Longitudinal High-Dimensional Data Analysis

Vadim Zipunnikov*, Sonja Greven, Brian Caffo,
Daniel S. Reich, Ciprian Crainiceanu

November 14, 2011

Abstract

We develop a flexible framework for modeling high-dimensional functional and imaging data observed longitudinally. The approach decomposes the observed variability of high-dimensional observations measured at multiple visits into three additive components: a subject-specific functional random intercept that quantifies the cross-sectional variability, a subject-specific functional slope that quantifies the dynamic irreversible deformation over multiple visits, and a subject-visit specific functional deviation that quantifies exchangeable or reversible visit-to-visit changes. The proposed method is very fast, scalable to studies including ultra-high dimensional data, and can easily be adapted to and executed on modest computing infrastructures. The method is applied to the longitudinal analysis of diffusion tensor imaging (DTI) data of the corpus callosum of multiple sclerosis (MS) subjects. The study includes 176 subjects observed at 466 visits. For each subject and visit the study contains a registered DTI scan of the corpus callosum at roughly 30,000 voxels.

Keywords: functional data analysis, principal components, linear mixed model, diffusion tensor imaging, brain imaging data, multiple sclerosis.

1 Introduction

An increasing number of longitudinal studies routinely acquire high-dimensional data, such as brain images or gene expression, at multiple visits. This led to increased interest in generalizing standard models designed for longitudinal data analysis to the case when the observed data are massively multivariate or functional. In this paper we propose

*Vadim Zipunnikov is a Postdoctoral Fellow, Brian Caffo and Ciprian Crainiceanu are Associate Professors, Department of Biostatistics, Johns Hopkins University, Baltimore, MD, 21205 (emails: vzipunni@jhsph.edu, bcaffo@jhsph.edu, ccrainic@jhsph.edu), Sonja Greven is Emmy Noether Junior Research Group Leader, Department of Statistics, Ludwig-Maximilians-Universität and München, 80539 Munich, Germany (e-mail: sonja.greven@stat.uni-muenchen.de), Daniel Reich is Chief of the Translational Neurology Unit, Neuroimmunology Branch, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD 20892, USA (email:daniel.reich@nih.gov)

to generalize the random intercept random slope mixed effect model to the case when instead of a scalar, such as systolic blood pressure, one measures a highly multivariate object, such as a brain image. The proposed methods can be applied to any longitudinal study that includes high-dimensional functional or imaging data.

This paper is motivated by a study of multiple sclerosis (MS) patients (Reich et al., 2010). Multiple sclerosis is a degenerative disease of the central nervous system. A hallmark of MS is damage to and degeneration of the myelin sheaths that surround and insulate nerve fibers in the brain. Such damage results in sclerotic plaques that distort the flow of electrical impulses along the nerves to different parts of the body (Raine et al., 2008). Secondary to the loss of myelin, but potentially as well a primary manifestation of the disease, MS directly affects the neurons themselves, resulting in accelerated brain atrophy.

Our data are derived from a natural history study of 176 MS cases drawn from a wide spectrum of disease severity. Subjects were scanned over a 5-year period up to 10 times per subject, for a total of 466 scans. The scans have been aligned using a 12 degrees of freedom transformation, meaning that we accounted for rotation, translation, scaling, and shearing, but not for nonlinear deformation. In this study we focus on fractional anisotropy (FA), a useful DTI summary which is a measure of tissue integrity and is sensitive to both axon fiber density and myelination in white matter. FA is measured on the scale between zero (isotropic diffusion characteristic of fluid-filled cavities) and one (anisotropic diffusion, characteristic of highly ordered white matter fiber bundles) and characterizes the degree of d of a diffusion process (Mori, 2007).

The goal of the study was to model longitudinal variability of FA in the corpus callosum. The primary region of interest (ROI), therefore is a central block of the brain containing the corpus callosum, the major bundle of neural fibers connecting the left and right cerebral hemispheres. We weight FA at each voxel in the block with a probability for the voxel to be in the corpus callosum, where the probability is derived from an atlas formed using healthy-volunteer scans, and study longitudinal changes of weighted FAs in the blocks (Reich et al., 2010). Figure 1 displays the ROI displayed as a blue block in a template brain. Each block is of size $38 \times 72 \times 11$, indicating that there are 38 sagittal, 72 coronal, and 11 axial slices, respectively. Figure 2 displays the 11 axial (horisontal) slices for one of the subjects from bottom to top. In this paper, we study the FA at every voxel of the blue blocks, which could be unfolded into an approximately $30,000$ dimensional vector that contains the corresponding FA value at each entry. The variability of these images over multiple visits and subjects will be described by the combination of: 1) a subject-specific functional random intercept that quantifies the cross-sectional variability; 2) a subject-specific functional slope that quantifies the dynamic irreversible deformation over multiple visits; and 3) a subject-visit specific functional deviation that quantifies exchangeable or reversible visit-to-visit changes.

High dimensional data sets have motivated the statistical and imaging communities to develop new methodological approaches to data analysis. Successful modeling approaches involving wavelets and splines have been reported in the literature (Mohamed and Davatzikos, 2004; Morris and Carroll, 2006; Guo, 2002; Morris et al., 2011; Zhu et al., 2011; Morris et al., 2011; Rodriguez et al., 2009; Bigelow and Dunson, 2009; Reiss

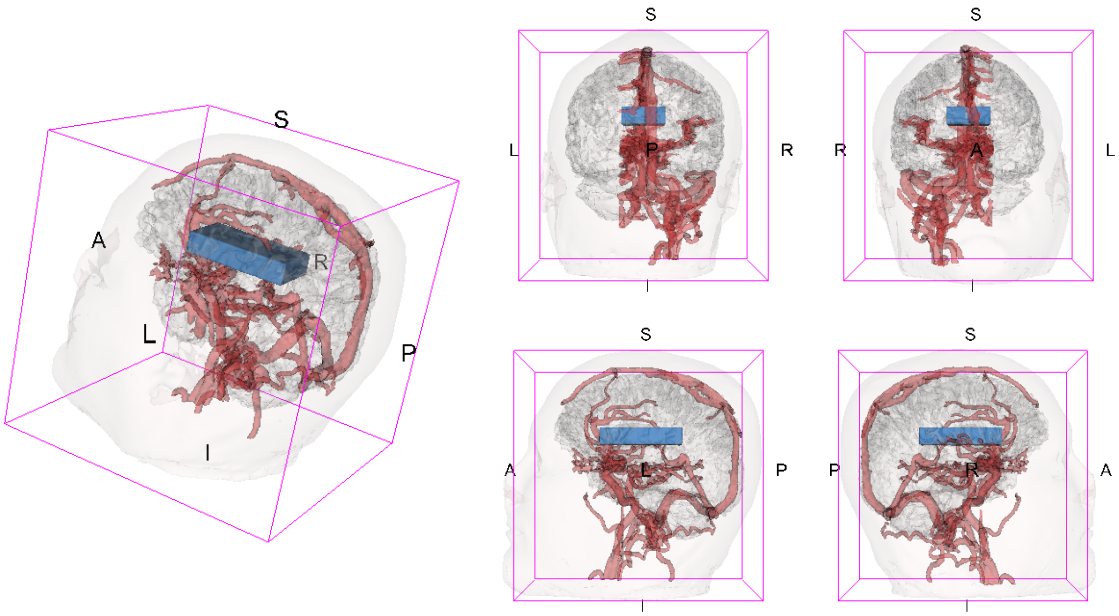

Figure 1: *The 3D-rendering of the template brain and the region of interest, a blue block containing corpus callosum. Views: R=Right, L=Left, S=Superior, I=Interior, A=Anterior, P=Posterior. For the purposes of orientation, major venous structures are displayed in red in the right half of the template brain. The 3D-renderings are obtained using 3D-Slicer (2011) and 3D reconstructions of the anatomy from Pujol (2010).*

et al., 2005; Reiss and Ogden, 2008, 2010). A different direction of research has focused on principal component decompositions (Di et al., 2008; Crainiceanu et al., 2009; Aston et al., 2010; Staicu et al., 2010; Greven et al., 2010; Di et al., 2010; Zipunnikov et al., 2011b; Crainiceanu et al., 2011), which led to several applications to imaging data (Shinohara et al., 2011; Goldsmith et al., 2011; Zipunnikov et al., 2011a). However, the high dimensionality of new data sets, the inherent complexity of sampling designs and data collection, and the diversity of new technological measurements raise multiple difficult challenges that are currently unaddressed.

Here we address directly the problem of analyzing populations of high dimensional images at multiple visits using high dimensional longitudinal functional principal components analysis (HD-LFPCA). The method decomposes the longitudinal functional/imaging data into a subject specific, longitudinal subject specific, and subject-visit specific components. The dimension reduction for all components is done using principal components of corresponding covariance operators. The estimation and inferential methods are fast and can be performed on standard personal computers to analyze hundreds or thousands of high-dimensional curves or images at multiple visits. This was achieved by a combination of statistical and computational methods: 1) relying only on matrix block calculations and sequential access to memory to avoid loading very large data sets into the computer memory (see Demmel, 1997 and Golub and Loan, 1996 for a comprehensive review of partitioned matrix techniques); 2) using SVD for matrices that have at least one dimension smaller than $10,000$ (Zipunnikov et al., 2011a) 3) obtaining best linear unbiased predictors (BLUPS) of principal scores as a by-product of SVD of the data matrix; 4) linking the high-dimensional space to a low-dimensional intrinsic space; this

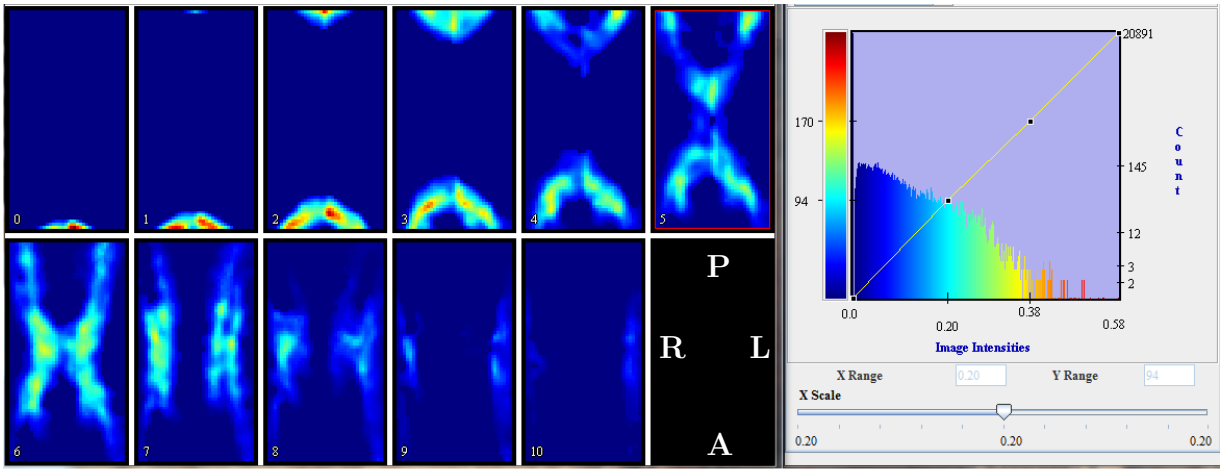

Figure 2: *The corpus callosum of a randomly chosen subject. Eleven axial slices are shown on the left. A histogram of the weighted FA values is on the right. Orientation: Interior(slice* 0*) to Superior(slice* 10*), Posterior (top) to Anterior(bottom), Right to Left. The pictures are obtained using MIPAV (2011).*

allows Karhunen-Loeve (KL) decompositions of covariance operators that cannot even be stored in the computer memory. The proposed methods are computationally linear in the dimension of the function or images. Thus, even though we analyze images with tens of thousands of voxels, the methods can easily be applied to images that are 2 or 3 orders of magnitude larger.

The rest of the manuscript is laid out as follows. Section 2 reviews LFPCA and discusses its limitation in high-dimensional settings. In Section 3 we introduce HD-LFPCA, which provides a new statistical and computational framework for LFPCA. This will circumvent the problems associated with LFPCA in high dimensional settings. Simulation studies are provided in Section 4. Our methods are applied to the MS data in Section 5. Section 6 concludes the paper with a discussion.

2 Longitudinal FPCA

In this section we review the LFPCA framework introduced by Greven et al. (2010). We develop an estimation procedure based on the original one in Greven et al. (2010) but heavily modified for application to high dimensional data. We emphasize the major reasons why the original methods can not be applied to high-dimensional data such as the voxel-wise FA in the corpus callosum for MS subjects described in Section 1.

2.1 Model

A brain imaging longitudinal study usually contains a sample of images $\mathbf{Y}_{ij}$, where $\mathbf{Y}_{ij}$ is a recorded brain image of the ith subject, $i = 1, \ldots, I$, scanned at times $T_{ij}, j = 1, \ldots, J_i$. The total number of subjects is denoted by I. The times T_{ij} are subject specific. Different subjects could have different number of visits (scans), J_i. The images are stored in 3-dimensional array structures of dimension $p = p_1 \times p_2 \times p_3$. For example, in the MS data

$p = 38 \times 72 \times 11 = 30,096$. Note that our approach is not limited to the case when data are in a 3 dimensional array. Instead, it can be applied to any data structure where the voxels (or pixels, or locations, etc.) are the same across subjects and visits and data can be unfolded into a vector. Following Greven et al. (2010) we consider the LFPCA model

$$Y_{ij}(v) = \eta(v, T_{ij}) + X_{i,0}(v) + X_{i,1}(v)T_{ij} + W_{ij}(v), \tag{1}$$

where v denotes a voxel, $\eta(v, T_{ij})$ is a fixed main effect, $X_{i,0}(v)$ is the random functional/imaging intercept for subject i, $X_{i,1}(v)$ is the random functional/imaging slope for subject i, T_{ij} is the time of visit j for subject i, $W_{ij}(v)$ is the random subject/visit-specific functional/imaging deviation.

In the remainder of the paper, we unfold the data $\mathbf{Y}_{ij}$ and represent it as a $p \times 1$ dimensional vector containing the voxels in a particular order, where the order is preserved across all subjects and visits. We assume that $\eta(v, T_{ij})$ is a fixed surface/image and the latent (unobserved) $2p$-dimensional process $X_i(v) = (X'_{i,0}(v), X'_{i,1}(v))'$ and the p-dimensional process $W_{ij}(v)$ are zero-mean second-order stationary. We also assume that $X_i(v)$ and $W_{ij}(v)$ are uncorrelated. We denote by $\mathbf{K}^X(v_1, v_2)$ and $\mathbf{K}^W(v_1, v_2)$ their covariance operators, respectively. Assuming that $\mathbf{K}^X(v_1, v_2)$ and $\mathbf{K}^W(v_1, v_2)$ are continuous, we can use the standard Karhunen-Loeve expansions of the random processes (Karhunen, 1947; Loeve, 1978) and represent $X_i(v) = \sum_{k=1}^{\infty} \xi_{ik}\phi_k^X(v)$ with $\phi_k^X(v) = (\phi_k^{X,0}(v), \phi_k^{X,1}(v))$ and $W_{ij}(v) = \sum_{l=1}^{\infty} \zeta_{ijl}\phi_l^W(v)$, where ϕ_k^X and ϕ_l^W are the eigenfunctions of the $\mathbf{K}^X$ and $\mathbf{K}^W$ operators, respectively. Note that $\mathbf{K}^X$ and $\mathbf{K}^W$ will be estimated by their sample counterparts on finite $2p \times 2p$ and $p \times p$ grids, respectively. Hence, we can always make a working assumption of continuity $\mathbf{K}^X$ and $\mathbf{K}^W$. The LFPCA model becomes the mixed effects model

$$\begin{cases} Y_{ij}(v) = \eta(v, T_{ij}) + \sum_{k=1}^{\infty} \xi_{ik}\mathbf{Z}'_{ij}\phi_k^X(v) + \sum_{l=1}^{\infty} \zeta_{ijl}\phi_l^W(v) \\ (\xi_{ik_1}, \xi_{ik_2}) \sim (0, 0; \lambda_{k_1}^X, \lambda_{k_2}^X, 0); (\zeta_{ijl_1}, \zeta_{ijl_2}) \sim (0, 0; \lambda_{l_1}^W, \lambda_{l_2}^W, 0), \end{cases} \tag{2}$$

where $\mathbf{Z}_{ij} = (1, T_{ij})'$ and " $\sim (0, 0; \lambda_{k_1}^X, \lambda_{k_2}^X, 0)$" indicates that a pair of variables is uncorrelated with mean zero and variances $\lambda_{k_1}^X$ and $\lambda_{k_2}^X$, respectively. Variances λ_k^X's are nonincreasing, that is $\lambda_{k_1}^X \geq \lambda_{k_2}^X$ if $k_1 \leq k_2$. No distributional assumptions on the scores are required. In addition, the assumption that $X_i(v)$ and $W_{ij}(v)$ are uncorrelated is ensured by the assumption that $\{\xi_{ik}\}_{k=1}^{\infty}$ and $\{\zeta_{ijl}\}_{l=1}^{\infty}$ are uncorrelated. Note that model (2) may be extended to include a more general vector of covariates $\mathbf{Z}_{ij}$. We discuss a general functional mixed model in Section 3.3.

In practice, model 2 is projected onto the first N_X and N_W components, respectively. Assuming that N_X and N_W are known, the model becomes

$$\begin{cases} Y_{ij}(v) = \eta(v, T_{ij}) + \sum_{k=1}^{N_X} \xi_{ik}\mathbf{Z}'_{ij}\phi_k^X(v) + \sum_{l=1}^{N_W} \zeta_{ijl}\phi_l^W(v) \\ (\xi_{ik_1}, \xi_{ik_2}) \sim (0, 0; \lambda_{k_1}^X, \lambda_{k_2}^X, 0); (\zeta_{ijl_1}, \zeta_{ijl_2}) \sim (0, 0; \lambda_{l_1}^W, \lambda_{l_2}^W, 0). \end{cases} \tag{3}$$

The choice of the number of principal components N_X and N_W is discussed in (Di et al., 2008; Greven et al., 2010). Typically, N_X and N_W are small and (3) provides significant dimension reduction of the family of images and their longitudinal dynamics. The main

reason why the LFPCA model (3) cannot be fit when data are high dimensional is that the empirical covariance matrices $\mathbf{K}^X$ and $\mathbf{K}^W$ can not be calculated, stored or diagonalized. Indeed, in our case these operators would be $30,000$ by $30,000$ dimensional, which would have around 1 billion entries. In other applications these operators would be even bigger.

2.2 Estimation

There are many efficient and flexible ways to parametrize and estimate $\eta(v, T_{ij})$. Some of them are discussed in detail in (Greven et al., 2010). Therefore, we assume that $\eta(v, T_{ij})$ is estimated as $\tilde{\eta}(v, T_{ij})$ and define the unexplained part of the image as $\tilde{Y}_{ij}(v) = Y_{ij}(v) - \tilde{\eta}(v, T_{ij})$. The computationally intensive part of fitting (3) is estimating the following massively multivariate model

$$\widetilde{\mathbf{Y}}_{ij} = \sum_{k=1}^{N_X} \xi_{ik}\boldsymbol{\phi}_k^{X,0} + T_{ij}\sum_{k=1}^{N_X} \xi_{ik}\boldsymbol{\phi}_k^{X,1} + \sum_{l=1}^{N_W} \zeta_{ijl}\boldsymbol{\phi}_l^{W}, \tag{4}$$

where $\widetilde{\mathbf{Y}}_{ij} = \{\tilde{Y}_{ij}(v_1), \ldots, \tilde{Y}_{ij}(v_p)\}$ is a $p \times 1$ dimensional vector, and $\boldsymbol{\phi}_k^{X,0}$, $\boldsymbol{\phi}_k^{X,1}$, and $\boldsymbol{\phi}_l^{W}$ are correspondingly vectorized eigenvectors. The model can be rewritten in matrix form as

$$\widetilde{\mathbf{Y}}_{ij} = \boldsymbol{\Phi}^{X,0}\boldsymbol{\xi}_i + T_{ij}\boldsymbol{\Phi}^{X,1}\boldsymbol{\xi}_i + \boldsymbol{\Phi}^{W}\boldsymbol{\zeta}_{ij}, \tag{5}$$

where $\boldsymbol{\Phi}^{X,0} = [\boldsymbol{\phi}_1^{X,0}, \ldots, \boldsymbol{\phi}_{N_X}^{X,0}]$ and $\boldsymbol{\Phi}^{X,1} = [\boldsymbol{\phi}_1^{X,1}, \ldots, \boldsymbol{\phi}_{N_X}^{X,1}]$ are $p \times N_X$ dimensional matrices, $\boldsymbol{\Phi}^{W} = [\boldsymbol{\phi}_1^{W}, \ldots, \boldsymbol{\phi}_{N_W}^{W}]$ is a $p \times N_W$ dimensional matrix, principal scores $\boldsymbol{\xi}_i = (\xi_{i1}, \ldots, \xi_{iN_X})'$ and $\boldsymbol{\zeta}_{ij} = (\zeta_{ij1}, \ldots, \zeta_{ijN_U})'$ are uncorrelated with diagonal covariance matrices $E(\boldsymbol{\xi}_i\boldsymbol{\xi}_i') = \boldsymbol{\Lambda}^X = \text{diag}(\lambda_1^X, \ldots, \lambda_{N_X}^X)$ and $E(\boldsymbol{\zeta}_{ij}\boldsymbol{\zeta}_{ij}') = \boldsymbol{\Lambda}^W = \text{diag}(\lambda_1^W, \ldots, \lambda_{N_W}^W)$, respectively. To obtain the eigenvectors and eigenvalues in model (5), the spectral decompositions of $\mathbf{K}^X$ and $\mathbf{K}^W$ need to be constructed. The first N_X and N_W eigenvectors and eigenvalues are retained after this, that is $\mathbf{K}^X \approx \boldsymbol{\Phi}^X\boldsymbol{\Lambda}^X\boldsymbol{\Phi}^{X\prime}$ and $\mathbf{K}^W \approx \boldsymbol{\Phi}^W\boldsymbol{\Lambda}^W\boldsymbol{\Phi}^{W\prime}$, where $\boldsymbol{\Phi}^X = [\boldsymbol{\Phi}^{X,0\prime}, \boldsymbol{\Phi}^{X,1\prime}]'$ denotes a $2p \times N_X$ matrix with orthonormal columns and $\boldsymbol{\Phi}^W$ is a $p \times N_W$ matrix with orthonormal columns.

Next, we describe how LFPCA estimates the covariance operators $\mathbf{K}^X$ and $\mathbf{K}^W$. First, we split the $2p \times 2p$ matrix $\mathbf{K}^X$ into four $p \times p$ blocks as $\mathbf{K}^X = [\mathbf{K}_X^{00}\vdots\mathbf{K}_X^{01}; \mathbf{K}_X^{10}\vdots\mathbf{K}_X^{11}]$, where $\mathbf{K}_X^{ks} = E\{\boldsymbol{\Phi}^{X,k}\boldsymbol{\xi}_i(\boldsymbol{\Phi}^{X,s}\boldsymbol{\xi}_i)'\}$ for $k, s \in \{0, 1\}$. Intuitively, $\mathbf{K}_X^{00}$ is the covariance operator of the random intercept process $X_{0i}(v)$, $\mathbf{K}_X^{11}$ is the covariance operator of the random slope process $X_{1i}(v)$, and $\mathbf{K}_X^{01}$ is the cross-covariance operator of the random slope process $X_{0i}(v)$ with the random slope process $X_{1i}(v)$. From (5), it follows that for subject i observations $\widetilde{\mathbf{Y}}_{ij_1}$ and $\widetilde{\mathbf{Y}}_{ij_2}$ "on average" interact according to the following identity

$$E(\widetilde{\mathbf{Y}}_{ij_1}\widetilde{\mathbf{Y}}_{ij_2}') = \mathbf{K}_X^{00} + T_{ij_2}\mathbf{K}_X^{01} + T_{ij_1}\mathbf{K}_X^{10} + T_{ij_1}T_{ij_2}\mathbf{K}_X^{11} + \delta_{j_1j_2}\mathbf{K}^W, \quad j_1, j_2 = 1, \ldots, J_i \tag{6}$$

where $\delta_{j_1j_2}$ is 1 if $j_1 = j_2$ and 0 otherwise. Identifiability of model (6) requires that some subjects have more than two visits, that is $J_i \geq 3$. Covariances $\mathbf{K}^X$ and $\mathbf{K}^W$ can be estimated by regressing $\widetilde{\mathbf{Y}}_{ij_1}\widetilde{\mathbf{Y}}_{ij_2}'$ on $1, T_{ij_2}, T_{ij_1}, T_{ij_1}T_{ij_2}$, and $\delta_{j_1j_2}$. We can rewrite (6) in matrix form as $E(\widetilde{\mathbf{Y}}_{ij_1j_2}^v) = \mathbf{K}^v\mathbf{f}_{ij_1j_2}$, where $\widetilde{\mathbf{Y}}_{ij_1j_2}^v = \widetilde{\mathbf{Y}}_{ij_2} \otimes \widetilde{\mathbf{Y}}_{ij_1}$ is a $p^2 \times 1$ dimensional vector, the parameter of interest is the $p^2 \times 5$ matrix $\mathbf{K}^v = [\text{vec}(\mathbf{K}_X^{00}), \text{vec}(\mathbf{K}_X^{01}), \text{vec}(\mathbf{K}_X^{10}), \text{vec}(\mathbf{K}_X^{11}),$

$\text{vec}(\mathbf{K}^W)]$, and the covariates are entries in the 5×1 vector $\mathbf{f}_{ij_1j_2} = (1, T_{ij_2}, T_{ij_1}, T_{ij_1}T_{ij_2}, \delta_{j_1j_2})'$. With these notations, (6) can be rewritten as $E\mathbf{Y}^v = \mathbf{K}^v\mathbf{F}$, where $\widetilde{\mathbf{Y}}^v$ is $p^2 \times m$ dimensional with $m = \sum_{i=1}^{I} J_i^2$ and $\mathbf{F}$ is a $5 \times m$ dimensional matrix with columns equal to $\mathbf{f}_{ij_1j_2}, i = 1, \ldots, I$ and $j_1, j_2 = 1, \ldots, J_i$. The ordinary least squares (OLS) estimator of $\mathbf{K}^v$ is thus $\hat{\mathbf{K}}^v = \widetilde{\mathbf{Y}}^v\mathbf{F}'(\mathbf{F}\mathbf{F}')^{-1}$ which provides unbiased estimators of the covariances $\mathbf{K}^X$ and $\mathbf{K}^W$. If we denote $\mathbf{H} = \mathbf{F}'(\mathbf{F}\mathbf{F}')^{-1}$ then each column of $\mathbf{H}$ is a vector of weights such that

$$\hat{\mathbf{K}}_X^{00} = \sum_{i,j_1,j_2} \widetilde{\mathbf{Y}}_{ij_1}\widetilde{\mathbf{Y}}'_{ij_2}h^1_{ij_1j_2}, \quad \hat{\mathbf{K}}_X^{01} = \sum_{i,j_1,j_2} \widetilde{\mathbf{Y}}_{ij_1}\widetilde{\mathbf{Y}}'_{ij_2}h^2_{ij_1j_2}, \quad \hat{\mathbf{K}}_X^{10} = \sum_{i,j_1,j_2} \widetilde{\mathbf{Y}}_{ij_1}\widetilde{\mathbf{Y}}'_{ij_2}h^3_{ij_1j_2}, \quad (7)$$
$$\hat{\mathbf{K}}_X^{11} = \sum_{i,j_1,j_2} \widetilde{\mathbf{Y}}_{ij_1}\widetilde{\mathbf{Y}}'_{ij_2}h^4_{ij_1j_2}, \quad \hat{\mathbf{K}}^W = \sum_{i,j_1,j_2} \widetilde{\mathbf{Y}}_{ij_1}\widetilde{\mathbf{Y}}'_{ij_2}h^5_{ij_1j_2}.$$

Thus, the OLS equations (7) define the symmetric matrices $\hat{\mathbf{K}}^X$ and $\hat{\mathbf{K}}^W$.

Estimating the covariance matrices is a crucial first step. However, constructing and storing these matrices requires $O(p^2)$ calculations and $O(p^2)$ memory units, respectively. Even if it were possible to calculate and store these covariances, obtaining the spectral decompositions would typically be infeasible. Indeed, $\mathbf{K}^X$ is a $2p \times 2p$ and $\mathbf{K}^W$ is a $p \times p$ dimensional matrix, which would require $O(p^3)$ operations, making diagonalization infeasible for $p > 10^4$. Therefore, LFPCA, which performs extremely well when the functional dimensionality is moderate, fails in very high and ultra high dimensional settings.

In the next section we develop a methodology capable of handling longitudinal models of very high dimensionality. The main reason why our methods work efficiently is because the intrinsic dimensionality of the model is controlled by the sample size of the study, which is much smaller compared to the number of voxels. The core part of the methodology is to carefully exploit this underlying low dimensional space.

3 HD-LFPCA

In this section we provide our statistical model and inferential methods. The main emphasis is given to providing a new methodological approach with the ultimate goal of solving the intractable computational problems discussed in the previous section. While our approach was motivated by high dimensional DTI data of the MS study, HD-LFPCA has a considerable potential for the analysis of many other longitudinal study dealing with high-dimensional data.

3.1 Eigenanalysis

In Section 2 we established that the main computational bottleneck for standard LFPCA of Greven et al. (2010) is constructing, storing, and decomposing the relevant covariance operators. In this section we propose an algorithm that allows efficient calculation of the eigenvectors and eigenvalues of these covariance operators. In addition, we demonstrate how all necessary calculations can be done using sequential access to data. One of the main assumptions of this section is that the sample size, $n = \sum_{j=1}^{I} J_i$, is moderate so

calculations of order $O(n^3)$ are feasible. In Section 3.4 we discuss ways to extend our approach to situations when this assumption is violated.

Denote by $\widetilde{\mathbf{Y}} = (\widetilde{\mathbf{Y}}_1, \ldots, \widetilde{\mathbf{Y}}_I)$, where $\widetilde{\mathbf{Y}}_i = (\widetilde{\mathbf{Y}}_{i1}, \ldots, \widetilde{\mathbf{Y}}_{iJ_i})$ is a centered $p \times J_i$ matrix and the column j, $j = 1, \ldots, J_i$, contains the unfolded image for subject i at visit j. Note that the matrix $\widetilde{\mathbf{Y}}_i$ contains all the data for subject i with each column corresponding to a particular visit. The matrix $\widetilde{\mathbf{Y}}$ is the $p \times n$ matrix is not directly observed and obtained by column-binding the centered subject-specific data matrices $\widetilde{\mathbf{Y}}_i$. Thus if $\widetilde{\mathbf{Y}}_i = (\widetilde{\mathbf{Y}}_{i1}, \ldots, \widetilde{\mathbf{Y}}_{iJ_i})$ then $\widetilde{\mathbf{Y}} = (\widetilde{\mathbf{Y}}_1, \ldots, \widetilde{\mathbf{Y}}_I)$. Our approach starts with constructing the SVD of the matrix $\widetilde{\mathbf{Y}}$

$$\widetilde{\mathbf{Y}} = \mathbf{V}\mathbf{S}^{1/2}\mathbf{U}'. \tag{8}$$

Here, the matrix $\mathbf{V}$ is $p \times n$ dimensional with n orthonormal columns, $\mathbf{S}$ is a diagonal $n \times n$ dimensional matrix and $\mathbf{U}$ is an $n \times n$ dimensional orthogonal matrix. Calculating the SVD of $\widetilde{\mathbf{Y}}$ requires only a number of operations linear in the number of parameters p. Indeed, consider the $n \times n$ symmetric matrix $\widetilde{\mathbf{Y}}'\widetilde{\mathbf{Y}}$ with its spectral decomposition $\widetilde{\mathbf{Y}}'\widetilde{\mathbf{Y}} = \mathbf{U}\mathbf{S}\mathbf{U}'$. Note that for high-dimensional p the matrix $\widetilde{\mathbf{Y}}$ cannot be loaded into the memory. The solution is to partition it into L slices as $\widetilde{\mathbf{Y}}' = [(\widetilde{\mathbf{Y}}^1)'|(\widetilde{\mathbf{Y}}^2)'|\ldots|(\widetilde{\mathbf{Y}}^L)']$, where the size of the lth slice, $\widetilde{\mathbf{Y}}^l$, is $(p/L) \times n$ and can be adapted to the available computer memory and optimized to reduce implementation time. The matrix $\widetilde{\mathbf{Y}}'\widetilde{\mathbf{Y}}$ is then calculated as $\sum_{l=1}^{L}(\widetilde{\mathbf{Y}}^l)'\widetilde{\mathbf{Y}}^l$.

From the SVD (8) the $p \times n$ matrix $\mathbf{V}$ can be obtained as $\mathbf{V} = \widetilde{\mathbf{Y}}\mathbf{U}\mathbf{S}^{-1/2}$. The actual calculations can be performed on the slices of the partitioned matrix $\widetilde{\mathbf{Y}}$ as $\mathbf{V}^l = \widetilde{\mathbf{Y}}^l\mathbf{U}\mathbf{S}^{-1/2}, l = 1, \ldots, L$. The concatenated slices $[(\mathbf{V}^1)'|(\mathbf{V}^2)'|\ldots|(\mathbf{V}^L)']$ form the matrix of the left singular vectors $\mathbf{V}'$. Therefore, the SVD (8) can be constructed with a sequential access to the data $\widetilde{\mathbf{Y}}$ with p-linear effort.

After obtaining the SVD decomposition of $\widetilde{\mathbf{Y}}$, each image $\widetilde{\mathbf{Y}}_{ij}$ can be represented as $\widetilde{\mathbf{Y}}_{ij} = \mathbf{V}\mathbf{S}^{1/2}\mathbf{U}_{ij}$, where $\mathbf{U}_{ij}$ is a corresponding column of matrix $\mathbf{U}'$. Therefore, the vectors $\widetilde{\mathbf{Y}}_{ij}$ differ only via the vector factors $\mathbf{U}_{ij}$ of dimension $n \times 1$. Comparing this SVD representation of $\widetilde{\mathbf{Y}}_{ij}$ with the right-hand side of (4), it follows that *cross-sectional and longitudinal variability controlled by the principal scores $\boldsymbol{\xi}_i$, $\boldsymbol{\zeta}_{ij}$, and time variables T_{ij} must be completely determined by the low-dimensional vectors $\mathbf{U}_{ij}$*. This is the key observation which makes our approach feasible. Below, we provide more intuition behind our approach. The formal argument is presented in Theorem 1.

First, we substitute the left-hand side of (5) with its SVD representation of $\widetilde{\mathbf{Y}}_{ij}$ to get $\mathbf{V}\mathbf{S}^{1/2}\mathbf{U}_{ij} = \boldsymbol{\Phi}^{X,0}\boldsymbol{\xi}_i + T_{ij}\boldsymbol{\Phi}^{X,1}\boldsymbol{\xi}_i + \boldsymbol{\Phi}^{W}\boldsymbol{\zeta}_{ij}$. Now we can multiply by $\mathbf{V}'$ both sides of the equation to get $\mathbf{S}^{1/2}\mathbf{U}_{ij} = \mathbf{V}'\boldsymbol{\Phi}^{X,0}\boldsymbol{\xi}_i + T_{ij}\mathbf{V}'\boldsymbol{\Phi}^{X,1}\boldsymbol{\xi}_i + \mathbf{V}'\boldsymbol{\Phi}^{W}\boldsymbol{\zeta}_{ij}$. If we denote $\mathbf{A}^{X,0} = \mathbf{V}'\boldsymbol{\Phi}^{X,0}$ of size $n \times N_X$, $\mathbf{A}^{X,1} = \mathbf{V}'\boldsymbol{\Phi}^{X,1}$ of size $n \times N_X$, and $\mathbf{A}^{W} = \mathbf{V}'\boldsymbol{\Phi}^{U}$ of size $n \times N_W$, we obtain

$$\mathbf{S}^{1/2}\mathbf{U}_{ij} = \mathbf{A}^{X,0}\boldsymbol{\xi}_i + T_{ij}\mathbf{A}^{X,1}\boldsymbol{\xi}_i + \mathbf{A}^{W}\boldsymbol{\zeta}_{ij}. \tag{9}$$

Conditionally on the observed data $\widetilde{\mathbf{Y}}$, models (5) and (9) are equivalent. Indeed, model (5) is a *linear* model that we want to fit to n vectors $\widetilde{\mathbf{Y}}_{ij}$'s. These vectors span an (at most) n-dimensional linear subspace. Hence, the columns of matrix $\mathbf{V}$, the right singular vectors of $\widetilde{\mathbf{Y}}$, could be thought of as an orthonormal basis with $\mathbf{S}^{1/2}\mathbf{U}_{ij}$ being the coordinates of $\widetilde{\mathbf{Y}}_{ij}$ with respect to this basis. Multiplication by $\mathbf{V}'$ can be seen as

a linear mapping from model (5) restricted to observed data $\widetilde{\mathbf{Y}}'_{ij}s$ to model (9). Thus, even though $\mathbf{V}\mathbf{V}' \neq \mathbf{I}_p$, the projection defined by $\mathbf{V}$ does not lose any information about the linear mixed model (5) as the original data vectors $\widetilde{\mathbf{Y}}_{ij}$'s can be recovered using the identity $\mathbf{V}\mathbf{V}'\widetilde{\mathbf{Y}}_{ij} = \widetilde{\mathbf{Y}}_{ij}$. Hence, model (9) has an "intrinsic" dimensionality that is induced by the study sample size n. We can estimate the low-dimensional model (9) using the LFPCA methods developed in Section 2. This step is now feasible as it requires only $O(n^3)$ calculations. The formal result presented below shows that fitting model (9) is an essential step for getting the high-dimensional principal components in a p-linear time.

Theorem 1: *The eigenvectors of the estimated covariance operators (7) can be calculated as* $\hat{\mathbf{\Phi}}^{X,0} = \mathbf{V}\hat{\mathbf{A}}^{X,0}$, $\hat{\mathbf{\Phi}}^{X,1} = \mathbf{V}\hat{\mathbf{A}}^{X,1}$, $\hat{\mathbf{\Phi}}^{W} = \mathbf{V}\hat{\mathbf{A}}^{W}$, *where matrices* $\hat{\mathbf{A}}^{X,0}$, $\hat{\mathbf{A}}^{X,1}$, $\hat{\mathbf{A}}^{W}$ *are obtained from fitting model (9). The estimated matrices of eigenvalues* $\hat{\mathbf{\Lambda}}^{X}$ *and* $\hat{\mathbf{\Lambda}}^{W}$ *are the same for both model (4) and model (9).*

The proof of the theorem is given in Appendix. Note that the construction above is a generalization of the HD-MFPCA result in Zipunnikov et al. (2011b), which was obtained in the case when there is no longitudinal component $\mathbf{\Phi}^{X,1}$. In the next section, we provide more insights into the intrinsic model (9).

3.2 Estimation of principal scores

Principal scores play the role of the coordinates of $\widetilde{\mathbf{Y}}_{ij}$ in the basis defined by the LFPCA model (5). Therefore, it is essential to estimate the principal scores to make inferences based on principal components. In this section, we propose an approach to calculating the BLUPS of the scores that is computationally feasible for samples of high-resolution images.

First, we introduce some notations. In Section 3.1, we showed that the SVD of the matrix $\widetilde{\mathbf{Y}}$ can be written in by-subject blocks as $\widetilde{\mathbf{Y}}_i = \mathbf{V}\mathbf{S}^{1/2}\mathbf{U}'_i$, where the $n \times J_i$ matrix $\mathbf{U}'_i$ corresponds to the subject i. Model (5) can be rewritten as

$$vec(\widetilde{\mathbf{Y}}_i) = \mathbf{B}_i\boldsymbol{\omega}_i, \tag{10}$$

where $\mathbf{B}_i = [\mathbf{B}_i^X : \mathbf{B}_i^W]$, $\mathbf{B}_i^X = \mathbf{1}_{J_i} \otimes \mathbf{\Phi}^{X,0} + \mathbf{T}_i \otimes \mathbf{\Phi}^{X,1}$, $\mathbf{B}_i^W = \mathbf{I}_{J_i} \otimes \mathbf{\Phi}^{W}$, $\mathbf{T}_i = (T_{i1}, \ldots, T_{iJ_i})'$, $\boldsymbol{\omega}_i = (\boldsymbol{\xi}_i', \boldsymbol{\zeta}_i')'$, the subject level principal scores $\boldsymbol{\zeta}_i = (\boldsymbol{\zeta}_{i1}', \ldots, \boldsymbol{\zeta}_{iJ_i}')'$, $\mathbf{1}_{J_i}$ is a $J_i \times 1$ vector of ones, $\otimes$ is the Kronecker product of matrices, and operation $vec(\cdot)$ stacks the columns of a matrix on top of each other. The following theorem contains the main result of this section; it shows how the estimated BLUPs can be efficiently calculated for the LFPCA model in the context of high dimensional data.

Theorem 2: *Under the LFPCA model (3), the estimated best linear unbiased predictor (EBLUP) of* $\boldsymbol{\xi}_i$ *and* $\boldsymbol{\zeta}_i$ *is given by*

$$\begin{pmatrix} \hat{\boldsymbol{\xi}}_i \\ \hat{\boldsymbol{\zeta}}_i \end{pmatrix} = (\hat{\mathbf{B}}_i'\hat{\mathbf{B}}_i)^{-1}\hat{\mathbf{B}}_i' vec(\widetilde{\mathbf{Y}}_i), \tag{11}$$

where matrix factors above can be written in terms of low-dimensional right singular

vectors as

$$\hat{\mathbf{B}}_i'\hat{\mathbf{B}}_i = \begin{pmatrix} J_i\hat{\mathbf{C}}_{00} + T_{i\cdot}^2\hat{\mathbf{C}}_{11} + T_{i\cdot}(\hat{\mathbf{C}}_{10} + \hat{\mathbf{C}}_{01}) & \mathbf{1}_{J_i}' \otimes \hat{\mathbf{C}}_{0W} + \mathbf{T}_i' \otimes \hat{\mathbf{C}}_{1W} \\ \mathbf{1}_{J_i} \otimes \hat{\mathbf{C}}_{W0} + \mathbf{T}_i \otimes \hat{\mathbf{C}}_{W1} & \mathbf{I}_{N_W J_i} \end{pmatrix}$$

and

$$\hat{\mathbf{B}}_i' vec(\widetilde{\mathbf{Y}}_i) = \begin{pmatrix} \hat{\mathbf{A}}^{X,0}\mathbf{S}^{1/2}\mathbf{U}_i'\mathbf{1}_{J_i} + \hat{\mathbf{A}}^{X,1}\mathbf{S}^{1/2}\mathbf{U}_i'\mathbf{T}_i \\ vec(\hat{\mathbf{A}}^{W}\mathbf{S}^{1/2}\mathbf{U}_i') \end{pmatrix},$$

where $T_{i\cdot} = \sum_{j=1}^{J_i} T_{ij}$, $T_{i\cdot}^2 = \sum_{j=1}^{J_i} T_{ij}^2$, $\hat{\mathbf{C}}_{ks} = (\hat{\mathbf{A}}^{X,k})'\hat{\mathbf{A}}^{X,s}$ *for* $k, s \in \{0,1\}$, $\hat{\mathbf{C}}_{sW} = (\hat{\mathbf{A}}^{X,s})'\hat{\mathbf{A}}^{W}$, $\hat{\mathbf{C}}_{Ws} = \hat{\mathbf{C}}_{sW}'$ *for* $s = \{0,1\}$.

The proof of the theorem is given in Appendix. It is interesting to note that the EBLUPs given by Theorem 1 for random effects $\boldsymbol{\omega}_i$ in the linear mixed model (10) coincide with the subject-specific OLS estimator which would be calculated if $\boldsymbol{\omega}_i$ were fixed parameters (more details are given in Appendix). The EBLUPs calculations are almost instantaneous as the matrices involved in (11) are low-dimensional and do not depend on the dimension p. This is due to the crucial link between the full model (4) and the intrinsic model (9). Although the proof is based on the full model (4), the exact same result can be obtained from the intrinsic model (9). Indeed, the BLUPs can be seen as a projection and does not require distributional assumptions. Model (9) is an orthogonal mapping of (4). Hence, the projection argument of the EBLUPs proof can be applied to (9). In other words, the intrinsic model is fully informative for calculating the EBLUPs of the full model (4).

Informally, the result of Theorem 2 can be obtained from the result given in Section 3.3 of Greven et al. (2010), under the assumption of no measurement noise. However, the formal derivation in this paper is based on a projection argument of Harville (1976) (see Appendix for more details). It is also worth mentioning that model (4) collapses to the MFPCA model developed in Di et al. (2008) if there is no longitudinal component $\boldsymbol{\Phi}^{X,1}$. In that case, the EBLUPs of Theorem 2 are exactly the EBLUPs derived in Zipunnikov et al. (2011b) for MFPCA.

3.3 The general functional mixed model

A natural way to generalize model (1) is to consider the following model

$$Y_{ij}(v) = \eta(v, T_{ij}) + Z_{ij,0}X_{i,0}(v) + Z_{ij,1}X_{i,1}(v) + \ldots + Z_{ij,q}X_{i,q}(v) + W_{ij}(v), \qquad (12)$$

where the $(q+1)$-dimensional vector of covariates $\mathbf{Z}_{ij} = (Z_{ij,0}, Z_{ij,1}, \ldots, Z_{ij,q})$ may include, for instance, polynomial terms of T_{ij} and/or some other covariates of interest. This model can be reduced to the form similar to (4) as

$$\widetilde{\mathbf{Y}}_{ij} = Z_{ij,0}\sum_{k=1}^{N_X}\xi_{ik}\boldsymbol{\phi}_k^{X,0} + Z_{ij,1}\sum_{k=1}^{N_X}\xi_{ik}\boldsymbol{\phi}_k^{X,1} + \ldots + Z_{ij,q}\sum_{k=1}^{N_X}\xi_{ik}\boldsymbol{\phi}_k^{X,q} + \sum_{l=1}^{N_W}\zeta_{ijl}\boldsymbol{\phi}_l^{W}. \qquad (13)$$

Computationally this model can be treated in the way described above in Sections 3.1 and 3.2. Particularly, one would need to get the SVD decomposition (8) for the matrix $\widetilde{\mathbf{Y}}$. As before, the right singular vectors $\mathbf{U}_{ij}$ would contain the longitudinal information about $\boldsymbol{\xi}_i$, $\boldsymbol{\zeta}_i$, and covariates $\mathbf{Z}_{ij}$. Hence, the intrinsic model

$$\mathbf{S}^{1/2}\mathbf{U}_{ij} = Z_{ij,0}\sum_{k=1}^{N_X}\xi_{ik}\mathbf{A}_k^{X,0} + Z_{ij,1}\sum_{k=1}^{N_X}\xi_{ik}\mathbf{A}_k^{X,1} + \ldots + Z_{ij,q}\sum_{k=1}^{N_X}\xi_{ik}\mathbf{A}_k^{X,q} + \sum_{l=1}^{N_W}\zeta_{ijl}\mathbf{A}_l^{W} \quad (14)$$

can provide all necessary quantities, which coupled with $\mathbf{V}$ allow to calculate high-dimensional principal components. With a small change in notations, model (14) can be rewritten in the form (10) and the EBLUPs for principal scores can be derived using the exact same argument as in Theorem 2.

3.4 Large sample size

The main assumption which has been made throughout this section is that the sample size, $n = \sum_{j=1}^{I} J_i$, is sufficiently small to guarantee that calculations of order $O(n^3)$ are feasible. Below we briefly describe how our framework can be adapted to settings with many more scans - on the order of tens or hundreds of thousands.

LFPCA equation (4) models each vector $\widetilde{\mathbf{Y}}_{ij}$ as a linear combination of columns of matrices $\boldsymbol{\Phi}^{X,0}, \boldsymbol{\Phi}^{X,1}, \boldsymbol{\Phi}^{W}$. Hence, each $\widetilde{\mathbf{Y}}_{ij}$ belongs to an at most $(2N_X+N_W)$-dimensional linear space $\mathfrak{L}(\boldsymbol{\Phi}^{X,0}, \boldsymbol{\Phi}^{X,1}, \boldsymbol{\Phi}^{W})$ spanned by those columns. Thus, if model (4) holds exactly the rank of the matrix $\widetilde{\mathbf{Y}}$ does not exceed $(2N_X + N_W)$ and at most $2N_X + N_W$ columns of $\mathbf{V}$ correspond to non-zero singular values. This implies that the intrinsic model (9) can be obtained by projecting onto the first $2N_X + N_W$ columns of $\mathbf{V}$ and the sizes of matrices $\mathbf{A}^{X,0}, \mathbf{A}^{X,1}, \mathbf{A}^{W}$ in (9) will be $(2N_X + N_W) \times N_X$, $(2N_X + N_W) \times N_X$, and $(2N_X + N_W) \times N_W$, respectively. Therefore, the most computationally intensive part would require finding the first $2N_X + N_W$ left singular vectors of $\widetilde{\mathbf{Y}}$. Of course, in practice, model (4) never holds exactly. Hence, the number of columns of matrix $\mathbf{V}$ should be chosen to be large enough to either reasonably exceed $(2N_X + N_W)$ or to capture the most of variability in data. The latter can be estimated by tracking down the sums of the squares of the corresponding first singular vectors. Thus, this provides a constructive way to handle situations when n is too large to calculate the SVD of $\widetilde{\mathbf{Y}}$.

There are a few computationally efficient ways to calculate the first k singular vectors of a large matrix. One way is to adapt streaming algorithms (Weng et al., 2003; Zhao et al., 2006; Budavari et al., 2009). These algorithms usually require only one pass through the data matrix $\widetilde{\mathbf{Y}}$ during which information about the first k singular vectors is accumulated sequentially. Their complexity is of order $O(k^3 p)$. An alternate approach is to use iterative power methods (see, for example, Roweis, 1997). As the dimension of the intrinsic model, $2N_X + N_W$, is not known in advance, the number of left singular vectors to keep and project onto can be adaptively estimated based on the singular values of the matrix $\widetilde{\mathbf{Y}}$. The further development in this direction is beyond the scope of this paper.

4 Simulations

In this section, two simulation studies illustrate the developed methods. In the first one, we replicate a simulation scenario in Greven et al. (2010) for functional curves, but we focus on using a grid that is two orders of magnitude larger than the one in the original scenario. This increase in dimensionality could not be handled by the original LFPCA approach. In the second simulation study, we explore how our methods recover 3D spatial bases in a case where the approach of Greven et al. (2010) cannot be implemented. Both studies were run on a four core i7-2.67Gz PC with 6Gb of RAM memory using Matlab 2010a. The software is available upon request.

In our first scenario we follow Greven et al. (2010) and generate data as follows

$$\begin{cases} Y_{ij}(v) = \sum_{k=1}^{N_X} \xi_{ik}\phi_k^{X,0}(v) + T_{ij}\sum_{k=1}^{N_X} \xi_{ik}\phi_k^{X,1}(v) + \sum_{l=1}^{N_W} \zeta_{ijl}\phi_l^W(v) + \varepsilon_{ij}(v), v \in \mathcal{V} \\ \xi_{ik} \overset{i.i.d.}{\sim} 0.5N(-\sqrt{\lambda_k^X}/2, \lambda_k^X/2) + 0.5N(\sqrt{\lambda_k^X}/2, \lambda_k^X/2), \\ \zeta_{ijl} \overset{i.i.d.}{\sim} 0.5N(-\sqrt{\lambda_l^W}/2, \lambda_l^W/2) + 0.5N(\sqrt{\lambda_l^W}/2, \lambda_l^W/2), \end{cases}$$

where $\xi_{ik} \overset{i.i.d.}{\sim} 0.5N(-\sqrt{\lambda_k^X}/2, \lambda_k^X/2) + 0.5N(\sqrt{\lambda_k^X}/2, \lambda_k^X/2)$ means that scores ξ_{ik} are drawn from a mixture of two normals, $N(-\sqrt{\lambda_k^X}/2, \lambda_k^X/2)$ and $N(\sqrt{\lambda_k^X}/2, \lambda_k^X/2)$ with equal probabilities, similarly for ζ_{ijl}. Scores ξ_{ik}'s and ζ_{ijl}'s are mutually independent. We set $I = 100$ and $J_i = 4, i = 1, \ldots, I$, the number of eigenfunctions $N_X = N_W = 4$. The true eigenvalues are the same, $\lambda_k^X = \lambda_k^W = 0.5^{k-1}, k = 1, 2, 3, 4$. The non-orthogonal bases were chosen as

$$\begin{array}{lll} \phi_1^{X,0}(v) = \sqrt{2/3}\sin(2\pi v), & \phi_1^{X,1}(v) = 1/2, & \phi_1^W = \sqrt{4}\phi_1^{X,1}, \\ \phi_2^{X,0}(v) = \sqrt{2/3}\cos(2\pi v), & \phi_2^{X,1}(v) = \sqrt{3}(2v-1)/2, & \phi_2^W = \sqrt{4/3}\phi_1^{X,0}, \\ \phi_3^{X,0}(v) = \sqrt{2/3}\sin(4\pi v), & \phi_3^{X,1}(v) = \sqrt{5}(6v^2-6v+1)/2, & \phi_3^W = \sqrt{4/3}\phi_2^{X,0}, \\ \phi_4^{X,0}(v) = \sqrt{2/3}\cos(4\pi v), & \phi_4^{X,1}(v) = \sqrt{7}(20v^3-30v^2+12v-1)/2, & \phi_4^W = \sqrt{4/3}\phi_3^{X,0}, \end{array}$$

which are measured on a regular grid of p equidistant points in the interval $[0, 1]$. Compared with Greven et al. (2010), we increased the sampling grid by a hundred times and set $p = 12,000$. Note that a brute-force extension of standard LFPCA would be at the edge of feasibility for such a large p. For each i, the first time T_{i1} is generated from the uniform distribution over interval (0,1) denoted by $U(0, 1)$. Then differences $(T_{ij+1} - T_{ij})$ are also generated from $U(0, 1)$ for $1 \leq j \leq 3$. The times $T_{i1}, \ldots, T_{i4}$ are normalized to have sample mean zero and variance one. Finally, $\varepsilon_{ij}(v)$ are assumed to be i.i.d $N(0, 0.01^2)$ for each i, j, v. Although, no measurement noise is assumed in (3), it is included in this simulation scenario. The purpose of having $\varepsilon_{ij}(v)$ is twofold. First, it is of interest to explore how the presence of measurement noise affects the performance of our methods which ignore this possibility. Second, we want to be as close as possible to the setup in Greven et al. (2010). However, the choice of the eigenfunctions in the original simulation scenario of Greven et al. (2010) makes the estimation problem ill-posed if there is no measurement noise. Therefore, the measurement noise essentially regularizes the estimation problem. In this study, we simulated 100 data sets according to the scenario above and ran our estimating procedures. The simulation study took 2.5

minutes, that is 1.5 seconds per data set.

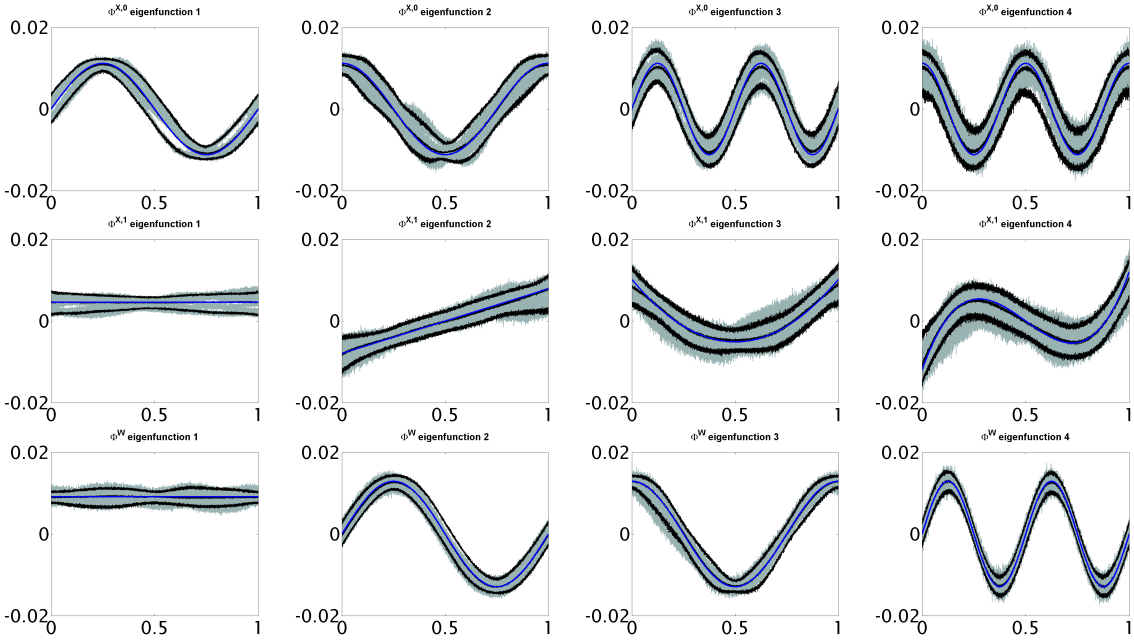

Figure 3: *True and estimated eigenfunctions* ($\phi_k^{X,0}$, $\phi_k^{X,1}$), *and* ϕ_l^W *for scenario* 1 *replicated* 100 *times. Each box shows* 10 *randomly chosen estimated functions (grey), the true function (solid blue line), the pointwise median and the* 5^{th} *and* 95^{th} *pointwise percentile curves (dashed black lines).*

First, we display the true and estimated eigenfunctions in Figure 3. The results for $\phi_k^{X,0}$ are displayed in the top panel, for $\phi_k^{X,1}$ in the middle panel, and for ϕ_k^W in the bottom panel. The grid in this study is very dense and includes 12,000 points. To make plots distinguishable, we show only 10 randomly selected estimated functions (gray lines). The true function (solid blue line), the pointwise median of estimated eigenvectors (indistinguishable from the true functions) and the pointwise 5th and 95th percentiles of estimated eigenvectors (black dashed lines). As we can see, if the amount of noise is moderate relative to the main signal, as in this scenario, then there is not any noticeable effect on our estimation procedure. Comparing Figure 3 with Figure 2 in Greven et al. (2010) we conclude that our estimation procedure completely reproduce the eigenfunction results obtained using the standard LFPCA approach.

The boxplots of the estimated eigenvalues of the processes $X_i(v)$ and $W_{ij}(v)$ are displayed in Figure 4. The centered and standardized eigenvalues, $(\hat{\lambda}_k^X - \lambda_k^X)/\lambda_k^X$ and $(\hat{\lambda}_l^W - \lambda_l^W)/\lambda_l^W$, are given on the left and the right panels of Figure 4, respectively. The amount of consistent bias coming from the presence of measurement noise is so small that it cannot be visually detected. This is explained by a small variance of the measurement noise relative to those of the eigenvalues of $X_i(v)$ and $W_{ij}(v)$. The results are consistent with those reported in Greven et al. (2010) and confirm the good performance of our estimation methods.

The scores ξ_{ik} and ζ_{ijl} are estimated using the EBLUP given by Theorem 2. The total number of the scores ξ_{ik} estimated in the study is $10,000$ for each k, as each generated data

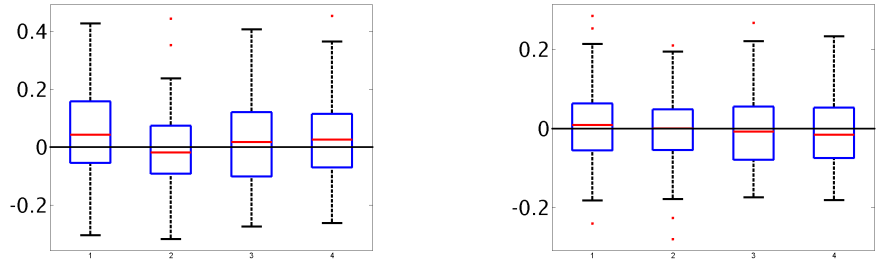

Figure 4: *Boxplots of the normalized estimated eigenvalues for process* $X_i(v)$, $(\hat{\lambda}_k^X - \lambda_k^X)/\lambda_k^X$, *(left box) and the normalized estimated eigenvalues for process* $W_{ij}(v)$, $(\hat{\lambda}_l^W - \lambda_l^W)/\lambda_l^W$, *(right box) based on scenario* 1 *with* 100 *replications. The zero is shown by the solid black line.*

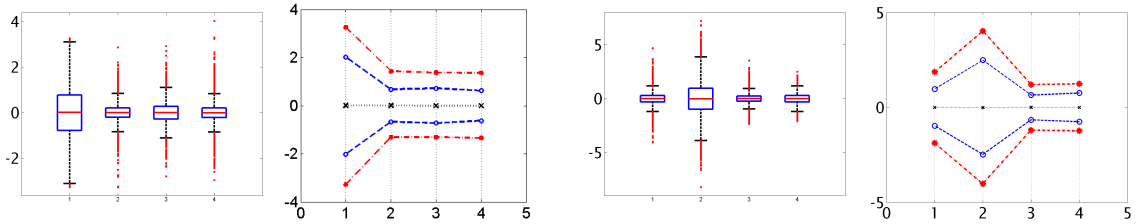

Figure 5: *The left two panels show the distribution of the normalized estimated scores of process* $X_i(v)$, $(\xi_{ik} - \hat{\xi}_{ik})/\sqrt{\lambda_k^X}$. *Boxplots are given in the left column. The right column shows the medians (black marker),* 5% *and* 95% *quantiles (blue markers), and* 0.5% *and* 99.5% *quantiles (red markers). Similarly, the distribution of the normalized estimated scores of process* $W_{ij}(v)$, $(\zeta_{ijl} - \hat{\zeta}_{ijl})/\sqrt{\lambda_l^X}$ *is provided at the right two panels.*

set provides exactly 100 estimates of scores $\xi_{ik}, k = 1, 2, 3, 4$. Similarly, the total number of the estimated scores ζ_{ijl} is $40,000$ for each l. Note that the estimated scores within each replication are dependent even if their theoretical counterparts are independent. Panels one and three of Figure 5 report the boxplots of the normalized estimated scores $(\xi_{ik} - \hat{\xi}_{ik})/\sqrt{\lambda_k^X}$ and $(\zeta_{ijl} - \hat{\zeta}_{ijl})/\sqrt{\lambda_l^W}$, respectively. The distribution of the normalized estimated scores corresponding to the first eigenfunction of $X_i(v)$ as well to the second eigenfunction of $W_{ij}(v)$ has a wider spread around zero. This is likely due to the fact that $\phi_2^W(v) = \sqrt{4/3}\phi_1^{X,0}(v)$ and the estimation of these components is harder than the estimation $\phi_1^W(v) = \sqrt{4}\phi_1^{X,1}(v)$, which has a more strong signal due to the multiplicative factor $\sqrt{4}$. Panels two and four in Figure 5 display the medians, 0.5%, 5%, 95% and 99.5% quantiles of the distribution of the normalized estimated scores. This demonstrates that the estimation procedures based on the theoretical results of Theorem 2 are unbiased. Due to the presence of the measurement noise, there is probably a slightly larger variation in the observed distributions. Some formal comparison of the results for the EBLUPs given in Theorem 2 and those in Greven et al. (2010) can provide a better quantification of those differences. Overall, this study demonstrated that HD-LFPCA replicates the results given by the standard LFPCA in a highly computationally efficient manner.

In the second scenario, our methods are evaluated on 3D data. Particularly, data sets

in this study replicate the 3D ROI blocks which we have in MS data set. We simulated 100 data sets from the model

$$\begin{cases} Y_{ij}(v) = \sum_{k=1}^{N_X} \xi_{ik}\phi_k^{X,0}(v) + T_{ij}\sum_{k=1}^{N_X} \xi_{ik}\phi_k^{X,1}(v) + \sum_{l=1}^{N_W} \zeta_{ijl}\phi_l^W(v), v \in \mathcal{V} \\ \xi_{ik} \overset{i.i.d.}{\sim} N(0, \lambda_k^X) \quad \text{and} \quad \zeta_{ijl} \overset{i.i.d.}{\sim} N(0, \lambda_l^W), \end{cases}$$

where $\mathcal{V} = [1, 38] \times [1, 72] \times [1, 11]$. Eigenimages ($\phi_k^{X,0}$, $\phi_k^{X,1}$), and ϕ_l^W are displayed in Figure 6. The images in this scenario can be thought of as 3D images with voxel intensities on the $[0, 1]$ scale. The voxels within each sub-block (eigenimage) are set to 1 and outside voxels set to zero. There are four blue and red sub-blocks corresponding to $\phi_k^{X,0}$ and $\phi_k^{X,1}$, respectively. The closest to Anterior are $\phi_1^{X,0}$ and $\phi_1^{X,1}$, which have the strongest signal proportional to the largest eigenvalue(variance) λ_1^X. Coming down to Posterior, the signal decreases. The sub-blocks closest to Posterior have the smallest signal, which is proportional to λ_4^X. Eigenimages ϕ_k^W shown in green are ordered in the same way. Note that $\phi_k^{X,0}$ are uncorrelated with ϕ_l^W. However, both $\phi_k^{X,0}$ and ϕ_l^W are correlated with the $\phi_k^{X,1}$'s describing the random slope $X_{i,1}(v)$. We assume that $I = 150$, $J_i = 6, i = 1, \ldots, I$, and the true eigenvalues $\lambda_k^X = 0.5^{k-1}, k = 1, 2, 3$, and $\lambda_l^W = 0.5^{l-1}, l = 1, 2$. The times T_{ij} were generated in the same way as in simulation scenario 1. To apply HD-LFPCA, we unfold each image $\mathbf{Y}_{ij}$ and obtain vectors of size $p = 38 \times 72 \times 11 = 30,096$. The simulation study took 20 minutes or approximately 12 seconds per data set. Figures 14, 15, and 16 display the medians of the estimated

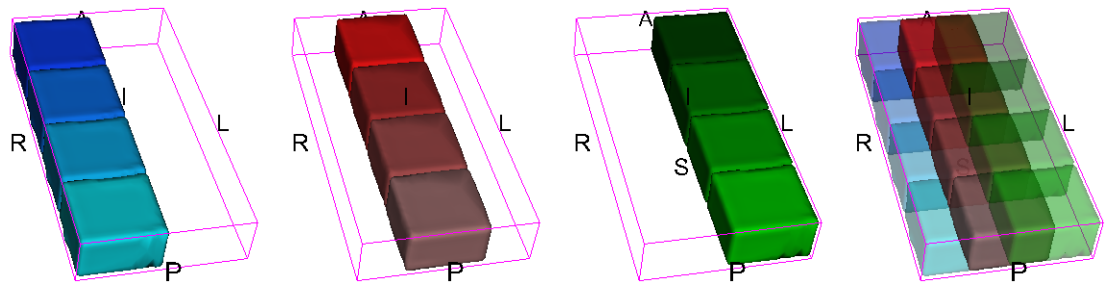

Figure 6: *3D eigenimages of the 2nd simulation scenario. From left to right: $\phi_k^{X,0}$ are in blue, $\phi_k^{X,1}$ are in red, ϕ_k^W are in green, the most right one shows the overlap of all eigenimages. Views: R=Right, L=Left, S=Superior, I=Interior, A=Anterior, P=Posterior. The 3D-renderings are obtained using 3D-Slicer (2011).*

eigenimages and the voxelwise 5th and 95th percentile images, respectively. All axial slices, or z slices in a x-y-z cartesian coordinate system, are the same. It is a result of the data generating process which has no noise and the eigenimages sharing the exact same intensity within each sub-block. Therefore, we display only one z-slice, which is representative of the entire 3D image. To obtain a grayscale image with voxel values in the $[0, 1]$ interval, each estimated eigenvector, $\hat{\boldsymbol{\phi}} = (\hat{\phi}_1, \ldots, \hat{\phi}_p)$, was normalized as $\hat{\boldsymbol{\phi}} \to (\hat{\boldsymbol{\phi}} - \min_s \hat{\phi}_s)/(\max_s \hat{\phi}_s - \min_s \hat{\phi}_s)$. Figure 14 in the web-appendix shows the voxelwise medians of the estimator. The method recovers the spatial configuration of both bases very well. The 5-percentile and 95-percentile images are displayed in Figures 15 and

16 in the web-appendix, respectively. Overall, the original pattern is recovered with some small distortions from the other interacting bases (please note the light gray patches). It is interesting to note that ϕ_l^W's seem not to interact with $\phi_k^{X,0}$'s and $\phi_k^{X,1}$'s. Whereas, $\phi^{X,1}$ seem to be affected by both ϕ_l^W and $\phi_k^{X,0}$. Finally, $\phi_k^{X,0}$ interacts only with ϕ_l^W. We conclude that the estimation of the 3D eigenimages is very good.

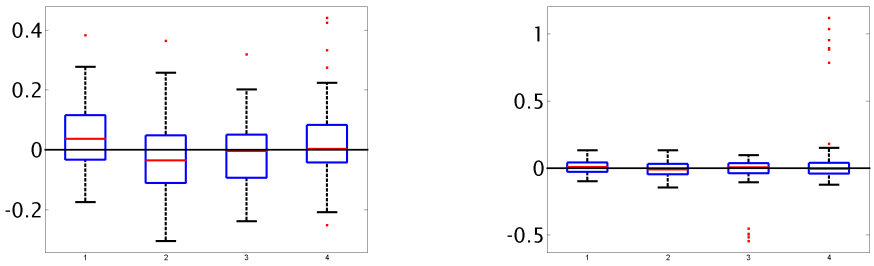

Figure 7: *Boxplots of the centered estimated eigenvalues for process* $X_i(v)$, $(\hat{\lambda}_k^X - \lambda_k^X)/\lambda_k^X$, *(left box) and the normalized estimated eigenvalues for process* $W_{ij}(v)$, $(\hat{\lambda}_l^W - \lambda_l^W)/\lambda_l^W$, *(right box) based on scenario* 2 *with* 100 *replications. The zero is shown by the solid black line.*

The boxplots of the estimated normalized eigenvalues, $(\hat{\lambda}_k^X - \lambda_k^X)/\lambda_k^X$ and $(\hat{\lambda}_l^W - \lambda_l^W)/\lambda_l^W$, are displayed in Figure 7. The eigenvalues are estimated consistently. However, in 6 out of 100 cases (extreme values shown in red), the estimation procedure had a hard time distinguishing between the 3^{rd} and the 4^{th} eigenimages, ϕ_3^W and ϕ_4^W. Remember, though, that the scales for eigenscores ζ_{ij3} and ζ_{ij4} are relative to the corresponding eigenvalues $\lambda_3^W = 0.25$ and $\lambda_4^W = 0.125$. We speculate that it is due to a relatively low signal, which is proportional to eigenvalues λ's. Increasing the sample size should address this issue. Overall, the results illustrate the good and consistent performance of the eigenvalue estimation procedure.

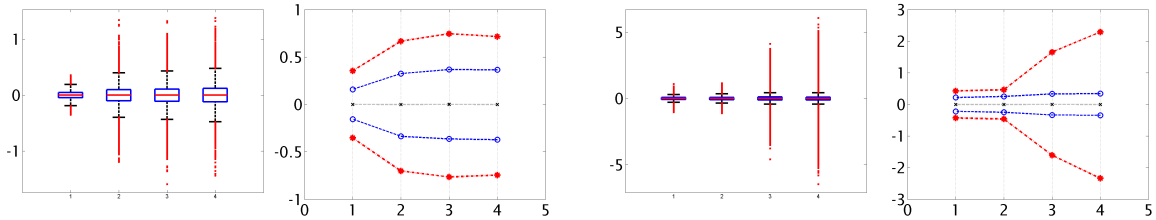

Figure 8: *Left two panels show the distribution of the normalized estimated scores,* $(\xi_{ik} - \hat{\xi}_{ik})/\sqrt{\lambda_k^X}$. *Boxplots are given in the left column. The right column shows the medians (black marker),* 5% *and* 95% *quantiles (blue markers), and* 0.5% *and* 99.5% *quantiles (red markers). Similarly, the distribution of the normalized estimated scores,* $(\zeta_{ijl} - \hat{\zeta}_{ijl})/\sqrt{\lambda_l^W}$ *is provided in the right two panels.*

The boxplots of the estimated eigenscores are displayed in Figure 8. In this scenario, the total number of the estimated scores ξ_{ik} is $15,000$ for each k and there are $90,000$ estimated scores ζ_{ijl} for each l. The distributions of the normalized estimated scores

Distribution of observations per subject												
# observations	1	2	3	4	5	6	7	8	9	10	Total:	466
# of subjects	57	55	15	18	13	14	1	2	0	1	Total:	176

Table 1: *Distribution of observations (scans) per subject*

$(\xi_{ik} - \hat{\xi}_{ik})/\sqrt{\lambda_k^X}$ and $(\zeta_{ijl} - \hat{\zeta}_{ijl})/\sqrt{\lambda_l^W}$ are displayed in first and third panels of Figure 8, respectively. We observe larger spreads of the distributions once a signal-to-noise ratio goes down. The second and fourth panels of Figure 8 display the medians, 0.5%, 5%, 95% and 99.5% quantiles of the distribution of the normalized estimated scores. Results show that the EBLUPs approximate true scores very well.

5 Application

In this section we apply HD-LFPCA to the DTI images of MS patients. The study population included individuals with no, mild, moderate, and severe disability, and over the follow-up period (as long as 5 years in some cases), there was little change in the median disability level of the cohort. Most of the subjects were taking disease-modifying therapies, but only a small fraction were being treated for acute relapses of their disease. Table 1 shows the distribution of number of scans per subject. Cohort characteristics are reported in Table 4 in the web-appendix. The scans have been aligned using a 12 degrees of freedom transformation, meaning that we accounted for rotation, translation, scaling, and shearing, but not for nonlinear deformation. As described in Section 1, the primary region of interest is a central block of the brain of size $38 \times 72 \times 11$ displayed in Figure 1. We weighted each voxel in the block with a probability for the voxel to be in the corpus callosum and study longitudinal changes of weighted voxels in the blocks (Reich et al., 2010). Probabilities less than 0.05 were set to zero. Below we model longitudinal variability of the weighted FA at every voxel of the blocks. The entire analysis performed in Matlab 2010a took only 3 seconds on a PC with a quad core i7-2.67Gz processor and 6Gb of RAM memory.

First, we unfolded each block into a $30,096$ dimensional vector that contained the corresponding weighted FA values. This gave us the data matrix $\mathbf{Y}$ of size 466 by $30,096$, where each row represented a subject/visit observation. In addition to high dimensionality, another difficulty of analyzing this study was the unbalanced distribution of scans across subjects (see Table 1); this is a typical problem in natural history studies.

After forming the data matrix $\mathbf{Y}$, we estimated the overall mean $\hat{\eta} = \frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{J_i}\mathbf{Y}_{ij}$ which is shown at Figure 9. The mean image appears to recover some template form of the corpus callosum averaged over all subjects and visits. In the analysis below, we de-mean the data by subtracting out $\hat{\eta}$ from $\mathbf{Y}_{ij}$'s. The times T_{ij}'s were transformed as follows. First, T_{ij} was set to the age of the person at the time of the jth visit minus the age of the person at the time of the first visit. So T_{i1} was set to zero for everyone. Then we followed a procedure described in Section 3.4 in Greven et al. (2010) and demeaned T_{ij}'s for each subject. After that all T_{ij} were normalized to have sample variance one. This normalizing

procedure allows us to compare the explained variabilities of processes $X_i(v)$ and $W_{ij}(v)$ on the same scale. More details can be found in Section 3.4 of Greven et al. (2010). Equations (7) give unbiased OLS estimators of covariance matrices $\hat{\mathbf{K}}^X$ and $\hat{\mathbf{K}}^W$. Even though the estimated covariance matrices estimate non-negative symmetric matrices, they are not necessarily non-negative definite themselves. We obtained small negative eigenvalues while calculating the spectral decompositions of the covariance operators $\hat{\mathbf{K}}^X$ and $\hat{\mathbf{K}}^W$. Following Hall et al. (2008) all the negative eigenvalues with corresponding eigenvectors were trimmed to zero for the analysis. After this, the total variation was decomposed into "subject-specific" part modeled by process X_i and "exchangeable visit-to-visit" part modeled by process W_{ij}. Most of the total variability, 70.8%, is explained by X_i (subject-specific variability) with the trace of $\mathbf{K}^X = 122.53$, while 29.2% is explained by W_{ij} (exchangeable visit-to-visit variability) with the trace of $\mathbf{K}^W = 50.47$. Two major contributions of our approach is to separate the processes X_i and W_{ij} and quantify their corresponding contributions to the total variability.

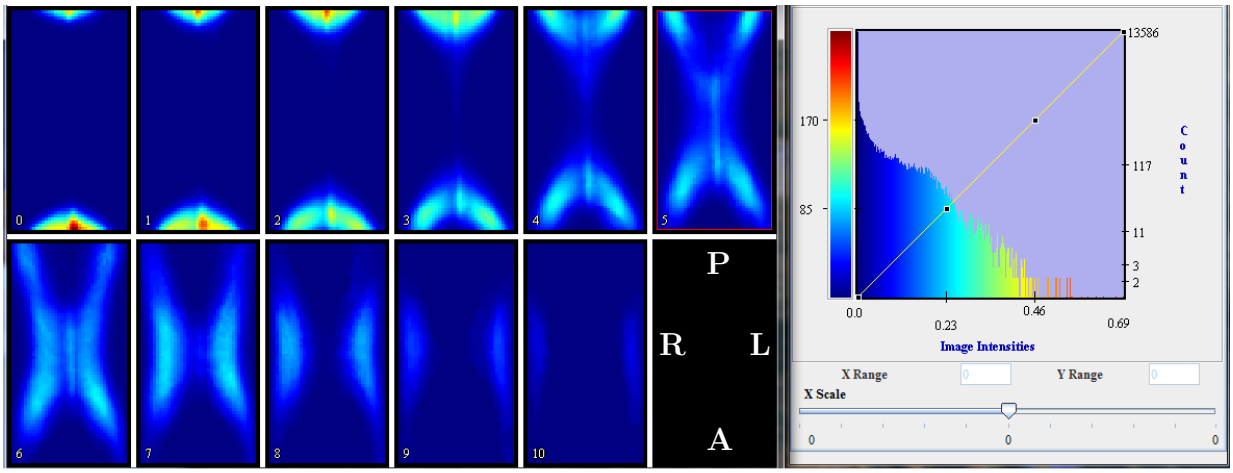

Figure 9: *Mean image $\hat{\eta}$. Eleven axial slices are shown. A histogram of the voxel intensities is on the right. The pictures are obtained using MIPAV (2011).*

Table 2 provides the first 10 eigenvalues of the estimated covariances $\hat{\mathbf{K}}^X$ and $\hat{\mathbf{K}}^W$. Table 3 reports the percentages (rounded to two decimal points) explained by these first 10 eigenimages. The first 10 random intercept eigenimages explain roughly 55% of the total variability. The effect of the random slope is accounting for only 0.81% of the total variability. However, the separation of variability within $X_i(v)$ between the random intercept and the random slope is completely determined by a choice of the scale for times T_{ij}. A different normalizing scheme for T_{ij}'s would result in a different percentage separation of variability within $X_i(v)$. The exchangeable variability captured by $W_{ij}(v)$ accounts for 17.5% of the total variation.

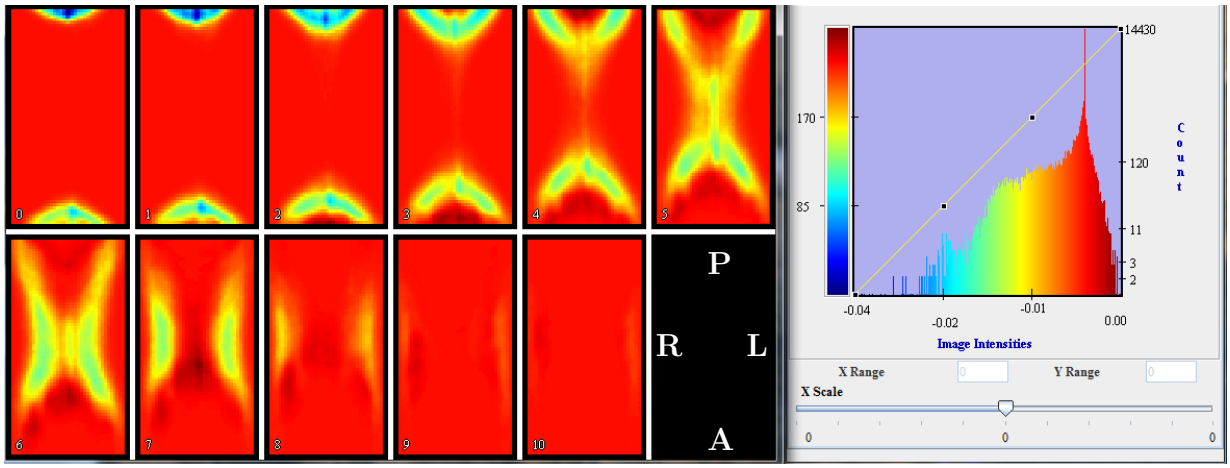

Figure 10: *Eleven slices of* $\hat{\phi}_1^{X,0}$. *A histogram of the voxel intensities is on the right. The pictures are obtained using MIPAV (2011).*

k	λ_k^X	λ_k^W
1	38.31	13.83
2	18.64	5.99
3	10.72	3.82
4	8.59	2.73
5	5.01	1.69
6	4.37	1.55
7	3.58	1.17
8	3.18	0.93
9	2.84	0.83
10	2.21	0.73

Table 2: *The first* 10 *eigenvalues of the estimated covariance matrices* $\hat{\mathbf{K}}^X$ *and* $\hat{\mathbf{K}}^W$.

k	$\phi_k^{X,0}$	$\phi_k^{X,1}$	ϕ_k^W	cumulative
1	22.13	0.08	7.12	29.33
2	10.66	0.11	3.20	43.29
3	5.99	0.13	2.04	51.44
4	4.84	0.08	1.44	57.80
5	2.80	0.06	0.90	61.56
6	2.39	0.07	0.83	64.85
7	1.94	0.10	0.63	67.52
8	1.72	0.08	0.50	69.82
9	1.55	0.05	0.45	71.86
10	1.20	0.05	0.39	73.50
	55.20	0.80	17.50	73.50

Table 3: *Cumulative variability explained by the first* 10 *eigenimages.*

The first three estimated random intercept and slope eigenimages are shown in pairs in Figures 10, 11, and 12, 13, and 17, 18 in the web-appendix, respectively. Figures 19, 20, and 21 in the web-appendix display the first three eigenimages of the exchangeable measurement error process $W_{ij}(v)$. Each eigenimage is accompanied with the histogram of its voxel values. Recall that the eigenimages were obtained by folding the unit length eigenvectors of $p \approx 3 \cdot 10^4$ voxels. Therefore, each voxel is represented by a small value between the negative and positive one. For principal scores, negative and positive voxel values correspond to opposite loadings (directions) of variation. Each histogram has a peak at zero due to the existence of the threshold for the probability maps applied indicating if a voxel is not in the corpus callosum. This peak is a convenient visual divider of the color spectrum into the loading specific colors. Because of the sign invariance of the SVD, the separation between positive and negative loadings is comparable only within the same eigenimage. However, the loadings of the random intercept and slope within an eigenimage of the process $X_i(v)$ can be compared as they share the same principal score. This allows us to contrast the time invariant random intercept with the longitudinal

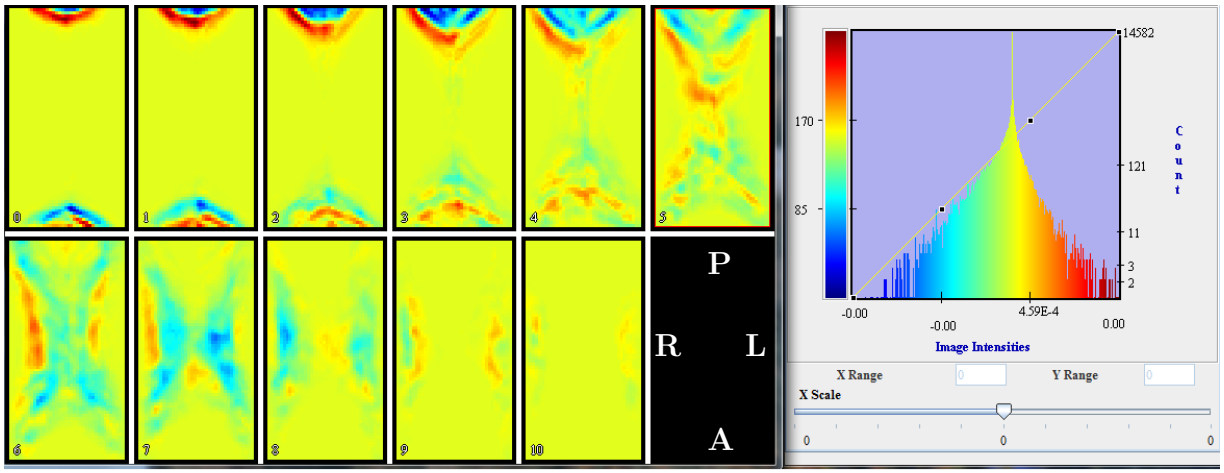

Figure 11: *Eleven slices of $\hat{\phi}_1^{X,1}$. A histogram of the voxel intensities is on the right. The pictures are obtained using MIPAV (2011).*

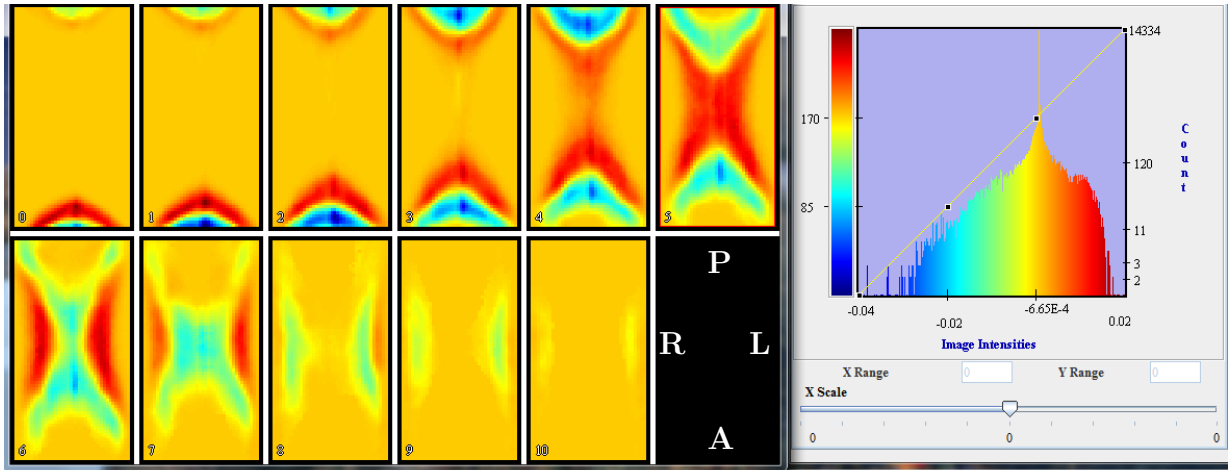

Figure 12: *Eleven slices of $\hat{\phi}_2^{X,0}$. A histogram of the voxel intensities is on the right. The pictures are obtained using MIPAV (2011).*

random slope and, thus, to localize regions that exhibit the largest longitudinal variability. This could be used to analyze the longitudinal changes of brain imaging in a particular disease or to help generate new scientific hypotheses.

We now interpret the random intercept and slope parts of the eigenimages obtained for the MS data. Figures 10 and 11 show the random intercept and slope parts of the first eigenimage ϕ_1^X, respectively. The negatively loaded voxels of the random intercept, $\phi_1^{X,0}$, essentially compose the entire corpus callosum. This indicates an overall shift in the mean FA of the corpus callosum. This is expected and is a widely observed empirical feature of principal components. The positively loaded background voxels of $\phi_1^{X,0}$ are not of any practical interest. The random slope part, $\phi_1^{X,1}$, has both positively and negatively loaded areas in the corpus callosum. The areas colored in blue shades of the spectrum share the sign of the random intercept $\phi_1^{X,0}$ whereas the areas colored in red have the opposite sign. The extreme colors of the spectrum of $\phi_1^{X,1}$ show a clear separation into negative and positive loadings especially accentuated in the splenium (posterior) and the genu (anterior) areas of the corpus callosum; please note the upper and lower areas

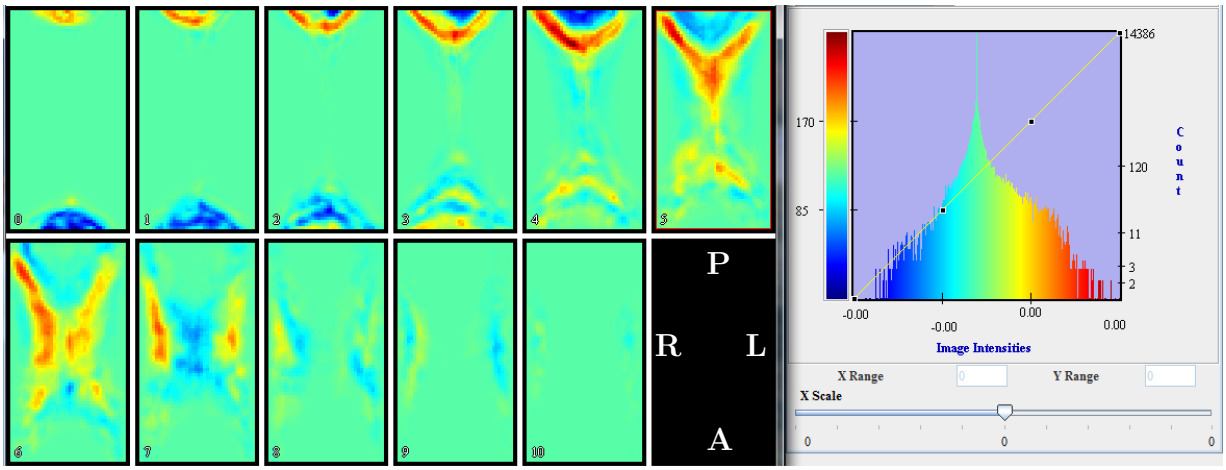

Figure 13: *Eleven slices of $\hat{\phi}_2^{X,1}$. A histogram of the voxel intensities is on the right. The pictures are obtained using MIPAV (2011).*

in panels 0 through 5 of Figure 11. This implies that a subject with a positive first component score $\xi_{i1} > 0$ would tend to have a smaller mean FA over the entire corpus callosum and the FA would tend to decrease with time in the negatively loaded parts of the splenium. The reverse will be true for a subject with a negative score ξ_{i1}. In the second eigenimage, the random slope part, $\phi_2^{X,0}$, (see Figure 12) has both the splenium and the genu areas negatively loaded and the truncus (the part connecting the two) positively loaded. Contrasting it with $\phi_2^{X,1}$ (shown in Figure 13) we see that both the splenium and the truncus are loaded positively while the genu is loaded negatively. For a subject with a positive second component score $\xi_{i2} > 0$ it would translate into a smaller mean FA in the splenium and the genu and a larger mean FA in the truncus. With time, the mean FA would increase in the most intensively colored parts of the splenium and the truncus and decrease in the genu; please see panels 0 through 7 in Figure 13. Again, the effect will be inverted for a negative score ξ_{i2}. The last eigenimage representing the random slope and the random intercept can be analyzed similarly.

In the first eigenimage, ϕ_1^W, representing the process of exchangeable deviations $W_{ij}(v)$, the entire corpus callosum is negatively loaded; please see Figure 19 in the web-appendix. Again, this is expected and can be seen as a shift in the mean FA of the corpus callosum. The intense red shades displayed as an envelop of the corpus callosum arguably aggregate both registration errors and longitudinal deviations that are not captured by the random slope. The second eigenimage, ϕ_2^W, shown in Figure 20 in the web-appendix, indicates a heavier positive loading of the splenium and several spatially continuous regions of the truncus. In the third one, ϕ_3^W, shown in Figure 21 in the web-appendix, only the splenium has some positive loading and the truncus is mostly negatively loaded.

These eigenimages could be used for group comparisons, as MS patients may have different eigenimage-specific loadings from the controls and the difference may depend on the disease stage. Scientific hypothesis could be tested by analyzing the scores describing individual patients. For that, the scores can be estimated by EBLUPs from Section 3.2. Thus, our methods make it possible to both identify and quantify the significant regions and perform necessary follow-up testing.

6 Discussion

The methods developed in this paper increase the scope and general applicability of LF-PCA to very high dimensional settings. The base model decomposes the longitudinal data into three main components: a subject-specific random intercept, a subject-specific random slope, and reversible visit-to-visit deviation. We described and addressed computational difficulties which arise with high-dimensional data in a powerful technique refered to as HD-LFPCA. We demonstrated ways of identifying a low-dimensional space induced by the data that contain all necessary information for estimation of the model. This significantly extended the previous related efforts in the clustered functional principal components models, MFPCA (Di et al., 2008) and HD-MFPCA (Zipunnikov et al., 2011b).

We applied HD-LFPCA to a novel imaging setting considering DTI and MS in a primary white matter structure. Our investigation characterized longitudinal and cross sectional variation in the corpus callosum. Though our research focused on fractional anisotropy, it is immediately applicable to other DTI summary metrics. For future research, it would be of interest to extend the result to multivariate image summaries, such as simultaneously considering FA and mean diffusivity, for example. Even more ambitious, would be functional PCA models on registered tensor images. However, such an approach presents numerous scientific and mathematical obstacles. Scientifically, obtaining registered tensors is difficult. Mathematically, the tensors are complex structures and difficult to analyze without further data reduction.

There are several outstanding issues for HD-LFPCA that continue to need to be addressed. First, a key assumption of our methods is a moderate sample size of the study, say, not exceeding ten thousands of images. This limitation can be circumvented by adapting methods discussed in Section 3.4. A rigorous treatment of this issue will be a goal for future research. Secondly, we have not formally included measurement noise in our model. A simulation study in Section 4 demonstrated that a moderate amount of measurement noise does not have any significant effect. However, a more systematic treatment of the related issues may be required. Smoothing techniques employed for functional curves which aggregate information in a covariance matrix and then smooth it can not be directly extended to images. In high-dimensional settings methods developed in Shabalin and Nobel (2010) could be considered as a powerful and feasible alternative.

In summary, we believe that HD-LFPCA is an important conceptual and practical step in the general applicability of functional principal component analysis to modern high dimensional longitudinal studies.

Supplemental materials

web-appendix.pdf: The file contains some additional images for Sections 4 and 5.

hd-lfpca-simulations-scenario-01.m: Matlab code for Scenario 1 of Simulations.

hd-lfpca-simulations-scenario-02.m: Matlab code for Scenario 2 of Simulations.

Acknowledgments

The authors would like to thank Jeff Goldsmith for his help with data management. The research of Vadim Zipunnikov, Brian Caffo and Ciprian Crainiceanu was supported by grant R01NS060910 from the National Institute of Neurological Disorders and Stroke and by Award Number EB012547 from the NIH National Institute of Biomedical Imaging and Bioengineering (NIBIB). The research of Sonja Greven was funded by the German Research Foundation through the Emmy Noether Programme, grant GR 3793/1-1. The research of Daniel S. Reich was supported by the Intramural Research Program of the National Institute of Neurological Disorders and Stroke. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Neurological Disorders and Stroke or or the National Institute of Biomedical Imaging and Bioengineering or the National Institutes of Health.

References

3D-Slicer (2011), "http://www.slicer.org/," *web-site*.

Aston, J., Chiou, J.-M., and Evans, J. P. (2010), "Linguistic Pitch Analysis using Functional Principal Component Mixed Effect Models," *Journal of the Royal Statistical Society, Series C*, 59, 297–317.

Bigelow, J. L. and Dunson, D. B. (2009), "Bayesian semiparametric joint models for functional predictors," *Journal of the American Statistical Association*, 104, 26–36.

Budavari, T., Wild, V., Szalay, A. S., Dobos, L., and Yip, C.-W. (2009), "Reliable Eigenspectra for New Generation Surveys," *Monthly Notices of the Royal Astronomical Society*, 394(3), 1496–1502.

Crainiceanu, C. M., Caffo, B. S., Luo, S., Zipunnikov, V. V., and Punjabi, N. M. (2011), "Population Value Decomposition, a framework for the analysis of image populations (with discussion)," *Journal of the American Statistical Association*, in press.

Crainiceanu, C. M., Staicu, A.-M., and Di, C.-Z. (2009), "Generalized Multilevel Functional Regression," *Journal of the American Statistical Association*, 104, 488, 1550–1561.

Demmel, J. W. (1997), *Applied Numerical Linear Algebra*, SIAM.

Di, C.-Z., Crainiceanu, C. M., Caffo, B. S., and Punjabi, N. M. (2008), "Multilevel Functional Principal Component Analysis," *Annals of Applied Statistics*, 3(1), 458–488.

Di, C.-Z., Crainiceanu, C. M., and Jank, W. (2010), "Multilevel Sparse Functional Principal Component Analysis," *Technical Report*.

Goldsmith, J., Crainiceanu, C. M., Caffo, B. S., and Reich, D. S. (2011), "Penalized functional regression analysis of white-matter tract profiles in multiple sclerosis," *NeuroImage*, 57(2), 431–439.

Golub, G. H. and Loan, C. V. (1996), *Matrix Computations*, The Johns Hopkins University Press.

Greven, S., Crainiceanu, C. M., Caffo, B. S., and Reich, D. (2010), "Longitudinal Functional Principal Component Analysis," *Electronic Journal of Statistics*, 4, 1022–1054.

Guo, W. (2002), "Functional mixed effects models," *Biometrics*, 58, 121–128.

Hall, P., Muller, H.-G., and Yao, F. (2008), "Modelling sparse generalized longitudinal observations with latent gaussian processes." *Journal of the Royal Statistical Society: Series B*, 70(4), 703–723.

Harville, D. (1976), "Extension of the Gauss-Markov theorem to include the estimation of random effects," *The Annals of Statistics*, 4(2), 384–395.

Karhunen, K. (1947), "Uber Lineare Methoden in der Wahrscheinlichkeitsrechnung," *Annales Academie Scientiarum Fennicae*, 37, 1–79.

Loeve, M. (1978), *Probability theory*, Springer-Verlag.

McCulloch, C. and Searle, S. (2001), *Generalized, Linear, and Mixed Models*, Wiley Interscience.

MIPAV (2011), "http://mipav.cit.nih.gov," *web-site*.

Mohamed, A. and Davatzikos, C. (2004), "Shape Representation via Best Orthogonal Basis Selection," *Medical Image Computing and Computer-Assisted Intervention*, 3216, 225–233.

Mori, S. (2007), *Introduction to Diffusion Tensor Imaging*, Elsevier.

Morris, J. S., Baladandayuthapani, V., Herrick, R. C., Sanna, P., and Gulstein, H. (2011), "Automated analysis of quantitative image data using isomorphic functional mixed models, with application to proteomics data," *Annals of Applied Statistics*, 5(2A), 894–923.

Morris, J. S. and Carroll, R. J. (2006), "Wavelet-based functional mixed models," *Journal of the Royal Statistical Society*, B 68, 179–199.

Pujol, S. (2010), "Data Loading and 3D Visualization," *tutorial, National Alliance for Medical Image Computing (NA-MIC)*.

Raine, C. S., McFarland, H., and Hohlfeld, R. (2008), *Multiple Sclerosis: A Comprehensive Text*, Saunders Ltd.

Reich, D. S., Ozturk, A., Calabresi, P. A., and Mori, S. (2010), "Automated vs conventional tractography in multiple sclerosis: Variablity and correlation with disability," *NeuroImage*, 49, 3047–3056.

Reiss, P. T. and Ogden, R. T. (2008), "Functional Generalized Linear Models with Applications to Neuroimaging," *Poster presentation Workshop on Contemporary Frontiers in High-Dimensional Statistical Data Analysis*.

— (2010), "Functional generalized linear models with images as predictors," *Biometrics*, 66.1, 61–69.

Reiss, P. T., Ogden, R. T., Mann, J., and Parsey, R. V. (2005), "Functional logistic regression with PET imaging data: A voxel-level clinical diagnostic tool," *Journal of Cerebral Blood Flow & Metabolism*, 25, s635.

Rodriguez, A., Dunson, D. B., and Gelfand, A. E. (2009), "Nonparametric functional data analysis through Bayesian density estimation," *Biometrika*, 96, 149–162.

Roweis, S. (1997), "EM algorithms for PCA and SPCA," *Advances in neural information processing systems*, 10, 626–632.

Shabalin, A. and Nobel, A. (2010), "Reconstruction of a Low-rank Matrix in the Presence of Gaussian Noise," *arXiv.org*, 34ps.

Shinohara, R., Crainiceanu, C., Caffo, B., Gaita, M. I., and Reich, D. S. (2011), "Population wide model-free quantification of Blood-Brain-Barrier Dynamics in Multiple Sclerosis," *NeuroImage*, in press, 37ps.

Staicu, A.-M., Crainiceanu, C. M., and Carroll, R. J. (2010), "Fast analysis of spatially correlated multilevel functional data," *Biostatistics*, 11(2), 177–194.

Weng, J., Zhang, Y., and Hwang, W.-S. (2003), "Candid Covariance-Free Incremental Principal Component Analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(8), 1034–1040.

Zhao, H., Yuen, P. C., and Kwok, J. T. (2006), "A Novel Incremental Principal Component Analysis and Its Application for Face Recognition," *IEEE Transactions on Systems, Man, and, Cybernatics, Part B: Cybernatics*, 36(4), 873–886.

Zhu, H., Brown, P. J., and Morris, J. S. (2011), "Robust Adaptive Functional Regression in Functional Mixed Model Framework," *Journal of the American Statistical Association*, 106(495), 1167–1179.

Zipunnikov, V., Caffo, B., Yousem, D. M., Davatzikos, C., Schwartz, B. S., and Crainiceanu, C. M. (2011a), "Functional Principal Component Models for High Dimensional Brain Volumetrics," *Neuroimage*, 58, 772–784.

— (2011b), "Multilevel Functional Principal Component Analysis for High-Dimensional Data," *Journal of the Computational and Graphical Statistics*, to appear.

Appendix

Proof of Theorem 1: Let us denote by $\hat{\mathbf{K}}_{\mathbf{U}}^X$ and $\hat{\mathbf{K}}_{\mathbf{U}}^W$ matrices defined by equations (7) with $\mathbf{S}^{1/2}\mathbf{U}_{ij_1}\mathbf{U}_{ij_2}'\mathbf{S}^{1/2}$ substituted for $\widetilde{\mathbf{Y}}_{ij_1}\widetilde{\mathbf{Y}}_{ij_2}'$. The $2n \times 2n$ dimensional matrix $\hat{\mathbf{K}}_{\mathbf{U}}^X$ and the $n \times n$ dimensional matrix $\hat{\mathbf{K}}_{\mathbf{U}}^W$ are low-dimensional counterparts of $\hat{\mathbf{K}}^X$ and $\hat{\mathbf{K}}^W$, respectively. Using SVD representation $\widetilde{\mathbf{Y}}_{ij} = \mathbf{V}\mathbf{S}^{1/2}\mathbf{U}_{ij}$, the estimated high dimensional covariance matrices can be represented as $\hat{\mathbf{K}}^X = \mathbf{D}\hat{\mathbf{K}}_{\mathbf{U}}^X\mathbf{D}'$ and $\hat{\mathbf{K}}^W = \mathbf{V}\hat{\mathbf{K}}_{\mathbf{U}}^W\mathbf{V}'$, where matrix $\mathbf{D}$ is $2p \times 2n$ dimensional with orthonormal columns defined as

$$\mathbf{D} = \begin{pmatrix} \mathbf{V} & \mathbf{0}_{p\times n} \\ \mathbf{0}_{p\times n} & \mathbf{V} \end{pmatrix}. \tag{15}$$

From the constructive definition of $\mathbf{H}$, it follows that matrices $\hat{\mathbf{K}}_{\mathbf{U}}^X$ and $\hat{\mathbf{K}}_{\mathbf{U}}^W$ are symmetric. Thus, we can construct their spectral decompositions, $\hat{\mathbf{K}}_{\mathbf{U}}^X = \hat{\mathbf{A}}^X\hat{\mathbf{\Lambda}}^X\hat{\mathbf{A}}^{X\prime}$ and $\hat{\mathbf{K}}_{\mathbf{U}}^W = \hat{\mathbf{A}}^W\hat{\mathbf{\Lambda}}^W\hat{\mathbf{A}}^{W\prime}$. Hence, high dimensional covariance matrices can be represented as $\hat{\mathbf{K}}^X = \mathbf{D}\hat{\mathbf{A}}^X\hat{\mathbf{\Lambda}}^X\hat{\mathbf{A}}^{X\prime}\mathbf{D}'$ and $\hat{\mathbf{K}}^W = \mathbf{V}\hat{\mathbf{A}}^W\hat{\mathbf{\Lambda}}^W\hat{\mathbf{A}}^{W\prime}\mathbf{V}'$, respectively. The result of the theorem now follows from the orthonormality of the columns of matrices $\mathbf{D}$ and $\mathbf{V}$.

Proof of Theorem 2. The main idea of the proof is similar to that of Zipunnikov et al. (2011b). We assume that function $\eta(v, T_{ij}) = 0$. From equation (10) it follows that $\boldsymbol{\omega}_i \sim (0, \mathbf{\Lambda}_{\boldsymbol{\omega}})$, where $\mathbf{\Lambda}_{\boldsymbol{\omega}}$ is a covariance matrix of $\boldsymbol{\omega}_i$. When $p \leq N_X + J_i N_W$ the BLUP of $\boldsymbol{\omega}_i$ is given by $\hat{\boldsymbol{\omega}}_i = Cov(\boldsymbol{\omega}_i, vec(\widetilde{\mathbf{Y}}_i))Var(vec(\widetilde{\mathbf{Y}}_i))^{-1}vec(\widetilde{\mathbf{Y}}_i) = \mathbf{\Lambda}_{\boldsymbol{\omega}}\mathbf{B}_i'(\mathbf{B}_i\mathbf{\Lambda}_{\boldsymbol{\omega}}\mathbf{B}_i')^{-1}vec(\widetilde{\mathbf{Y}}_i)$ (see McCulloch and Searle, 2001, Section 9). BLUP is essentially a projection and thus it does not require any distributional assumptions. It may be defined in terms of projection matrix. If $\boldsymbol{\xi}_i$ and $\boldsymbol{\zeta}_{ij}$ are normal then BLUP is the best predictor. When $p > N_X + J_i N_W$ matrix $\mathbf{B}_i\mathbf{\Lambda}_{\boldsymbol{\omega}}\mathbf{B}_i'$ is not invertible and the generalized inverse of $\mathbf{B}_i\mathbf{\Lambda}_{\boldsymbol{\omega}}\mathbf{B}_i'$ is used (Harville, 1976). In that case, $\hat{\boldsymbol{\omega}}_i = \mathbf{\Lambda}_{\boldsymbol{\omega}}\mathbf{B}_i'(\mathbf{B}_i\mathbf{\Lambda}_{\boldsymbol{\omega}}\mathbf{B}_i')^{-}vec(\widetilde{\mathbf{Y}}_i) = \mathbf{\Lambda}_{\boldsymbol{\omega}}^{1/2}(\mathbf{\Lambda}_{\boldsymbol{\omega}}^{1/2}\mathbf{B}_i'\mathbf{B}_i\mathbf{\Lambda}_{\boldsymbol{\omega}}^{1/2})^{-1}\mathbf{\Lambda}_{\boldsymbol{\omega}}^{1/2}\mathbf{B}_i'vec(\widetilde{\mathbf{Y}}_i) = (\mathbf{B}_i'\mathbf{B}_i)^{-1}\mathbf{B}_i'vec(\widetilde{\mathbf{Y}}_i)$. Note that it coincides with the OLS estimator for $\boldsymbol{\omega}_i$ if $\boldsymbol{\omega}_i$ were a fixed parameter. Thus, the estimated BLUPs are given by $\hat{\boldsymbol{\omega}}_i = (\hat{\mathbf{B}}_i'\hat{\mathbf{B}}_i)^{-1}\hat{\mathbf{B}}_i'vec(\widetilde{\mathbf{Y}}_i)$, where

$$\hat{\mathbf{B}}_i'\hat{\mathbf{B}}_i = \begin{pmatrix} J_i\hat{\mathbf{C}}_{00} + T_{i\cdot}^2\hat{\mathbf{C}}_{11} + T_{i\cdot}(\hat{\mathbf{C}}_{10} + \hat{\mathbf{C}}_{01}) & \mathbf{1}_{J_i}' \otimes \hat{\mathbf{C}}_{0W} + \mathbf{T}_i' \otimes \hat{\mathbf{C}}_{1W} \\ \mathbf{1}_{J_i} \otimes \hat{\mathbf{C}}_{W0} + \mathbf{T}_i \otimes \hat{\mathbf{C}}_{W1} & \mathbf{I}_{N_W J_i} \end{pmatrix}$$

and

$$\hat{\mathbf{B}}_i'vec(\widetilde{\mathbf{Y}}_i) = \begin{pmatrix} \hat{\mathbf{A}}^{X,0}\mathbf{S}^{1/2}\mathbf{U}_i'\mathbf{1}_{J_i} + \hat{\mathbf{A}}^{X,1}\mathbf{S}^{1/2}\mathbf{U}_i'\mathbf{T}_i \\ vec(\hat{\mathbf{A}}^W\mathbf{S}^{1/2}\mathbf{U}_i') \end{pmatrix}.$$

Web-appendix to
"**Longitudinal High-Dimensional Data**"

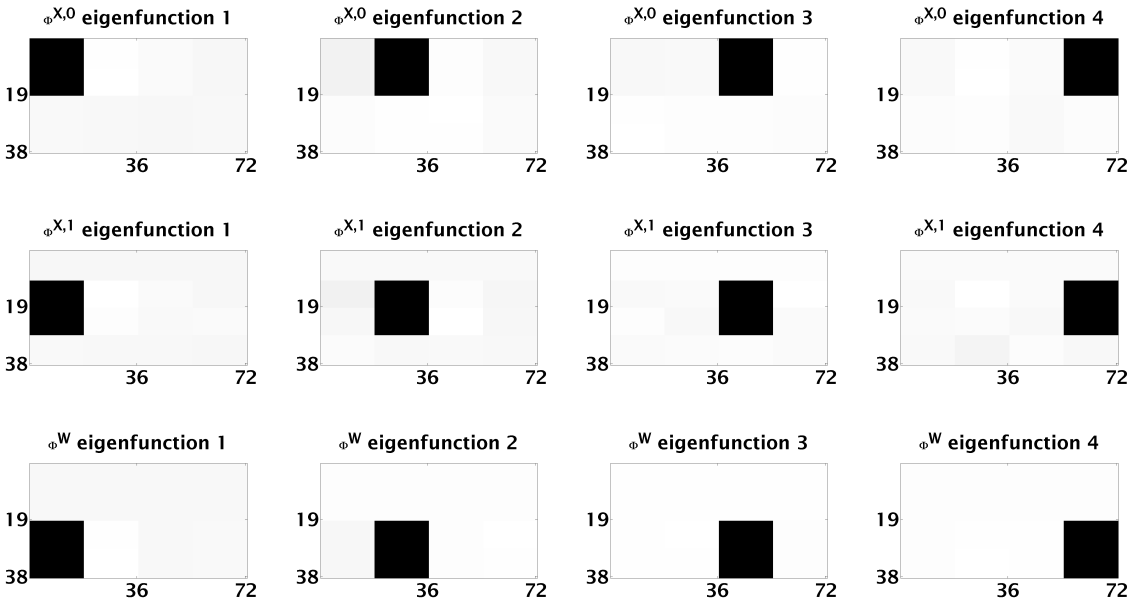

Figure 14: *Voxel-wise medians of estimated* $\phi_k^{X,0}$, $\phi_k^{X,1}$, *and* ϕ_l^W *in the 2nd scenario.*

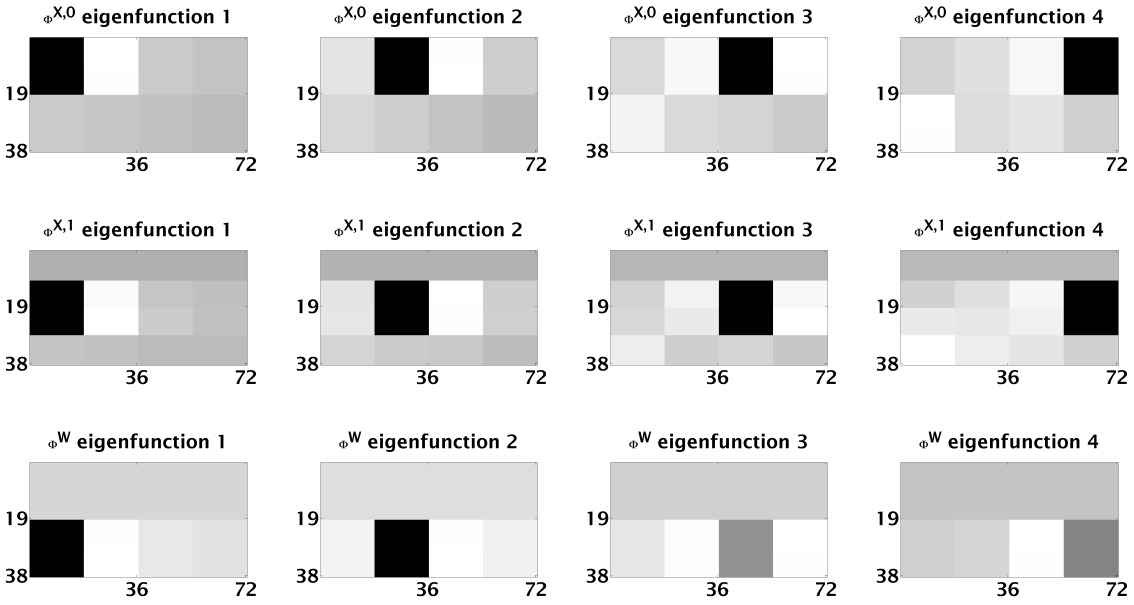

Figure 15: *5th voxel-wise quantiles of estimated* $\phi_k^{X,0}$, $\phi_k^{X,1}$, *and* ϕ_l^W *in the 2nd scenario.*

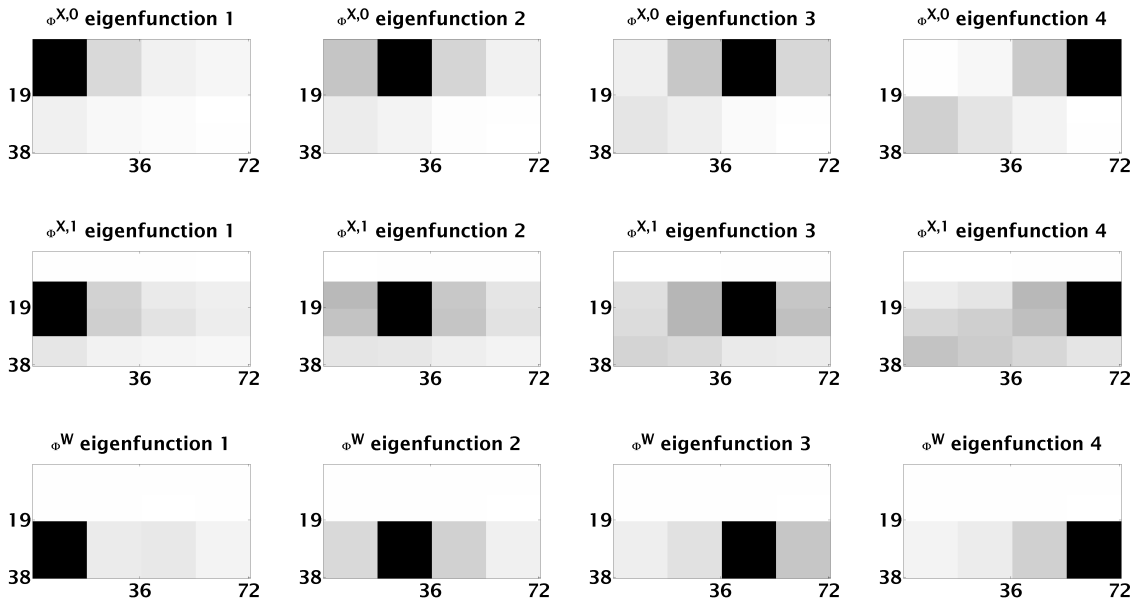

Figure 16: *95th voxel-wise quantiles of estimated $\phi_k^{X,0}$, $\phi_k^{X,1}$, and ϕ_l^W in the 2nd scenario.*

Cohort characteristics		
	baseline	follow-up
subjects #	176	
women #(%)	120 (68%)	
disease subtype # (%)		
clinically isolated syndrome	9 (5%)	5 (3%)
relapsing remitting	102 (58%)	104 (59%)
secondary progressive	40 (23%)	42 (24%)
primary progressive	25 (14%)	25 (14%)
age, mean SD (range)	44+/-12 (20-69)	45+/-12 (20-70)
disease duration, mean +/-SD (range)	10+/-10(0-42)	11+/-10(0-44)
receiving disease-modifying treatment, # (%)	115 (67%)	125(71%)
received steroids for a relapse within 30 days of the scan, # (%)	21 (12%)	12 (7%)
expanded disability status scale, median (range)	3.5 (0-8.5)	3(0-8)

Table 4: *Cohort characteristics of the subjects involved in the MS study.*

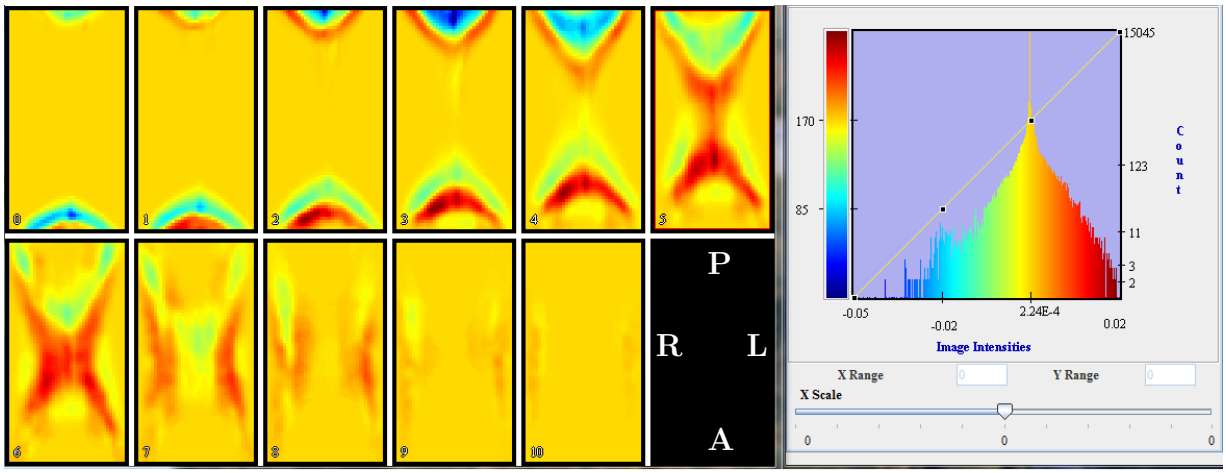

Figure 17: *Eleven slices of $\hat{\phi}_3^{X,0}$. A histogram of the voxel intensities is on the right. The pictures are obtained using MIPAV (2011).*

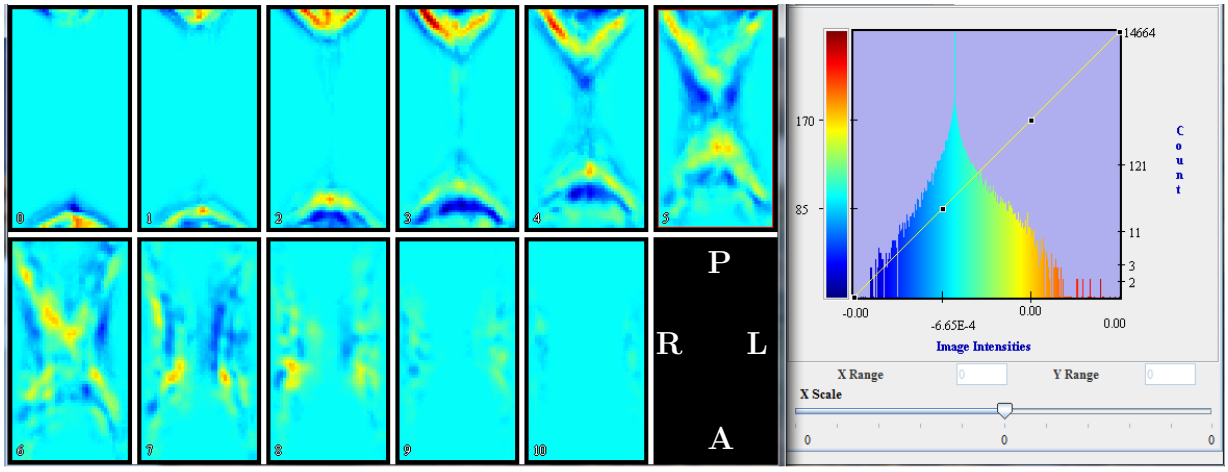

Figure 18: *Eleven slices of $\hat{\phi}_3^{X,1}$. A histogram of the voxel intensities is on the right. The pictures are obtained using MIPAV (2011).*

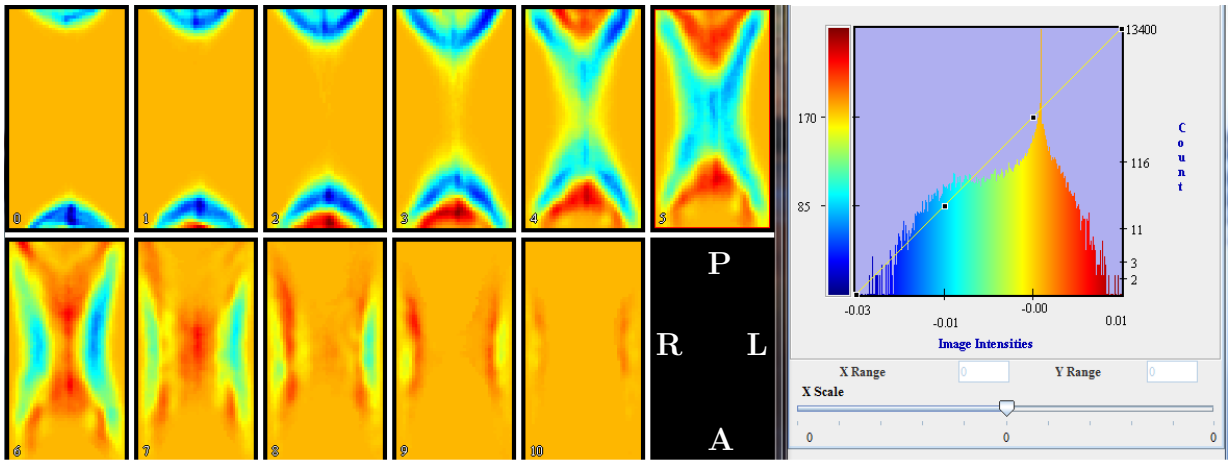

Figure 19: *Eleven slices of $\hat{\phi}_1^{W}$. A histogram of the voxel intensities is on the right. The pictures are obtained using MIPAV (2011).*

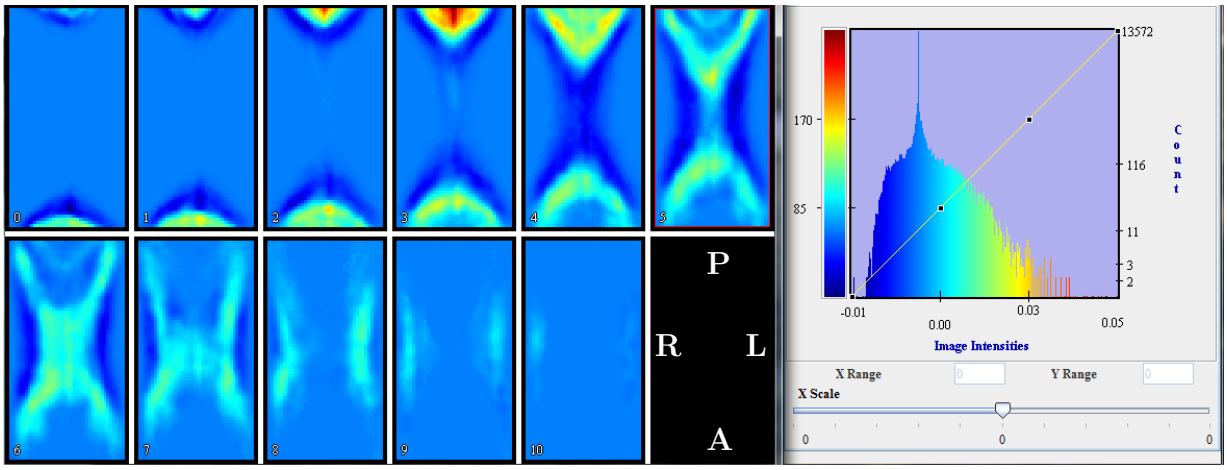

Figure 20: *Eleven slices of $\hat{\phi}_2^W$. A histogram of the voxel intensities is on the right. The pictures are obtained using MIPAV (2011).*

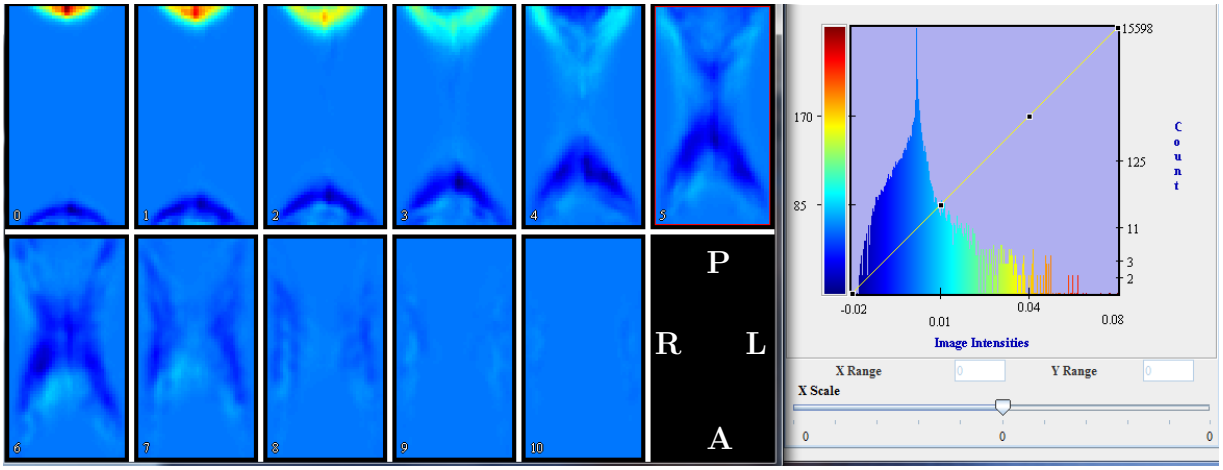

Figure 21: *Eleven slices of $\hat{\phi}_3^W$. A histogram of the voxel intensities is on the right. The pictures are obtained using MIPAV (2011).*