, t_{ij}) + x_{i,0}(v) + x_{i,1}(v)t_{ij} + W_{ij}(v), \qquad (2.1)$$

where $y_{ij}(v)$ denotes the image intensity at voxel v, $\eta(v, t_{ij})$ is a fixed main effect, and $x_{i,0}(v)$ and $x_{i,1}(v)$ denote the random intercept and random slope for subject i, respectively. The term $W_{ij}(v)$ is a random subject-visit specific imaging deviation, which is assumed to be a zero mean, second-order stationary random process uncorrelated with $\mathbf{X}_i(v) = \left(x_{i,0}(v), x_{i,1}(v) \right)^\top$. The covariance operators of $\mathbf{X}_i(v)$ and $W_{ij}(v)$ are denoted as $\mathbf{K}^X(v_1, v_2)$ and $\mathbf{K}^W(v_1, v_2)$, respectively.

While this is a natural and relatively simple model for longitudinally observed data, the scale of the problem requires aggressive dimensionality reduction. LFPCA reduces dimensionality by projecting onto the subspaces which explain principal directions of variation in the data. In model (2.1), there are two sources of variation: subject-to-subject, captured by $\mathbf{X}_i$, and visit-to-visit within a subject, captured by $\mathbf{W}_{ij}$ and the model assumption on $\mathbf{X}_i$ and $\mathbf{W}_{ij}$ in (2.1) allows us to partition the variation of the data and LFPCA models latent processes $\mathbf{X}_i$ and $\mathbf{W}_{ij}$ using a Karhunen-Loeve (K-L) expansion (Karhunen, 1947; Loève, 1948).

The K-L expansion decomposes the two latent processes as $\mathbf{X}_i(v) = \sum_{k=1}^{\infty} \xi_{ik} \boldsymbol{\phi}_k^X(v)$ and $W_{ij}(v) = \sum_{l=1}^{\infty} \zeta_{ijl} \phi_l^W(v)$, where $\boldsymbol{\phi}_k^X = \left(\phi_k^{X,0}, \phi_k^{X,1} \right)$ and ϕ_l^W are the eigenfunctions of $\mathbf{K}^X(v_1, v_2)$ and $\mathbf{K}^W(v_1, v_2)$, respectively, such that

$$\mathbf{K}^X(v_1, v_2) = \begin{pmatrix} \mathbf{K}_{00}^X(v_1, v_2) & \mathbf{K}_{10}^X(v_1, v_2) \\ \mathbf{K}_{01}^X(v_1, v_2) & \mathbf{K}_{11}^X(v_1, v_2) \end{pmatrix}$$
$$= \sum_{k=1}^{N_X} \lambda_k^X \boldsymbol{\phi}_k^X(v_1) \left\{ \boldsymbol{\phi}_k^X(v_2) \right\}^\top.$$

LFPCA truncates K-L representations and represents observed data through a linear mixed-effects model:

$$y_{ij}(v) = \eta(v, t_{ij}) + \sum_{k=1}^{N_X} \xi_{ik} \phi_k^{X,0}(v) + t_{ij} \sum_{k=1}^{N_X} \xi_{ik} \phi_k^{X,1}(v)$$
$$+ \sum_{l=1}^{N_W} \zeta_{ijl} \phi_l^W(v),$$
$$(\xi_{ik_1}, \xi_{ik_2}) \sim (0, 0, \lambda_{k_1}^X, \lambda_{k_2}^X, 0); (\zeta_{il_1}, \zeta_{il_2}) \sim (0, 0, \lambda_{l_1}^W, \lambda_{l_2}^W, 0),$$
$$(2.2)$$

where "$\cdot \sim (\mu_1, \mu_2; \sigma_1^2; \sigma_2^2; \rho)$" denotes that a pair of variables has a distribution with mean (μ_1, μ_2), variance σ_1^2, σ_2^2, and correlation ρ. We assume that $\lambda_{k_1} \geq \lambda_{k_2}$ if $k_1 \leq k_2$. Since $\mathbf{X}_i(v)$ and $W_{ij}(v)$ are uncorrelated, the scores $\{\xi_{ik}\}_{k=1}^{\infty}$ and $\{\zeta_{ijl}\}_{l=1}^{\infty}$ are also uncorrelated. A very important characteristic of model (2.2) is that both N_X and N_W are expected to be small in most applications.

For the unfolded vector, (2.2) can be rewritten as $\mathbf{y}_{ij} = \boldsymbol{\eta}(t_{ij}) + \boldsymbol{\Phi}_X^0 \boldsymbol{\xi}_i + t_{ij} \boldsymbol{\Phi}_X^1 \boldsymbol{\xi}_i + \boldsymbol{\Phi}_W \boldsymbol{\zeta}_{ij}$, where $\mathbf{y}_{ij} = (y_{ij}(v_1), \ldots, y_{ij}(v_p))^\top$ is a $p \times 1$ dimensional vector; $\boldsymbol{\phi}_k^{X,0}, \boldsymbol{\phi}_k^{X,1}$, and $\boldsymbol{\phi}_l^W$ are $p \times 1$ eigenvectors; $\boldsymbol{\Phi}_X^s = (\boldsymbol{\phi}_1^{X,s}, \ldots, \boldsymbol{\phi}_{N_X}^{X,s})$ for $s = 0, 1$; $\boldsymbol{\Phi}_W = (\boldsymbol{\phi}_1^W, \ldots, \boldsymbol{\phi}_{N_W}^W); \boldsymbol{\xi}_i = (\xi_{i1}, \ldots, \xi_{iN_X})^\top; \boldsymbol{\zeta}_i = (\zeta_{i1}, \ldots, \zeta_{iN_W})^\top$.

In brain imaging data analysis, LFPCA can separate biological signals from non-biological artifacts. For example, registration errors due to structural differences between subjects can be captured by baseline subject-specific components $\boldsymbol{\Phi}_X^0$ and scanner variability can be captured by visit-to-visit components $\boldsymbol{\Phi}_W$. This will be illustrated via an extensive simulation experiment in Section 3.1.

The fixed effect $\eta(v, t_{ij})$ can be estimated in a number of ways (Greven et al., 2011). The analyses in the later sections simply use the sample mean across all the image observations. Once $\eta(v, t_{ij})$ is estimated by the sample mean $\tilde{\eta}(v, t_{ij})$, the longitudinal eigenanalysis is applied to the residual images $\tilde{y}_{ij}(v) = y_{ij}(v) - \tilde{\eta}(v, t_{ij})$ that are modeled as follows:

$$\tilde{\mathbf{y}}_{ij} = \boldsymbol{\Phi}_X^0 \boldsymbol{\xi}_i + t_{ij} \boldsymbol{\Phi}_X^1 \boldsymbol{\xi}_i + \boldsymbol{\Phi}_W \boldsymbol{\zeta}_{ij}. \tag{2.3}$$

2.3.2. LFPCA Estimation

Zipunnikov et al. (2014) modified the original approach of Greven et al. (2011) and developed a method of moments estimator based on quadratics of $\tilde{\mathbf{y}}_{ij}$. The $p \times p$-covariance of $\tilde{\mathbf{y}}_{ij_1}$ and $\tilde{\mathbf{y}}_{ij_2}$ is given by

$$\mathsf{E}\left\{\tilde{\mathbf{y}}_{ij_1} \tilde{\mathbf{y}}_{ij_2}^\top\right\} = \mathbf{K}_{00}^X + t_{ij_1} \mathbf{K}_{10}^X + t_{ij_2} \mathbf{K}_X^{10} + t_{ij_1} t_{ij_2} \mathbf{K}_X^{11} + \delta_{j_1, j_2} \mathbf{K}^W,$$
$$j_1, j_2 = 1, \ldots, J_i, \tag{2.4}$$

where $\delta_{i,j} = 1$ if $i = j$ and $\delta_{i,j} = 0$ otherwise. Model (2.4) can be rewritten in terms of unfolded vectors $\mathbf{K}^v = \{\text{vec}\mathbf{K}_{00}, \text{vec}\mathbf{K}_{01}, \text{vec}\mathbf{K}_{10}, \text{vec}\mathbf{K}_{11}, \text{vec}\mathbf{K}^W\}$ and $\mathbf{f}_{ij_1 j_2} = (1, t_{ij_2}, t_{ij_1}, t_{ij_1} t_{ij_2}, \delta_{j_1, j_2})^\top$ such that $\mathsf{E}\text{vec}\tilde{\mathbf{y}}_{ij_1} \tilde{\mathbf{y}}_{ij_2}^\top = \mathbf{K}^v \mathbf{f}_{ij_1 j_2}$. By concatenating all vectors across all subjects and visits we obtain a

moment matrix identity for the $p^2 \times m$ matrix $\mathbf{Y}$: $\mathsf{E}\mathbf{Y} = \mathbf{K}^v \mathbf{F}$, where $m = \sum_{i=1}^N J_i^2$. Then covariance parameters $\mathbf{K}^v$ can be unbiasedly estimated by using ordinary least squares (OLS): $\widehat{\mathbf{K}}^v = \mathbf{Y}\mathbf{F}^\top (\mathbf{F}\mathbf{F}^\top)^{-1}$.

The covariance operators $\mathbf{K}^X$ and $\mathbf{K}^W$ are $2p \times 2p$ and $p \times p$ dimensional, respectively. For high-dimensional functional data, storing or diagonalizing these matrices is not feasible. Zipunnikov et al. (2014) proposed HD-LFPCA, a novel estimation approach that takes advantage of an intrinsically small dimension of the space spanned by high-dimensional data vectors. First we form a $p \times J_i$ dimensional matrix $\tilde{\mathbf{y}}_i$, where column j corresponds to a demeaned-RAVENS image obtained for subject i at visit j. The $p \times J$ dimensional data matrix $\tilde{\mathbf{y}} = (\tilde{\mathbf{y}}_1; \ldots; \tilde{\mathbf{y}}_n)$ is formed by column-binding the blocks of data corresponding to each subject, where $J = \sum_{i=1}^N J_i$. The data matrix can be decomposed as $\tilde{\mathbf{y}} = \mathbf{VSU}^\top$ using a singular value decomposition (SVD) approach. In the RAVENS image application, $J = 148$. Equation (2.3) can be rewritten as

$$\tilde{\mathbf{y}}_{ij} = \mathbf{VSU}_{ij} = \boldsymbol{\Phi}_X^0 \boldsymbol{\xi}_i + t_{ij} \boldsymbol{\Phi}_X^1 \boldsymbol{\xi}_i + \boldsymbol{\Phi}_W \boldsymbol{\zeta}_{ij}. \tag{2.5}$$

By multiplying with $\mathbf{V}^\top$ to the left, we have

$$\mathbf{SU}_{ij} = \mathbf{V}^\top \boldsymbol{\Phi}_X^0 \boldsymbol{\xi}_i + t_{ij} \mathbf{V}^\top \boldsymbol{\Phi}_X^1 \boldsymbol{\xi}_i + V^\top \boldsymbol{\Phi}_W \boldsymbol{\zeta}_{ij}$$
$$= \mathbf{A}_X^0 \boldsymbol{\xi}_i + \mathbf{A}_X^1 \boldsymbol{\xi}_i + \mathbf{A}_W \boldsymbol{\zeta}_{ij}. \tag{2.6}$$

We estimate $\hat{\mathbf{A}}_X^0, \hat{\mathbf{A}}_X^1$, and $\hat{\mathbf{A}}_W$ as described earlier, and estimate $\hat{\boldsymbol{\Phi}}_X^0 = \mathbf{V}\hat{\mathbf{A}}_X^0, \hat{\boldsymbol{\Phi}}_X^1 = \mathbf{V}\hat{\mathbf{A}}_X^1$, and $\hat{\boldsymbol{\Phi}}_W = \mathbf{V}\hat{\mathbf{A}}_W$. Note that multiplying by $\mathbf{V}^\top$ in Equation (2.5) reduces the model to its low-dimensional form (2.6), without losing the original correlation structure of the data. Once inference is conducted in model (2.6), then quantities of interest from model (2.5) can be estimated by pre-multiplying Equation (2.6) by $\mathbf{V}$.

Principal scores $\boldsymbol{\xi}_i$ and $\boldsymbol{\zeta}_{ij}$ are estimated via Best Linear Unbiased Predictions (BLUPs) as follows. The stacked vector of ith subject data, $\text{vec}\tilde{\mathbf{y}}_i = (\tilde{\mathbf{y}}_{i1}^\top, \ldots, \tilde{\mathbf{y}}_{iJ_i}^\top)^\top$, can be rewritten as $\text{vec}\tilde{\mathbf{y}}_i = \mathbf{B}_i \boldsymbol{\omega}_i$, where $\mathbf{B}_i = (\mathbf{B}_i^X; \mathbf{B}_i^W), \mathbf{B}_i^X = \mathbf{1}_{J_i} \otimes \boldsymbol{\Phi}_X^0 + \mathbf{T}_i \otimes \boldsymbol{\Phi}_X^1$, $\mathbf{B}_i^W = \mathbf{I}_{J_i} \otimes \boldsymbol{\Phi}_W$, where $\mathbf{T}_i = (t_{i1}, \ldots, t_{iJ_i})^\top, \boldsymbol{\omega}_i = (\boldsymbol{\xi}_i^\top, \boldsymbol{\zeta}_i^\top)^\top$, the subject level principal scores $\boldsymbol{\zeta}_i = (\boldsymbol{\zeta}_{i1}^\top, \ldots, \boldsymbol{\zeta}_{iJ_i}^\top)^\top$, and $\mathbf{1}_{J_i}$ is a $J_i \times 1$ vector of ones. Then the scores can be estimated as $\hat{\boldsymbol{\omega}}_i = (\hat{\mathbf{B}}_i^\top \hat{\mathbf{B}}_i)^{-1} \hat{\mathbf{B}}_i^\top \text{vec}\tilde{\mathbf{y}}_i$. Due to linearity the estimated scores are the same in both models (2.5) and (2.6). Details of the matrix calculation and additional theoretical results of HD-LFPCA can be found in Zipunnikov et al. (2014).

The computed subject-specific principal component scores $\boldsymbol{\xi}_i$ are the derived composite scores computed for each linear trajectories based on the eigenvectors for subject-specific PCs. These scores can be used as predictors or outcomes in subsequent regression analyses to evaluate relationships between high-dimensional longitudinal trajectories and other variables of interest. Also, we can apply cluster analysis on the scores to uncover latent structure in the sample.

2.4. Classical VBM Analysis using Linear Mixed Model

First, we applied traditional VBM analysis using a linear mixed model to find a longitudinal trend. Many previous longitudinal studies have applied pairwise comparisons between two time points (Driemeyer et al., 2008). This study attempts to discover constant longitudinal trends over the time, i.e., focusing on the atrophy or enlargement rates. This may elucidate disease progression patterns of the patients. For the ith subject jth visit, the RAVENS image at voxel v follows the model:

$$y_{ij}(v) = \beta^0(v) + \beta^1(v)t_{ij} + b_i^0 + b_i^1 t_{ij} + \epsilon_{ij}(v),$$
$$b_i^0 \sim N(0, \sigma_0^2(v)), \ b_i^1(v) \sim N(0, \sigma_1^2(v)), \qquad (2.7)$$
$$\text{Cov}(b_i^0(v), b_i^1(v)) = \sigma_{12}, \ \epsilon_{ij}(v) \sim N(0, \sigma_\epsilon^2(v)),$$

where $\beta^0(v)$ and $\beta^1(v)$ are the fixed-effect coefficients, b_i^0 and $b_i^1(v)$ are the random-effect coefficients for subject i, $\epsilon_{ij}(v)$ is the error. The parameters are estimated based on maximum likelihood estimation and the p-values of the fixed effect parameters are computed controlling for false discovery rate using (Benjamini and Yekutieli, 2001). We perform the statistical analysis in R (version 2.15.1).

3. RESULTS

3.1. Simulated Images

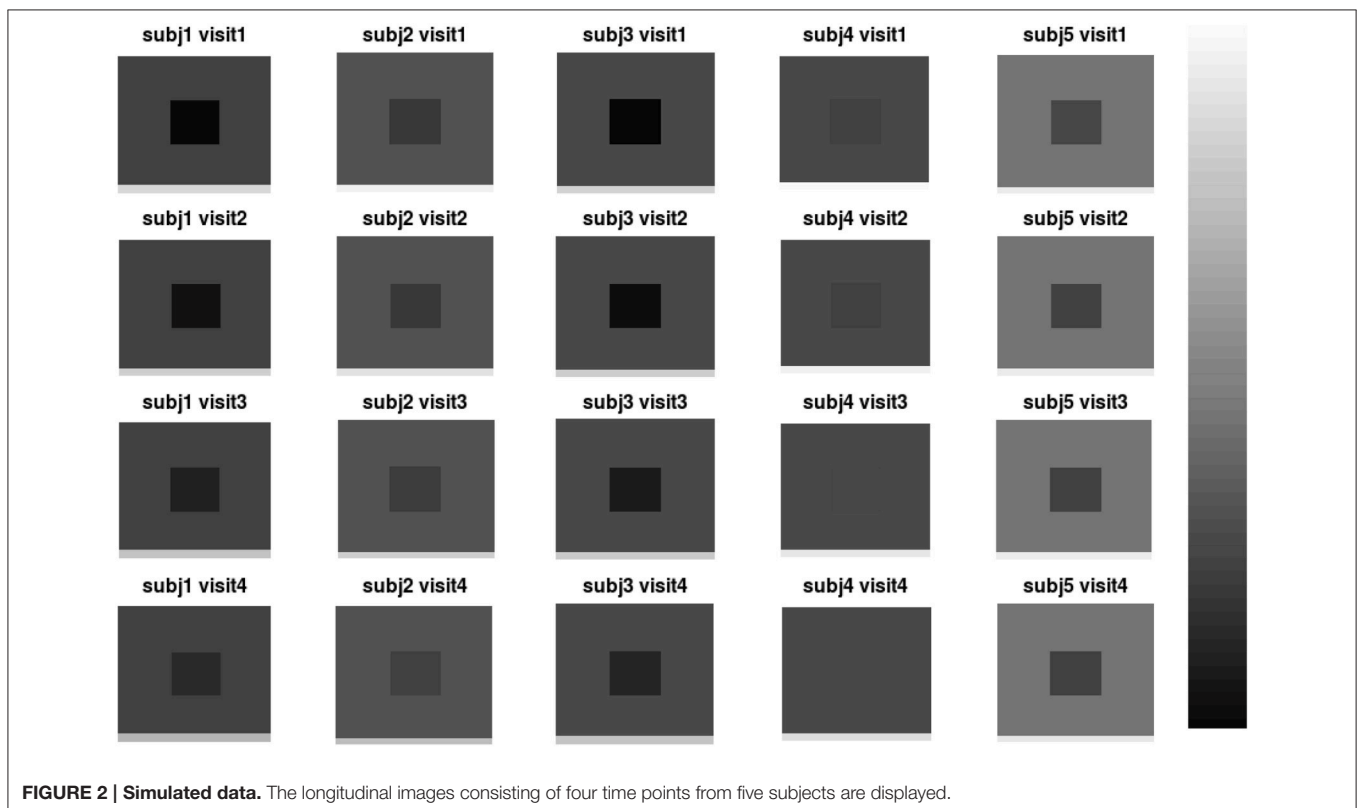In this section, we present a simulation study to test the performance of LFPCA in RAVENS-VBM analysis. We investigate if LFPCA can identify subject-specific signals from noise, particularly registration errors, which often dominate signals in VBM analyses. Also, we identify cross-sectional and longitudinal variation when they exist.

We design a simulation study to mimic longitudinal analysis of RAVENS images. For the purpose of illustration, we use 2D images with $200 \times 200 = 40,000$ pixels. We generate images from 50 subjects ($N = 50$) with three follow-ups. To replicate RAVENS processing routine, we assume that all images are registered to a template space. **Figure 2** displays simulated RAVENS images from 5 randomly chosen subjects. Each column represents four longitudinally collected images of the same subject.

Each image mimics four canonical brain structures: background (B), white matter (W), ventricles (V), and gray matter (G). Those four components are simplified and shown as a background, a big square, a small square inside the big square, and a rectangle at the bottom, respectively. Registration errors are introduced via random rigid shifts of simulated structures as described below.

In **Figure 2**, the images from the first subject, which are displayed at the first column, show the longitudinal patterns. In the images, the color of V changes from darker gray to brighter gray, which represents longitudinal enlargements of V. Similarly, the colors of W and G changed to darker colors, which represent longitudinal atrophy.

Figure 3 shows the first five pairs of subject-specific components (Φ_X). The baseline components (Φ_X^0) are displayed in the top row and their corresponding longitudinal components (Φ_X^1) are displayed in the second row. Each image is colored with
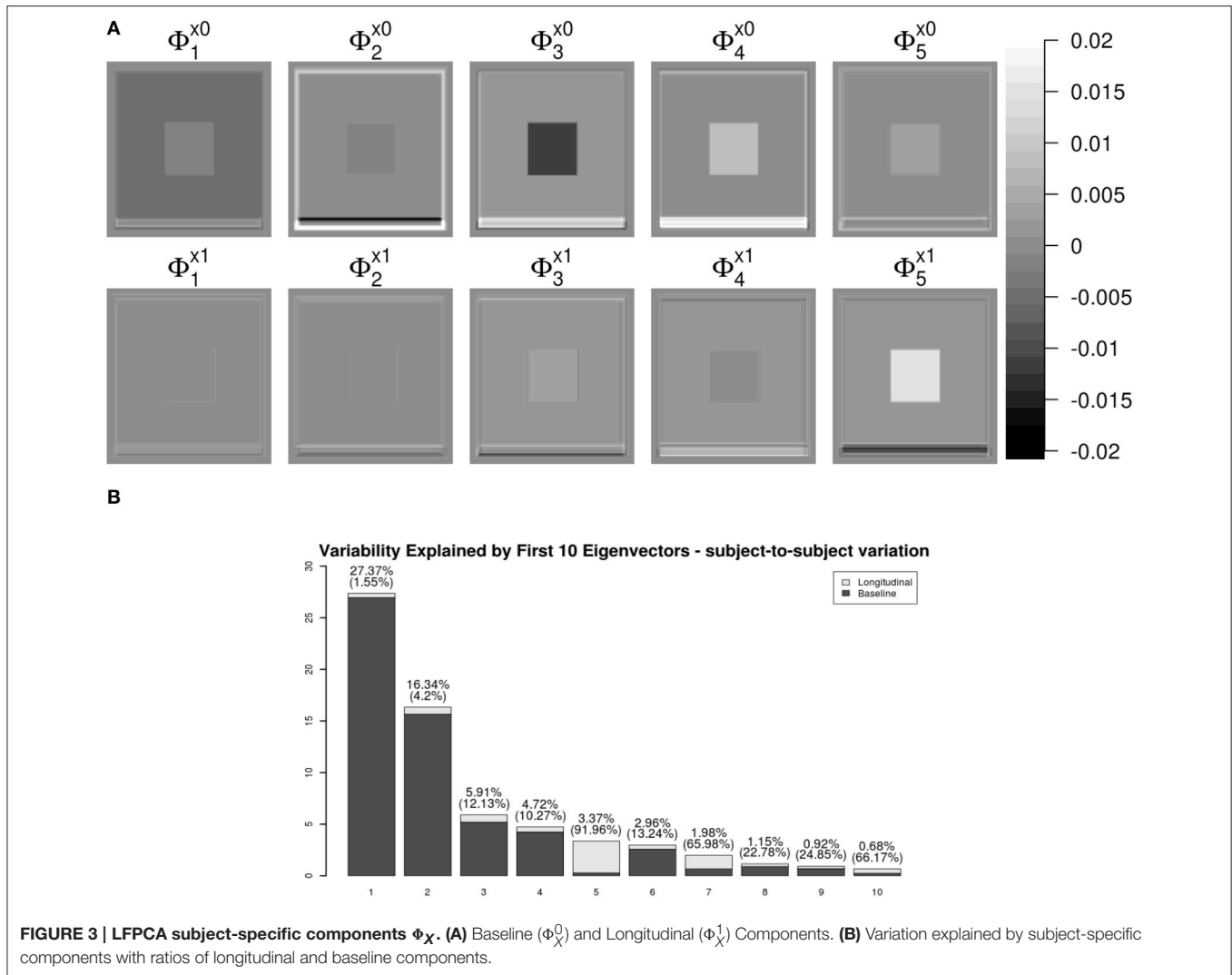

FIGURE 2 | Simulated data. The longitudinal images consisting of four time points from five subjects are displayed.

FIGURE 3 | LFPCA subject-specific components Φ_X. (A) Baseline (Φ_X^0) and Longitudinal (Φ_X^1) Components. **(B)** Variation explained by subject-specific components with ratios of longitudinal and baseline components.

a black(negative)-gray(0)-white(positive) color scheme. The first subject-specific component (**Figure 3A** first column) represents cross-sectional variations of the intensities of the W. The second component captures subject-specific registration errors, which only depend on cross-sectional variation. The third and fourth components represent the size of V and G. The fifth component shows longitudinal patterns of V and G. For a subject with positive score, the area V enlarges over the time and that of G shrinks, matching the truth used in simulation.

One useful feature of LFPCA is that contributions of the longitudinal and baseline components within each subject-specific component can be quantified on a $[0, 1]$ scale. A subject-specific eigenvector is the stacked vector of baseline and longitudinal components: $\boldsymbol{\Phi}_{X,k} = \left\{ \boldsymbol{\Phi}_{X,k}^{0}{}^{\top}, \boldsymbol{\Phi}_{X,k}^{1}{}^{\top} \right\}^{\top}$, such that $\|\boldsymbol{\Phi}_{X,k}\|^2 = \|\boldsymbol{\Phi}_{X,k}^0\|^2 + \|\boldsymbol{\Phi}_{X,k}^1\|^2 = 1$. For each component, the variation or the contribution of the longitudinal component can be calculated as $\frac{\|\boldsymbol{\Phi}_{X,j}^1\|^2}{\|\boldsymbol{\Phi}_{X,k}^0\|^2 + \|\boldsymbol{\Phi}_{X,k}^1\|^2}$. Combined with the contribution of each subject-specific component to

the total variation, **Figure 3B** displays variations explained by the first 10 subject-specific components with the proportion of the longitudinal components within each subject. Each bar plot intensity represents the amount of variation explained by each subject-specific component and is comprised of variations explained by the longitudinal component (dark) and the baseline component (bright). The top of each bar displays numerical values of the variation explained by the subject-specific component with the variation explained by the longitudinal component within the subject-specific component in parenthesis. Note that the fifth principal component has the highest longitudinal-baseline ratio among all 10 components. This provides a strong indication that the fifth component should be essentially treated as a longitudinal component. Using both visual and quantitative methods, we can conclude that the first four components represent baseline variation and registration error and the fifth component reveals longitudinal variation. In the data set, the longitudinal variation and baseline variations are independent, which agrees with the simulation setting.

An advantage of LFPCA is its ability to couple baseline and longitudinal variation. The longitudinal component is added to the baseline with the time used as a multiplicative weight. **Figure 4** illustrates the temporal trajectory of principal component loadings. We display only the first component, which does not appear to change over time. This pattern is replicated in the first four components. This indicates that the first four components mostly represent baseline variation. The fifth component loading does appear to change over time, while the baseline loading has relatively lower intensities compared to the longitudinal loading.

To summarize, our simulation studies convincingly demonstrate the power and flexibility of LFPCA to address some of key challenges of brain imaging. In particular, LFPCA managed to estimate and separate longitudinal and cross-sectional variation in a complex imaging simulation design with registration errors. The main part of the analysis can be

automated and performed robustly with no operator input. We also applied a classical VBM-linear mixed effect model for the simulated data. As we expected, the linear mixed effect model could identify linear trend in the ventricular area (V), but it did not find significant trend in other areas (W and G) due to low longitudinal changes in signal and high visit-to-visit variation.

3.2. Classical VBM Analysis using Linear Mixed Model

In this section, we apply a standard VBM analysis to the MS cohort described in Section 2. This analysis focuses on the population mean of the longitudinal trend $\beta_1^1(v)$. After an FDR correction (Benjamini and Yekutieli, 2001) combined with cluster level thresholding, there are significant clusters with spatial extent more than 20 voxels. **Table 1** shows information about the significant clusters, including cluster size, maximum or minimum t-values within each
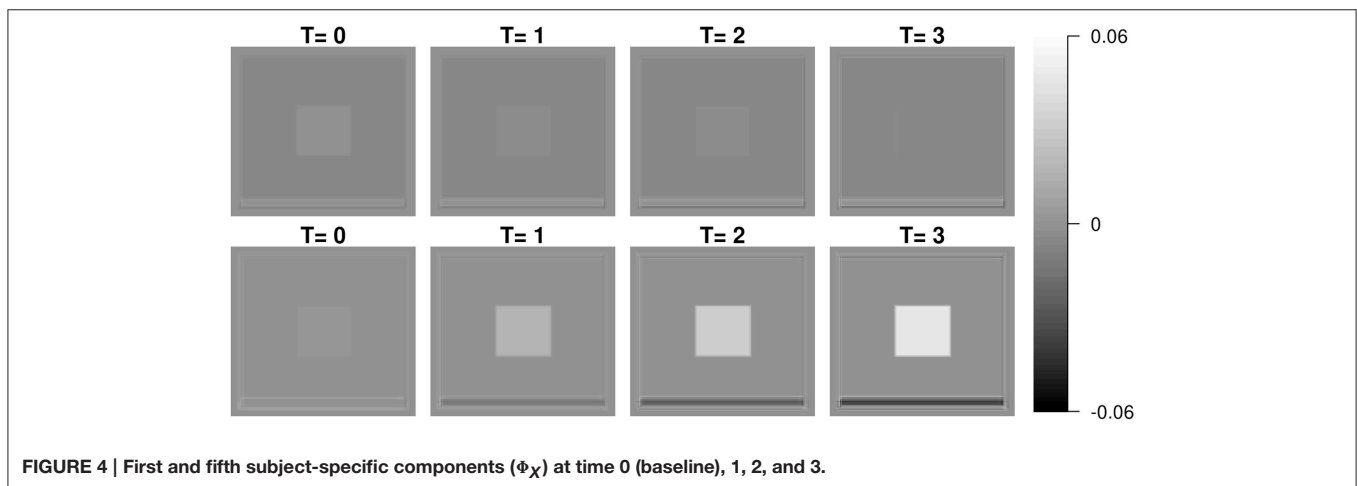

FIGURE 4 | First and fifth subject-specific components (Φ_X) at time 0 (baseline), 1, 2, and 3.

TABLE 1 | Significant clusters of GM/WM/VN VBM results.

		Cluster size	t-MAX[a] (or MIN)	t-MAX(or MIN)[b]			t-COG[c]		
				X	Y	Z	X	Y	Z
GM	Atrophy	112	−6.34	139	86	67	139	90	68.9
		79	−6.36	118	158	35	116	158	35.0
		40	−5.84	122	166	41	121	166	41.0
		28	−6.13	136	148	80	136	149	79.8
		15	−5.36	143	167	89	143	166	89.3
VN	Enlargement	111	5.59	161	151	70	157	152	72.4
WM	Enlargement	154	6.30	118	150	76	117	150	75.6
		100	5.75	127	111	85	128	111	84.8
	Atrophy	210	−5.97	126	83	59	124	83.7	57.6
		157	−5.77	106	98	83	107	98	83.5
		31	−5.46	118	154	37	118	154	36.7

All clusters are small (< 250) and spatially scatter.
[a]*Maximum t-value for positive values and minimum t-value for negative values.*
[b]*Location of maximum (minimum) Z-value (Z-MAX(MIN)).*
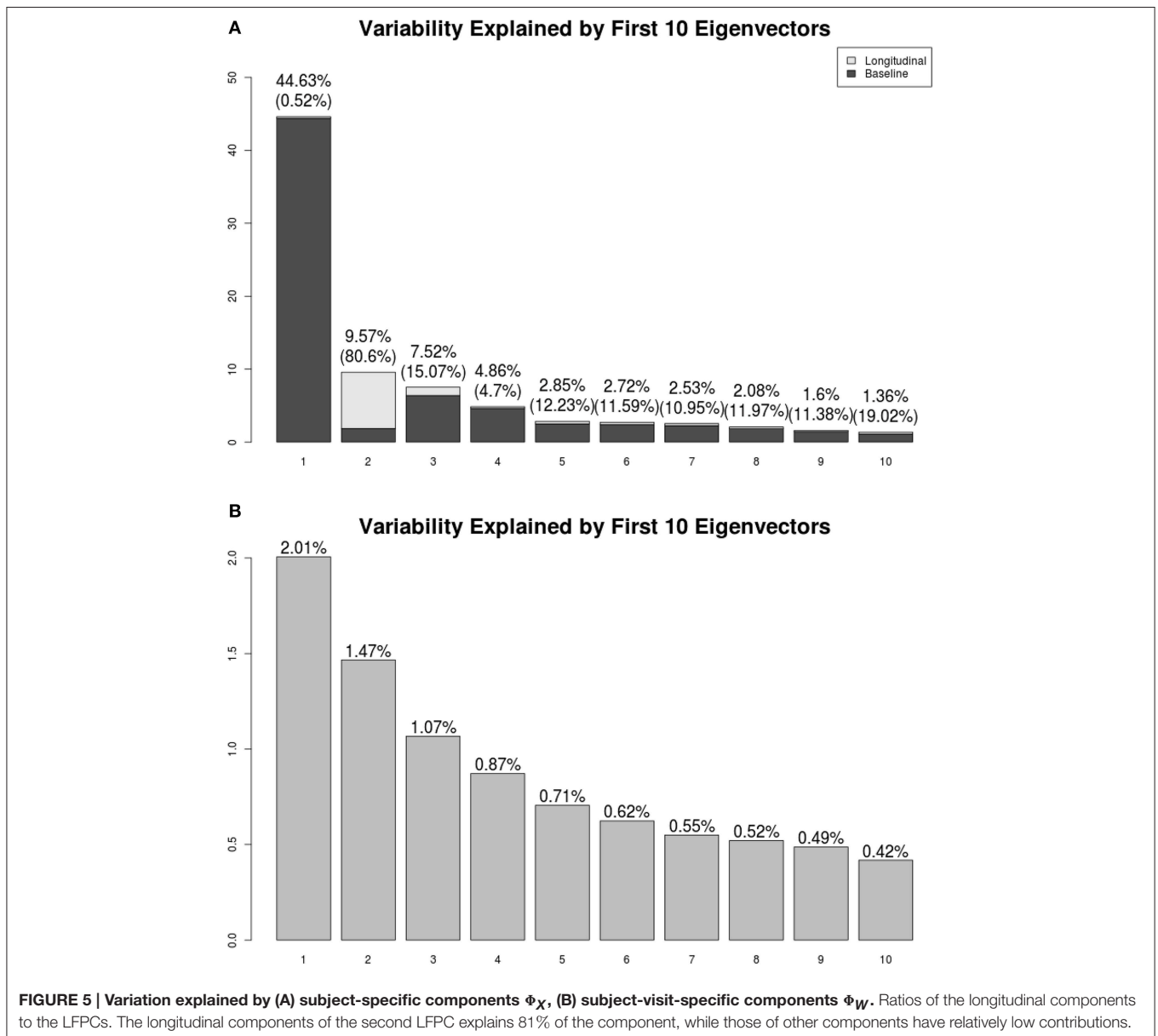[c]*Centre of gravity (COG) of the cluster.*

cluster and its location, and center of gravity of the clusters. We do not include an image of VBM results, since the clusters are very sparse having small cluster sizes. The results show that the longitudinal patterns do not appear to be significant for the most of brain regions. We suspect that it is because of the large variation within and between images due to real anatomic variation as well as registration error. We apply HD-LFPCA to uncover more subtle underlying variation.

3.3. HD-LFPCA Results

We present the LFPCA results for ventricular RAVENS images in **Figure 5**. **Figures 5A,B** display the amount of variation explained by subject-specific components and subject and visit-specific components to the total variation, respectively. **Figure 5A** displays variation explained by the first 10 subject-specific

components along the proportion of longitudinal variation represented within each subject-specific component. Each bar's height represents the percent of variation explained by each subject-specific component. It is color coded by the proportion of the variation explained by the longitudinal component and the baseline component. The top of each bar displays the variation explained by the subject-specific component; the fraction of that variation that is explained by the longitudinal component is given in parentheses.

The first subject-specific LFPC explains 45% of the overall variation, almost completely due to the cross-sectional part. The longitudinal part explains 81% of the variation within the second subject-specific LFPC. **Figure 5B** displays variation explained by the first 10 subject-visit-specific components to the total variation. The remaining LFPCs explain less than 1% of the total variation.

FIGURE 5 | Variation explained by (A) subject-specific components Φ_X, (B) subject-visit-specific components Φ_W. Ratios of the longitudinal components to the LFPCs. The longitudinal components of the second LFPC explains 81% of the component, while those of other components have relatively low contributions.

Most of the subject-specific LFPCs are driven by cross-sectional variation, which possibly include registration errors. The longitudinal changes are mainly captured by the second LFPC, which explains about 8% of the total variation. This provides an explanation as to why traditional VBM using linear mixed models did not find meaningful longitudinal patterns.

Figures 6, **7** shows the first two pairs of LFPCs of ventricles. **Figures 6A,B** show the baseline and longitudinal components of the first LFPC. The LFPC loadings are color-coded with a red-yellow color scheme for positive values and blue-cyan for negative values. The first components reveal baseline ventricular morphometric variation, while the longitudinal component has relatively lower intensities. To investigate the characteristics of the first component, we fit the linear regression with covariates of interest and volumes of 6 ROIs obtained by the Lesion-TOADS segmentation algorithm. **Figure 7C** displays scatter plots of the LFPC scores with covariates, baseline age, baseline Expanded Disability Status Scale (EDSS) score, and volumes of 6 ROIs (thalamus, ventricle, cortical gray matter volume, caudate, sulcal CSF, putamen). The dashed lines overlaid on the scatter plots are the linear regression lines and are colored as red when the linear trend is significant and colored as green otherwise.

The significant correlation between the first subject-specific score and baseline VN volume (first row, fourth column) confirms that the first component represents baseline variation (R^2: 0.9684), i.e., a subject with a positive score has larger ventricles at the baseline. The scores are significantly correlated to the subject's baseline age (R^2: 0.1402) and three gray matter ROIs (thalamus, caudate, and putamen).

Figures 7A,B display the baseline and longitudinal components of the second subject-specific component, respectively. A subject with a positive score tends to have a larger regional volume at the baseline (yellow-red colored voxels in **Figure 7A**) and longitudinal enlargement. The second subject-specific scores have significant correlations with the baseline age, EDSS, thalamus volume, VN volumes, sulcal CSF volume and putamen volume. The second component scores have higher correlation with baseline age (R^2: 0.2371) and EDSS (R^2: 0.2053) than the first component scores that represent cross-sectional variation. This indicates that the spatial patterns of longitudinal enlargement in ventricles are superior for modeling disease progression and age compared to simple ventricular volume measures.

We have applied a similar analysis to gray matter and white matter RAVENS images. **Figure 8** shows variation explained by first 10 subject-specific LFPCs in gray matter and white matter RAVENS images. In gray matter, around 20% variation comes from the longitudinal part across all subject-specific LFPCs. Lower proportions of variation, around 15%, are explained by the longitudinal part in white matter. Unlike the ventricles, any subject-specific component of gray and white matter is not dominated by the longitudinal part. We speculate it is due to spatial heterogeneity of longitudinal brain atrophy, which may depend on subject-specific disease progression. In the correlation analyses with age and EDSS scores, the first LFPC component of the gray matter was significantly associated with age ($r = -0.48$, $p < 0.001$) and EDSS ($r = -0.57$, $p < 0.001$) indicating gray

matter thinning is highly associated with age and MS progression, while other components were not significantly associated with age or EDSS. For white matter, LFPC1 was not significantly associated with age ($r = -0.07$, $p = 0.63$) but with EDSS ($r = -0.32$, $p = 0.03$). LFPC2 was significantly associated with both age ($r = -0.36$, $p = 0.01$) and EDSS ($r = -0.34$, $p = 0.02$). Although those LFPC components did not contain substantial longitudinal changes, the results indicate that local atrophy in the white matter can inform disease progression.

As described above, LFPCA is a useful dimension reduction tool for high-dimensntional longitudinal data. In this section, we illustrated how the LFPC scores an be used in the correlation analyses. Further, LFPCA scores can be used as predictors or outcomes in regression analyses, classification or cluster analysis.

4. DISCUSSION

In this manuscript, we described and evaluated a coherent methodology for the study of longitudinal RAVENS—or other methodological—volumetric imaging studies. Our simulation studies demonstrate that LFPCA tightly links the analysis methodology with the morphometric image processing stream. We demonstrated that LFPCA can uncover interesting, yet subtle, directions of longitudinal variation in a case where independent voxel-level investigations fail. Of note, this study represents the first application of the high dimensional variation of LFPCA to voxel-based morphometric analysis. Related work includes Zipunnikov et al. (2014), who investigated DTI imaging data and Zipunnikov et al. (2011a) and Zipunnikov et al. (2011b), who studied RAVENS images with cross-sectional and clustered (but not longitudinal) settings. Moreover, this manuscript represents the first application of LFPCA to morphometric imaging in multiple sclerosis.

A key insight from the simulation studies is the ability of LFPCA to uncover interesting directions of variation in the presence of errors from registration to a template. Previously, registration errors were handled via either extremely aggressive smoothing during post-registration processing or by improved normalization algorithms. While improved algorithms are certainly a desirable goal, all normalization algorithms must be tuned and suffer from trade offs (such as bias and variance). Our results suggest the possibility of employing less aggressive normalization.

The performance of LFPCA depends on the number of subjects, the number of time points, and time span over which data is collected. In designing imaging studies for LFPCA, having both a large number of subjects and a large number of visits may be challenging to obtain. Simulation studies we have conducted during the process of examining LFPCA showed that LFPCA performed well as long as we have either many subjects with smaller number of visits or smaller number of subjects with many visits. It is recommended to make the time span over which data is collected roughly similar across subjects, and long enough to observe longitudinal changes.

In our study of MS, we found that the majority of variation is focused in cross-sectional components. This will likely be true in any study of adults, as variation in head size, brain
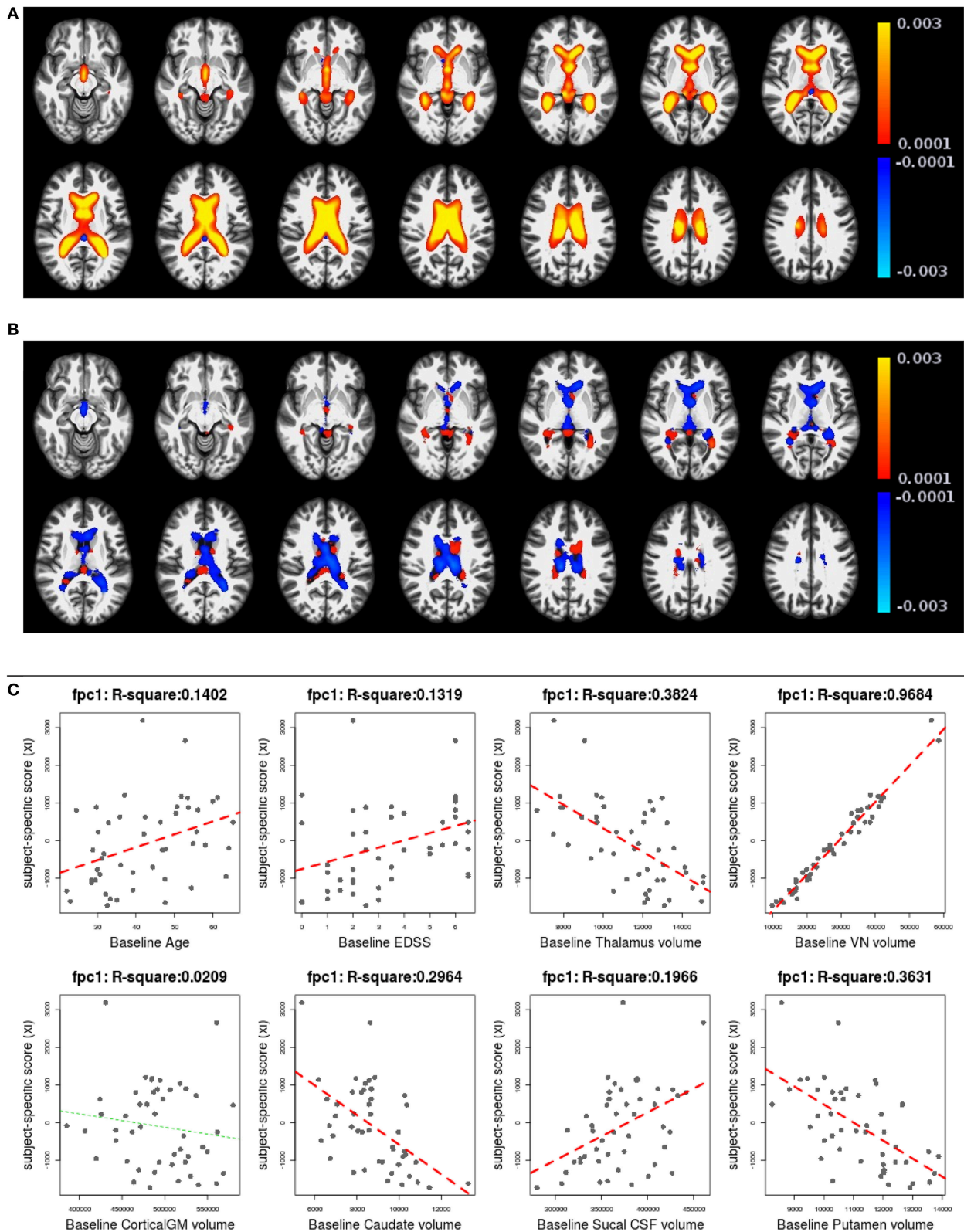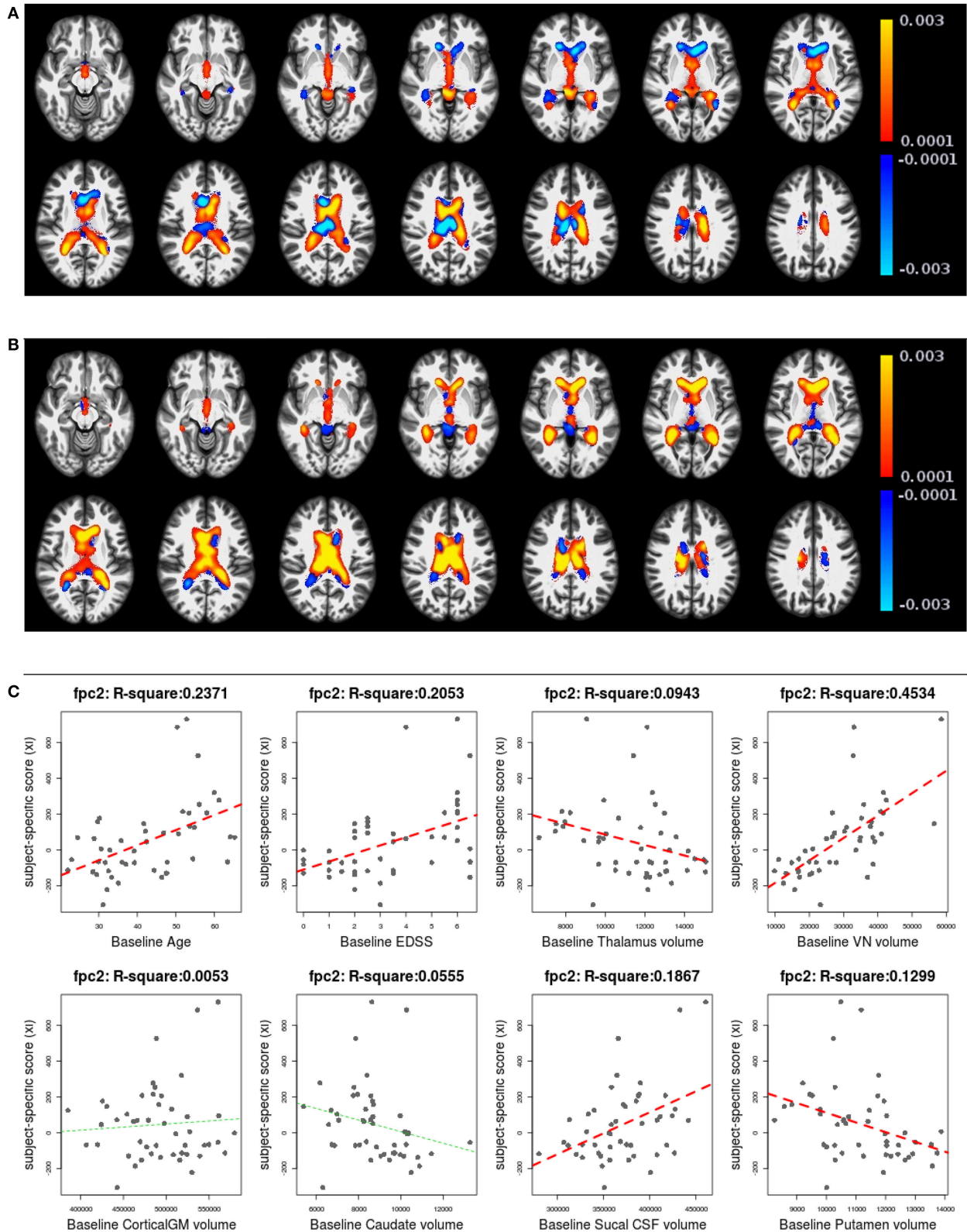
FIGURE 6 | The first subject-specific LFPC and scores. (A) The baseline map, **(B)** the longitudinal map, **(C)** the first subject-specific LFPC scores and covariates of interest. The baseline map represents cross-sectional variation while the longitudinal map has lower loadings. The LFPC scores have higher correlations with baseline age, EDSS, volumes of GM substructures (thalamus, caudate, putamen), ventricle, and sulcal CSF. The correlation with ventricluar volume is very high, which confirms that the first LFPC represents cross-sectional ventricle size.

FIGURE 7 | The second subject-specific LFPC and scores. (A) The baseline map, (B) the longitudinal map, (C) the second subject-specific LFPC scores and covariates of interest. The baseline map has relatively lower loadings while the longitudinal map shows an enlargement pattern in the ventricles. The LFPC scores have higher correlations with baseline age, EDSS, volumes of gray matter substructures (thalamus, putamen), ventricle, and sulcal CSF. The correlations with baseline age and EDSS scores are higher than those of the first LFPC scores.

FIGURE 8 | Variation explained by first 10 subject-specific LFPCs in RAVENS maps. (A) Gray matter, (B) white matter.
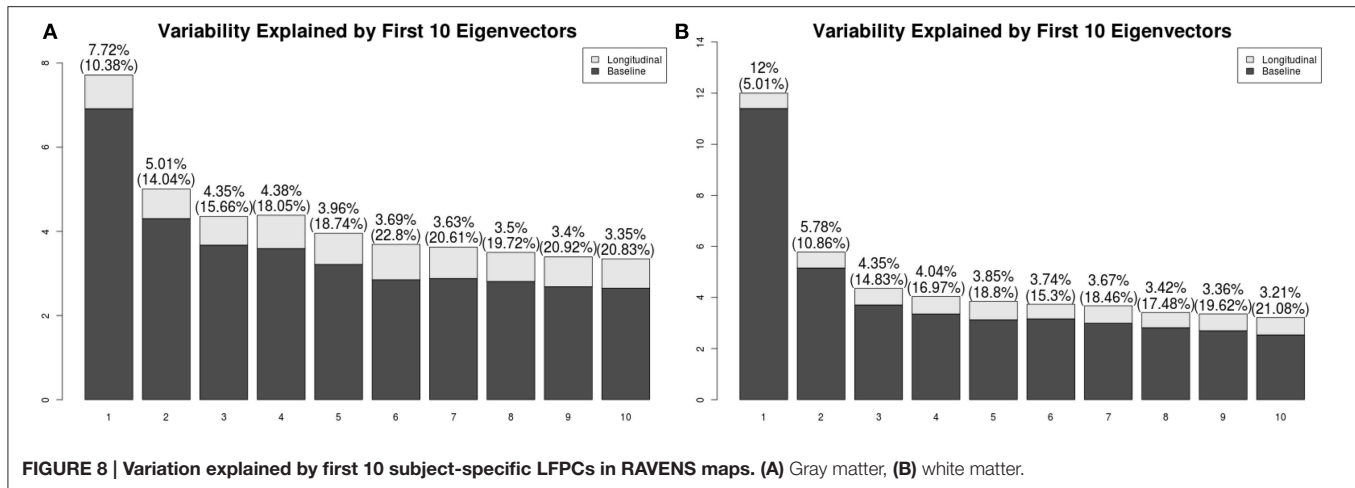
size and intracranial volume will vary far more substantially than longitudinal decline, not unlike if one were to study adult cross-sectional and longitudinal trends in heights. It would be of interest to apply LFPCA to developmental populations or populations with severe progressive brain disorders significantly after disease onset.

The correlation between subject-specific LFPC scores of ventricles and EDSS indicates that EDSS is better associated with longitudinal ventricular enlargement than baseline ventricular size. This implies ventricular enlargement is a sensitive measurement of disease progression. Some cross-sectional MS patient studies have reported that brain atrophy is related to irreversible clinical disability in (MS) and ventricular enlargement may be a sensitive marker of this tissue loss that is seen at all stages of MS (Turner et al., 2001; Benedict et al., 2005; Hildebrandt et al., 2006; Tekok-Kilic et al., 2007). In existing longitudinal studies, longitudinal ventricular enlargement and gray matter atrophy have been detected in both ROI and VBM with paired *t*-test or factor models (Simon et al., 1999; Dalton et al., 2002, 2006; Kalkers et al., 2002; Sepulcre et al., 2006; Lukas et al., 2010; Bendfeldt et al., 2009), which agrees with our finding. Unlike other methods, LFPCA is able to show spatially heterogeneous patterns of longitudinal enlargement, which ROI based methods cannot provide.

In the manuscript, we employed a registration strategy similar to Ashburner and Ridgway (2012). Recent developments in longitudinal registration algorithms (e.g., 4DHammer: Shen et al., 2003, CLASSIC: Shen et al., 2005, GLIRT: Wu et al., 2010, Quarc: Holland et al., 2011) are potentially capable of providing higher accuracy in tracking within subject anatomical changes. Improvements in registration would likely increase the sensitivity of LFPCA to subject-specific signals. However, visit-to-visit variation caused by image acquisition inconsistencies or large anatomical differences between subjects often cannot be corrected by registration. An advantage of the proposed method is that it can simultaneously quantify and characterize both cross-sectional and longitudinal signals of

interest in the presence of potentially large amounts of visit-to-visit variation.

As demonstrated previously for longitudinal diffusion imaging analysis, and here for longitudinal voxel-based morphometry, LFPCA is a compelling alternative to linear mixed model analysis for exploring spatial patterns of anatomical variation within and across subjects. We emphasize that this approach is not limited to a specific brain modality. Besides neuroimaging, we look forward to seeing this method is applied to many other exciting studies including epigenetics. For example, genome-wide DNA methylation data collected at multiple time points could be analyzed to study mechanisms of epigenetic changes related to certain diseases (Martino et al., 2014) or environmental exposure (Martino et al., 2013). Another potential domain of application is for analyzing dynamic imaging data, such as functional MRI or motion imaging. Such data often possess much larger numbers of time points, which would be needed to model the more complex variations in signal.

The LFPCA method described here is designed to model a linear trajectory over time. Given a relatively small number of visits (e.g., three visits on average) it is not feasible to model non-linear trends. However, if the data are collected over greater than 5 time points, the modeling of non-linear trajectories is possible. Currently, we are under a preliminary development of a method to extend LFPCA for non-linear trends modeled using spline functions. Further investigation on the numerical stability and performance will be conducted in the near future.

FUNDING

REFERENCES

Ashburner, J., and Friston, K. (2000). Voxel-based morphometry–the methods. *Neuroimage* 11, 805–821. doi: 10.1006/nimg.2000.0582

Ashburner, J., and Friston, K. (2001). Why voxel-based morphometry whould be used. *Neuroimage* 14, 1238–1243. doi: 10.1006/nimg.2001.0961

Ashburner, J., and Ridgway, G. R. (2012). Symmetric diffeomorphic modeling of longitudinal structural mri. *Front. Neurosci.* 6:197. doi: 10.3389/fnins.2012.00197

Avants, B. B., Tustison, N. J., Song, G., Cook, P. A., Klein, A., and Gee, J. C. (2011). A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage* 54, 2033–2044. doi: 10.1016/j.neuroimage.2010.09.025

Avants, B. B., Yushkevich, P., Pluta, J., Minkoff, D., Korczykowski, M., Detre, J., et al. (2010). The optimal template effect in hippocampus studies of diseased populations. *Neuroimage* 49, 2457–2466. doi: 10.1016/j.neuroimage.2009.09.062

Bartzokis, G., Beckson, M., Lu, P. H., Nuechterlein, K. H., Edwards, N., and Mintz, J. (2001). Age-related changes in frontal and temporal lobe volumes in men: a magnetic resonance imaging study. *Arch. Gen. Psychiat.* 58:461. doi: 10.1001/archpsyc.58.5.461

Bendfeldt, K., Hofstetter, L., Kuster, P., Traud, S., Mueller-Lenke, N., Naegelin, Y., et al. (2011). Longitudinal gray matter changes in multiple sclerosis: differential scanner and overall disease-related effects. *Hum. Brain Mapp.* 33, 1225–1245. doi: 10.1002/hbm.21279

Bendfeldt, K., Kuster, P., Traud, S., Egger, H., Winklhofer, S., Mueller-Lenke, N., et al. (2009). Association of regional gray matter volume loss and progression of white matter lesions in multiple sclerosis–a longitudinal voxel-based morphometry study. *Neuroimage* 45, 60–67. doi: 10.1016/j.neuroimage.2008.10.006

Benedict, R. H., Zivadinov, R., Carone, D. A., Weinstock-Guttman, B., Gaines, J., Maggiore, C., et al. (2005). Regional lobar atrophy predicts memory impairment in multiple sclerosis. *Am. J. Neuroradiol.* 26, 1824–1831.

Benjamini, Y., and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* 29, 1165–1188.

Bermel, R. H., Innus, M. D., Tjoa, C. W., and Bakshi, R. (2003). Selective caudate atrophy in multiple sclerosis: a 3D MRI parcellation study. *Neuroreport* 14, 335–339. doi: 10.1097/00001756-200303030-00008

Chen, S., and Bowman, F. (2011). A novel support vector classifier for longitudinal high-dimensional data and its application to neuroimaging data. *Stat. Anal. Data Mining* 4, 604–611. doi: 10.1002/sam.10141

Dalton, C. M., Brex, P. A., Jenkins, R., Fox, N. C., Miszkiel, K. A., Crum, W. R., et al. (2002). Progressive ventricular enlargement in patients with clinically isolated syndromes is associated with the early development of multiple sclerosis. *J. Neurol. Neurosurg. Psychiatry* 73, 141–147. doi: 10.1136/jnnp.73.2.141

Dalton, C. M., Miszkiel, K. A., O'Connor, P. W., Plant, G. T., Rice, G. P., and Miller, D. H. (2006). Ventricular enlargement in ms. *Neurology* 66, 693–698. doi: 10.1212/01.wnl.0000201183.87175.9f

Derado, G., Bowman, F. D., Zhang, L., and Alzheimer's Disease Neuroimaging Initiative. (2013). Predicting brain activity using a Bayesian spatial model. *Stat. Methods Med. Res.* 22, 382–397. doi: 10.1177/0962280212448972

Di, C., Crainiceanu, C., Caffo, B., and Punjabi, N. (2009). Multilevel functional principal component analysis. *Ann. Appl. Stat.* 3, 458–488. doi: 10.1214/08-AOAS206

Driemeyer, J., Boyke, J., Gaser, C., Büchel, C., and May, A. (2008). Changes in gray matter induced by learning–revisited. *PLoS ONE* 3:E2669. doi: 10.1371/journal.pone.0002669

Greven, S., Crainiceanu, C., Caffo, B., and Reich, D. (2011). "Longitudinal functional principal component analysis," in *Recent Advances in Functional Data Analysis and Related Topics*, ed F. Ferraty (Berlin; Heidelberg: Physica-Verlag HD), 149–154. doi: 10.1007/978-3-7908-2736-1_23

Hildebrandt, H., Hahn, H. K., Kraus, J. A., Schulte-Herbrüggen, A., Schwarze, B., and Schwendemann, G. (2006). Memory performance in multiple sclerosis patients correlates with central brain atrophy. *Mult. Scl.* 12, 428–436. doi: 10.1191/1352458506ms1286oa

Holland, D., Dale, A. M., and Alzheimer's Disease Neuroimaging Initiative. (2011). Nonlinear registration of longitudinal images and measurement

of change in regions of interest. *Med. Image Anal.* 15, 489–497. doi: 10.1016/j.media.2011.02.005

Jenkinson, M., Bannister, P., Brady, M., and Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* 17, 825–841. doi: 10.1006/nimg.2002.1132

Kalkers, N. F., Ameziane, N., Bot, J. C., Minneboo, A., Polman, C. H., and Barkhof, F. (2002). Longitudinal brain volume measurement in multiple sclerosis: rate of brain atrophy is independent of the disease subtype. *Arch. Neurol.* 59:1572. doi: 10.1001/archneur.59.10.1572

Karhunen, K. (1947). Über lineare methoden in der wahrscheinlichkeitsrechnung. *Ann. Acad. Sci. Fennicae. A Math. Phys.* 1947:79.

Loève, M. (1948). *Fonctions Aléatoires du Second Ordre*. Supplement to P. Lévy, Processus Stochastiques et Mouvement Brownien. Paris: Gauthier-Villars.

Lukas, C., Minneboo, A., de Groot, V., Moraal, B., Knol, D. L., Polman, C. H., et al. (2010). Early central atrophy rate predicts 5 year clinical outcome in multiple sclerosis. *J. Neurol. Neurosurg. Psychiatry* 81, 1351–1356. doi: 10.1136/jnnp.2009.199968

Martino, D., Joo, J. E., Sexton-Oates, A., Dang, T., Allen, K., Saffery, R., et al. (2014). Epigenome-wide association study reveals longitudinally stable dna methylation differences in cd4+ t cells from children with ige-mediated food allergy. *Epigenetics* 9, 998–1006. doi: 10.4161/epi.28945

Martino, D., Loke, Y. J., Gordon, L., Ollikainen, M., Cruickshank, M. N., Saffery, R., et al. (2013). Longitudinal, genome-scale analysis of dna methylation in twins from birth to 18 months of age reveals rapid epigenetic change in early life and pair-specific effects of discordance. *Genome Biol.* 14:R42. doi: 10.1186/gb-2013-14-5-r42

Sepulcre, J., Sastre-Garriga, J., Cercignani, M., Ingle, G. T., Miller, D. H., and Thompson, A. J. (2006). Regional gray matter atrophy in early primary progressive multiple sclerosis: a voxel-based morphometry study. *Arch. Neurol.* 63:1175. doi: 10.1001/archneur.63.8.1175

Shen, D., and Davatzikos, C. (2002). HAMMER: hierarchical attribute matching mechanism for elastic registration. *Med. Imaging IEEE Trans.* 21, 1421–1439. doi: 10.1109/TMI.2002.803111

Shen, D., Resnick, S. M., and Davatzikos, C. (2003). "4D HAMMER image registration method for longitudinal study of brain changes," in *Proceedings of the Human Brain Mapping* (New York, NY).

Shen, D., Sundar, H., Xue, Z., Fan, Y., and Litt, H. (2005). "Consistent estimation of cardiac motions by 4D image registration," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2005*, eds J. S. Duncan and G. Gerig (Berlin; Heidelberg: Springer), 902–910. doi: 10.1007/11566489_111

Shiee, N., Bazin, P. L., Ozturk, A., Reich, D. S., Calabresi, P. A., and Pham, D. L. (2010). A topology-preserving approach to the segmentation of brain images with multiple sclerosis lesions. *Neuroimage* 49, 1524–1535. doi: 10.1016/j.neuroimage.2009.09.005

Simon, J. H., Jacobs, L. D., Campion, M. K., Rudick, R. A., Cookfair, D. L., Herndon, R. M., et al. (1999). A longitudinal study of brain atrophy in relapsing multiple sclerosis. *Neurology* 53:139. doi: 10.1212/WNL.53.1.139

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E., Johansen-Berg, H., et al. (2004). Advances in functional and structural mr image analysis and implementation as fsl. *Neuroimage* 23, S208–S219. doi: 10.1016/j.neuroimage.2004.07.051

Tekok-Kilic, A., Benedict, R. H., Weinstock-Guttman, B., Dwyer, M. G., Carone, D., Srinivasaraghavan, B., et al. (2007). Independent contributions of cortical gray matter atrophy and ventricle enlargement for predicting neuropsychological impairment in multiple sclerosis. *Neuroimage* 36, 1294–1300. doi: 10.1016/j.neuroimage.2007.04.017

Turner, B., Ramli, N., Blumhardt, L. D., and Jaspan, T. (2001). Ventricular enlargement in multiple sclerosis: a comparison of three-dimensional and linear mri estimates. *Neuroradiology* 43, 608–614. doi: 10.1007/s002340000457

Worsley, K., Taylor, J., Carbonell, F., Chung, M., Duerden, E., Bernhardt, B., et al. (2009). Surfstat: a matlab toolbox for the statistical analysis of univariate and multivariate surface and volumetric data using linear mixed effects models

and random field theory. *Neuroimage* 47:S102. doi: 10.1016/S1053-8119(09)70882-1

Wu, G., Wang, Q., Jia, H., and Shen, D. (2010). "Registration of longitudinal image sequences with implicit template and spatial-temporal heuristics," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010*, Vol. 6362, eds T. Jiang, N. Navab, J. P. W. Pluim, and M. A. Viergever (Berlin; Heidelberg: Springer), 618–625. doi: 10.1007/978-3-642-15745-5_76

Zipunnikov, V., Caffo, B., Yousem, D., Davatzikos, C., Schwartz, B., and Crainiceanu, C. (2011a). Functional principal components model for high-dimensional brain imaging. *Neuroimage* 58, 772–784. doi: 10.1016/j.neuroimage.2011.05.085

Zipunnikov, V., Caffo, B., Yousem, D. M., Davatzikos, C., Schwartz, B. S., and Crainiceanu, C. (2011b). Multilevel functional principal component analysis for high-dimensional data. *J. Comput. Graph. Stat.* 20, 852–873. doi: 10.1198/jcgs.2011.10122

Zipunnikov, V., Greven, S., Shou, H., Caffo, B., Reich, D. S., Crainiceanu, C., et al. (2014). Longitudinal high-dimensional principal components analysis with application to diffusion tensor imaging of multiple sclerosis. *Ann. Appl. Stat.* 8, 2175–2202. doi: 10.1214/14-AOAS748

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.