Published in final edited form as:
Ann Appl Stat. 2014 ; 8(4): 2175–2202.

Longitudinal High-Dimensional Principal Components Analysis with Application to Diffusion Tensor Imaging of Multiple Sclerosis

Vadim Zipunnikov [Assistant Professor],
Department of Biostatistics, Johns Hopkins University, Baltimore, MD, 21205

Sonja Greven [Professor],
Department of Statistics, Ludwig-Maximilians-Universität and Miinchen, 80539 Munich, Germany

Haochang Shou, **Brian Caffo**,
Department of Biostatistics, Johns Hopkins University, Baltimore, MD, 21205

Daniel S. Reich [Chief], and
Translational Neurology Unit, Neuroimmunology Branch, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD 20892, USA

Ciprian Crainiceanu [Full Professors]
Department of Biostatistics, Johns Hopkins University, Baltimore, MD, 21205

Vadim Zipunnikov: vzipunni@jhsph.edu; Sonja Greven: sonja.greven@stat.uni-muenchen.de; Brian Caffo: bcaffo@jhsph.edu; Daniel S. Reich: daniel.reich@nih.gov; Ciprian Crainiceanu: ccrainic@jhsph.edu

Abstract

We develop a flexible framework for modeling high-dimensional imaging data observed longitudinally. The approach decomposes the observed variability of repeatedly measured high-dimensional observations into three additive components: a subject-specific imaging random intercept that quantifies the cross-sectional variability, a subject-specific imaging slope that quantifies the dynamic irreversible deformation over multiple realizations, and a subject-visit specific imaging deviation that quantifies exchangeable effects between visits. 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

principal components; linear mixed model; diffusion tensor imaging; brain imaging data; multiple sclerosis

Correspondence to: Vadim Zipunnikov, vzipunni@jhsph.edu.

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. In this paper we propose to generalize the random intercept random slope mixed effects model to the case when instead of a scalar one measures a high-dimensional object, such as a brain image. The proposed methods can be applied to longitudinal studies that include high-dimensional imaging observations without missing data that can be unfolded into a long vector.

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). MS also affects the neurons themselves and is associated with accelerated brain atrophy.

Our data are derived from a natural history study of 176 MS cases selected from a population with 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 (registered) using a 12 degrees of freedom transformation which accounts for rotation, translation, scaling, and shearing, but not for nonlinear deformation. In this study we focus on fractional anisotropy (FA), a useful voxel-level summary of diffusion tensor imaging (DTI), a type of structural Magnetic Resonance Imaging (MRI). FA is viewed as a measure of tissue integrity and is thought to be sensitive both to axon fiber density and myelination in white matter. It is measured on a scale between zero (isotropic diffusion characteristic of fluid-filled cavities) and one (anisotropic diffusion, characteristic of highly ordered white matter fiber bundles) (Mori, 2007).

The goal of the study was to quantify the location and size of longitudinal variability of FA along the corpus callosum. The primary region of interest (ROI) 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 that contains corpus callosum together with its relative location 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 imaging random intercept that quantifies the cross-sectional variability; 2) a subject-specific imaging slope that quantifies the dynamic

irreversible deformation over multiple visits; and 3) a subject-visit specific imaging 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 and adaptive kernels have been reported in the literature (Mohamed and Davatzikos, 2004; Morris and Carroll, 2006; Guo, 2002; Morris et al., 2011; Zhu et al., 2011; Rodriguez et al., 2009; Bigelow and Dunson, 2009; Reiss et al., 2005; Reiss and Ogden, 2008, 2010; Li et al., 2011; Hua et al., 2012; Yuan et al., 2014). 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 challenges that are currently unaddressed.

Here we address the problem of exploring and analyzing populations of high dimensional images at multiple visits using high dimensional longitudinal functional principal components analysis (HD-LFPCA). The method decomposes the longitudinal imaging data into subject specific, longitudinal subject specific, and subject-visit specific components. The dimension reduction for all components is done using principal components of the corresponding covariance operators. Note that we are interested in imaging applications and do not perform smoothing. However, in Section 3.4, we discuss how the proposed approach can be paired with smoothing and applied to high-dimensional functional data. The estimation and inferential methods are fast and can be performed on standard personal computers to analyze hundreds or thousands of high-dimensional images at multiple visits. This was achieved by the following 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; and 4) linking the high-dimensional space to a low-dimensional intrinsic space, which allows Karhunen-Loève (KL) decompositions of covariance operators that cannot even be stored in the computer memory. Thus, the proposed methods are computationally linear in the dimension of images.

The rest of the manuscript is organized 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 we heavily modify it to make it practical for applications to imaging high dimensional data. We also present the major reasons why the original methods can not be applied to high-dimensional data.

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 directly 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

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

where υ denotes a voxel, $\eta(\upsilon)$ is a fixed main effect, $X_{i,0}(\upsilon)$ is the random imaging intercept for subject i, $X_{i,1}(\upsilon)$ is the random imaging slope for subject i, T_{ij} is the time of visit j for subject i, $W_{ij}(\upsilon)$ is the random subject/visit-specific imaging deviation. For simplicity, the main effect $\eta(\cdot)$ does not depend on i and j. As discussed in (Greven et al., 2010), model (1) and the more general model (8) in Section 3.2 are similar to functional models with uncorrelated (Guo, 2002) and correlated (Morris and Carroll, 2006) random functional effects. Instead of using smoothing splines and wavelets as in (Guo, 2002; Morris and Carroll, 2006), our approach models the covariance structures using functional principal component analysis; we have found this approach to lead to the major computational advantages, as further discussed in Section 3.

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 $\eta(\upsilon)$ is a fixed surface/image and the latent (unobserved) bivariate process $X_i(\upsilon) = (X'_{i,0}(\upsilon), X'_{i,1}(\upsilon))'$ and process $W_{ij}(\upsilon)$ are square-integrable stochastic processes. We also assume that $X_i(\upsilon)$ and $W_{ij}(\upsilon)$ are uncorrelated. We denote by $\mathbf{K}^X(\upsilon_1, \upsilon_2)$ and $\mathbf{K}^W(\upsilon_1, \upsilon_2)$ their covariance operators, respectively. Assuming that $\mathbf{K}^X(\upsilon_1, \upsilon_2)$ and $\mathbf{K}^W(\upsilon_1, \upsilon_2)$ are continuous, we can use the standard Karhunen-Loève expansions of the random processes (Karhunen, 1947; Loeve, 1978) and represent $X_i(\upsilon) = \sum_{k=1}^{\infty} \xi_{ik}\phi_k^X(\upsilon)$ with $\phi_k^X(\upsilon) = (\phi_k^{X,0}(\upsilon), \phi_k^{X,1}(\upsilon))$ and $W_{ij}(\upsilon) = \sum_{l=1}^{\infty} \zeta_{ijl}\phi_l^W(\upsilon)$ 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 for $\mathbf{K}^X$ and $\mathbf{K}^W$. The LFPCA model becomes the mixed effects model

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

where $\mathbf{Z}_{ij} = (1, T_{ij})'$ and " ~ (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 $k_1 \leq k_2$. We do not require normality of the scores in the model. The only assumption is the existence of second order moments of the distribution of scores. In addition, the assumption that $X_i(\upsilon)$ and $W_{ij}(\upsilon)$ 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 3.2.

In practice, model 2 is projected onto the first N_X and N_W components of $\mathbf{K}^X$ and $\mathbf{K}^W$, respectively. Assuming that N_X and N_W are known, the model becomes

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

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

Our estimation is based on the methods of moments (MoM) for pairwise quadratics $E(\boldsymbol{Y}_{ij_1}\boldsymbol{Y}'_{kj_2})$. The computationally intensive part of fitting (3) is estimating the following massively multivariate model

$$\boldsymbol{Y}_{ij}=\eta+\sum_{k=1}^{N_X}\xi_{ik}\phi_k^{X,0}+T_{ij}\sum_{k=1}^{N_X}\xi_{ik}\phi_k^{X,1}+\sum_{l=1}^{N_W}\zeta_{ijl}\phi_l^W=\eta+\boldsymbol{\Phi}^{X,0}\boldsymbol{\xi}_i+T_{ij}\boldsymbol{\Phi}^{X,1}\boldsymbol{\xi}_i+\boldsymbol{\Phi}^W\boldsymbol{\zeta}_{ij}, \quad (4)$$

where $\eta = (\eta(\upsilon_1), \ldots, \eta(\upsilon_p))$, $\mathbf{Y}_{ij} = \{Y_{ij}(\upsilon_1), \ldots, Y_{ij}(\upsilon_p)\}$ are $p \times 1$ dimensional vectors, $\phi_k^{X,0}, \phi_k^{X,1}$ and ϕ_l^W are correspondingly vectorized eigenvectors, $\boldsymbol{\Phi}^{X,0}=[\phi_1^{X,0},\ldots,\phi_{N_X}^{X,0}]$ and $\boldsymbol{\Phi}^{X,1}=[\phi_1^{X,1},\ldots,\phi_{N_X}^{X,1}]$ are $p \times N_X$ dimensional matrices, $\boldsymbol{\Phi}^W=[\phi_1^W,\ldots,\phi_{N_W}^W]$ 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 $E(\boldsymbol{\xi}_i\boldsymbol{\xi}_i')=\boldsymbol{\Lambda}^X=\mathrm{diag}(\lambda_1^X,\ldots,\lambda_{N_X}^X)$ and $E(\boldsymbol{\zeta}_{ij}\boldsymbol{\zeta}_{ij}')=\boldsymbol{\Lambda}^W=\mathrm{diag}(\lambda_1^W,\ldots,\lambda_{N_W}^W)$, respectively.

To obtain the eigenvectors and eigenvalues in model (4), 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.

Lemma 1: *The MoM estimators of the covariance operators and the mean in (4) are unbiased and given by*

$$\begin{aligned}&\hat{\mathbf{K}}_X^{00}=\sum_{i,j_1,j_2}\tilde{\mathbf{Y}}_{ij_1}\tilde{\mathbf{Y}}_{ij_2}'h^1_{ij_1j_2},\quad \hat{\mathbf{K}}_X^{01}=\sum_{i,j_1,j_2}\tilde{\mathbf{Y}}_{ij_1}\tilde{\mathbf{Y}}_{ij_2}'h^2_{ij_1j_2},\quad \hat{\mathbf{K}}_X^{10}=\sum_{i,j_1,j_2}\tilde{\mathbf{Y}}_{ij_1}\tilde{\mathbf{Y}}_{ij_2}'h^3_{ij_1j_2},\\ &\hat{\mathbf{K}}_X^{11}=\sum_{i,j_1,j_2}\tilde{\mathbf{Y}}_{ij_1}\tilde{\mathbf{Y}}_{ij_2}'h^4_{ij_1j_2},\quad \hat{\mathbf{K}}^W=\sum_{i,j_1,j_2}\tilde{\mathbf{Y}}_{ij_1}\tilde{\mathbf{Y}}_{ij_2}'h^5_{ij_1j_2},\quad \hat{\eta}=\frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{J_i}\boldsymbol{Y}_{ij},\end{aligned} \tag{5}$$

where $\tilde{\mathbf{Y}}_{ij} = \mathbf{Y}_{ij} - \hat{\boldsymbol{\eta}}$, *the* $2p \times 2p$ *matrix* $\mathbf{K}^X=[\mathbf{K}_X^{00}\vdots\mathbf{K}_X^{01};\mathbf{K}_X^{10}\vdots\mathbf{K}_X^{11}]$, *with* $\mathbf{K}_X^{ks}=E\{\boldsymbol{\Phi}^{X,k}\boldsymbol{\xi}_i(\boldsymbol{\Phi}^{X,s}\boldsymbol{\xi}_i)'\}$ *for* $k, s \in \{0, 1\}$, *the weights* $h^l_{ij_1j_2}$ *are elements of the* l^{th} *column of the matrix* $\mathbf{H}_{m\times 5} = \mathbf{F}'(\mathbf{F}\mathbf{F}')^{-1}$ *the matrix* $\mathbf{F}_{5\times m}$ *has columns equal to* $\mathbf{f}_{ij_1j_2} = (1, T_{ij_2}, T_{ij_1}, T_{ij_1}, T_{ij_2}, \delta_{j_1j_2})'$, *and* $m=\sum_{i=1}^{I}J_i^2$.

The proof of the lemma is given in Appendix. The MoM estimators (5) define the symmetric matrices $\mathbf{K}^{\hat{X}}$ and $\mathbf{K}^{\hat{W}}$. Identifiability of model (4) requires that some subjects have more than two visits, that is $J_i \geq 3$. Note that if one is only interested in estimating covariances, η can be eliminated as a nuisance parameter by using MoMs for quadratics of differences $E\{\mathbf{Y}_{ij_1} - \mathbf{Y}_{kj_2})(\mathbf{Y}_{ij_1} - \mathbf{Y}_{kj_2})'$ as in Shou et al. (2013).

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. Even if it were possible to calculate and store these covariances, obtaining the spectral decompositions would 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 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 these 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 on providing a new methodological approach with the ultimate goal of solving the intractable computational problems discussed in the previous section.

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 without either calculating or storing the 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 6 we discuss ways to extend our approach to situations when this assumption is violated.

Write $\tilde{\mathbf{Y}} = (\tilde{\mathbf{Y}}_1, \ldots, \tilde{\mathbf{Y}}_I)$, where $\tilde{\mathbf{Y}}_i = (\tilde{\mathbf{Y}}_{i1},\ldots, \tilde{\mathbf{Y}}_{iJ_i})$ 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 $\tilde{\mathbf{Y}}_i$ contains all the data for subject i with each column corresponding to a particular visit. The matrix $\tilde{\mathbf{Y}}$ is the $p \times n$ matrix obtained by column-binding the centered subject-specific data matrices $\tilde{\mathbf{Y}}_i$. Thus if $\tilde{\mathbf{Y}}_i = (\tilde{\mathbf{Y}}_{i1}, \ldots, \tilde{\mathbf{Y}}_{iJ_i})$ then $\tilde{\mathbf{Y}} = (\tilde{\mathbf{Y}}1, \ldots, \tilde{\mathbf{Y}}_I)$. Our approach starts with constructing the SVD of the matrix $\tilde{\mathbf{Y}}$

$$\tilde{\mathbf{Y}}=\mathbf{V}\mathbf{S}^{1/2}\mathbf{U}'. \quad (6)$$

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 $\tilde{\mathbf{Y}}$ requires only a number of operations linear in the number of parameters p. Indeed, consider the $n \times n$ symmetric matrix $\tilde{\mathbf{Y}}'\tilde{\mathbf{Y}}$with its spectral decomposition $\tilde{\mathbf{Y}}'\tilde{\mathbf{Y}} = \mathbf{U}\mathbf{S}\mathbf{U}'$. Note that for high-dimensional p the matrix $\tilde{\mathbf{Y}}$ cannot be loaded into the memory. The solution is to partition it into L slices as $\tilde{\mathbf{Y}}' = [(\tilde{\mathbf{Y}}^1)'|(\tilde{\mathbf{Y}}^2)'| \ldots |(\tilde{\mathbf{Y}}^L)']$, where the size of the lth slice, $\tilde{\mathbf{Y}}^l$, is $(p/L) \times n$ and can be adapted to the available computer memory and optimized to reduce implementation time. The matrix $\tilde{\mathbf{Y}}'\tilde{\mathbf{Y}}$ is then calculated as $\sum_{l=1}^{L}(\tilde{\mathbf{Y}}^l)'\tilde{\mathbf{Y}}^l$ by streaming the individual blocks. This step calculates singular value decomposition of the $p \times n$ matrix $\tilde{\mathbf{Y}}$. Note that for any permutation of components υ, model (3) will be valid and the covariance structure imposed by the model can be recovered by doing the inverse permutation. If smoothing of the covariance matrix is desirable, then this step can be efficiently combined with Fast Covariance Estimation (FACE, Xiao et al. (2013)), a computationally efficient smoother of (low-rank) high-dimensional covariance matrices with p up to 100,000.

From the SVD (6) the $p \times n$ matrix $\mathbf{V}$ can be obtained as $\mathbf{V} = \tilde{\mathbf{Y}}\mathbf{U}\mathbf{S}^{-1/2}$. The actual calculations can be performed on the slices of the partitioned matrix $\tilde{\mathbf{Y}}$ as $\mathbf{V}^l = \tilde{\mathbf{Y}}^l\mathbf{U}\mathbf{S}^{-1/2}$, $l = 1, \ldots,L$. 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 (6) can be constructed with sequential access to the data $\tilde{\mathbf{Y}}$ with p-linear effort.

After obtaining the SVD of $\tilde{\mathbf{Y}}$, each image $\tilde{\mathbf{Y}}_{ij}$ can be represented as $\tilde{\mathbf{Y}}_{ij} = \mathbf{V}\mathbf{S}^{1/2}\mathbf{U}_{ij}$, where $\mathbf{U}_{ij}$ is a corresponding column of matrix $\mathbf{U}'$. Therefore, the vectors $\tilde{\mathbf{Y}}_{ij}$ differ only through the vector factors $\mathbf{U}_{ij}$ of dimension $n \times$ l. Comparing this SVD representation of $\tilde{\mathbf{Y}}_{ij}$ with the right-hand side of (4), it follows that *cross-sectional and longitudinal variability controlled by the principal scores* ξ_i, ζ_{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 the approach feasible. Below, we provide more intuition behind our approach. The formal argument is presented in Lemma 2.

First, we substitute the left-hand side of (4) with its SVD representation of $\tilde{\mathbf{Y}}_{ij}$ to get $\mathbf{V}\mathbf{S}^{1/2}\mathbf{U}_{ij} = \mathbf{\Phi}^{X,0}\xi_i + T_{ij}\mathbf{\Phi}^{X,1}\xi_i + \mathbf{\Phi}^{W}\zeta_{ij}$. Now we can multiply by $\mathbf{V}'$ both sides of the equation to get $\mathbf{S}^{1/2}\mathbf{U}_{ij} = \mathbf{V}'\mathbf{\Phi}^{X,0}\xi_i + T_{ij}\mathbf{V}'\mathbf{\Phi}^{X,1}\xi_i + \mathbf{V}'\mathbf{\Phi}^{W}\zeta_{ij}$. If we denote $\mathbf{A}^{X,0} = \mathbf{V}'\mathbf{\Phi}^{X,0}$ of size $n \times N_X$, $\mathbf{A}^{X,1} = \mathbf{V}'\mathbf{\Phi}^{X,1}$ of size $n \times N_X$, and $\mathbf{A}^{W} = \mathbf{V}'\mathbf{\Phi}^{U}$ of size $n \times N_W$, we obtain

$$\mathbf{S}^{1/2}\mathbf{U}_{ij} = \mathbf{A}^{X,0}\boldsymbol{\xi}_i + T_{ij}\mathbf{A}^{X,1}\boldsymbol{\xi}_i + \mathbf{A}^{W}\boldsymbol{\zeta}_{ij}. \quad (7)$$

Conditionally on the observed data, $\tilde{\mathbf{Y}}$, models (4) and (7) are equivalent. Indeed, model (4) is a *linear* model for the n vectors $\tilde{\mathbf{Y}}_{ij}$'s. These vectors span an (at most) n-dimensional linear subspace. Hence, the columns of the matrix $\mathbf{V}$, the right singular vectors of $\tilde{\mathbf{Y}}$, could be thought of as an orthonormal basis, while $\mathbf{S}^{1/2}\mathbf{U}_{ij}$ are the coordinates of $\tilde{\mathbf{Y}}_{ij}$ in this basis. Multiplication by $\mathbf{V}'$ can be seen as a linear mapping from model (4) for the high-dimensional observed data $\tilde{\mathbf{Y}}_{ij}'s$ to model (7) for the low dimensional data $\mathbf{S}^{1/2}\mathbf{U}_{ij}$. Additionally, even though $\mathbf{V}\mathbf{V}' \neq \mathbf{I}_p$, the projection defined by $\mathbf{V}$ is lossless in the sense that model (4) can be recovered from model (7) using the identity $\mathbf{V}\mathbf{V}'\tilde{\mathbf{Y}}_{ij} = \tilde{\mathbf{Y}}_{ij}$. Hence, model (7) has an "intrinsic" dimensionality induced by the study sample size, n. We can estimate the low-dimensional model (7) using the LFPCA methods described 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 (7) is an essential step for getting the high-dimensional principal components in p-linear time.

Lemma 2: *The eigenvectors of the estimated covariance operators (5) can be calculated as* $\mathbf{\Phi}^{\hat{X},0} = \mathbf{V}\hat{\mathbf{A}}^{X,0}$, $\mathbf{\Phi}^{\hat{X},1} = \mathbf{V}\hat{\mathbf{A}}^{X,1}$, $\mathbf{\Phi}^{\hat{W}} = \mathbf{V}\hat{\mathbf{A}}^{W}$, *where the matrices* $\hat{\mathbf{A}}^{X,0}$, $\hat{\mathbf{A}}^{X,1}$, $\hat{\mathbf{A}}^{W}$ *are obtained from fitting model (7). The estimated matrices of eigenvalues* $\mathbf{\Lambda}^{\hat{X}}$ and $\mathbf{\Lambda}^{\hat{W}}$ *are the same for both model (4) and model (7).*

The proof of the lemma is given in Appendix. This result 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 (7).

3.2 The general functional mixed model

A natural way to generalize model (3) is to consider the following model

$$\boldsymbol{Y}_{ij}=\eta+Z_{ij,0}\sum_{k=1}^{N_X}\xi_{ik}\phi_k^{X,0}+Z_{ij,1}\sum_{k=1}^{N_X}\xi_{ik}\phi_k^{X,1}+\ldots+Z_{ij,q}\sum_{k=1}^{N_X}\xi_{ik}\phi_k^{X,q}+\sum_{l=1}^{N_W}\zeta_{ijl}\phi_l^{W}, \quad (8)$$

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 other covariates of interest.

The fitting approach is essentially the same as the one described for the LFPCA model in Section 3.1. As before, the right singular vectors $\mathbf{U}_{ij}$ contain the longitudinal information about $\boldsymbol{\xi}_i$, $\boldsymbol{\zeta}_i$, and covariates $\mathbf{Z}_{ij}$. The following two results are direct generalizations of Lemmas 1 and 2.

Lemma 3: *The MoM estimators of the covariance operators and the mean in (8) are unbiased and given by*

$$\hat{\mathbf{K}}_X^{ks}=\sum_{i,j_1,j_2}\tilde{\mathbf{Y}}_{ij_1}\tilde{\mathbf{Y}}_{ij_2}^{'}h_{ij_1j_2}^{1+s+k(q+1)},\hat{\mathbf{K}}^W=\sum_{i,j_1,j_2}\tilde{\mathbf{Y}}_{ij_1}\tilde{\mathbf{Y}}_{ij_2}^{'}h_{ij_1j_2}^{(q+1)^2+1},\hat{\eta}=\frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{J_i}\boldsymbol{Y}_{ij},$$

where $\tilde{\mathbf{Y}}_{ij} = \mathbf{Y}_{ij} - \hat{\boldsymbol{\eta}}$, *the* $(q + 1)p \times (q + 1)p$ *block-matrix* $\mathbf{K}^X$ *is composed of* $p \times p$ *matrices* $\mathbf{K}_X^{ks}=E\{\boldsymbol{\Phi}^{X,k}\boldsymbol{\xi}_i(\boldsymbol{\Phi}^{X,s}\boldsymbol{\xi}_i)'\}$ *for* $k, s \in \{0, 1, \ldots, q\}$, *the weights* $h_{ij_1j_2}^{l}$ *are elements of the* l^{th} *column of matrix* $\mathbf{H}_{m\times((q+1)^2+1)} = \mathbf{F}'(\mathbf{F}\mathbf{F}')^{-1}$, *the matrix* $\mathbf{F}_{((q+1)^2+1)\times m}$ *has columns equal to* $\mathbf{f}_{ij_1j_2} = (vec(\mathbf{Z}_{ij_1} \otimes \mathbf{Z}_{ij_2}), \delta_{j_1j_2})'$, and $m=\sum_{i=1}^{I}J_i^2$.

Lemma 4: *The eigenvectors of the estimated covariance operators for (8) can be calculated as* $\boldsymbol{\Phi}^{\hat{X},k} = \mathbf{V}\hat{\mathbf{A}}^{X,k}$, $k = 0, 1, \ldots, q$, $\boldsymbol{\Phi}^{\hat{W}} = \mathbf{V}\hat{\mathbf{A}}^{W}$, *where the matrices* $\hat{\mathbf{A}}^{X,k}$, $k = 0, 1, \ldots, q$ *and* $\hat{\mathbf{A}}^{W}$ *are obtained from fitting the intrinsic model*

$$\mathbf{S}^{1/2}\mathbf{U}_{ij}=Z_{ij,0}\sum_{k=1}^{N_X}\xi_{ik}\mathbf{A}_k^{X,0}+Z_{ij,1}\sum_{k=1}^{N_X}\xi_{ik}\mathbf{A}_k^{X,1}+\ldots+Z_{ij,q}\sum_{k=1}^{N_X}\xi_{ik}\mathbf{A}_k^{X,q}+\sum_{l=1}^{N_W}\zeta_{ijl}\mathbf{A}_l^{W}. \quad (9)$$

The estimated matrices of eigenvalues $\boldsymbol{\Lambda}^{\hat{X}}$ *and* $\boldsymbol{\Lambda}^{\hat{W}}$ *are the same for both model (8) and model (9).*

3.3 Estimation of principal scores

The principal scores are the coordinates of $\tilde{\mathbf{Y}}_{ij}$ in the basis defined by the LFPCA model (8). In this section, we propose an approach to calculating BLUP of the scores that is computationally feasible for samples of high-resolution images.

First, we introduce some notation. In Section 3.1, we showed that the SVD of the matrix $\tilde{\mathbf{Y}}$ can be written as $\tilde{\mathbf{Y}}_i=\mathbf{V}\mathbf{S}^{1/2}\mathbf{U}_i^{'}$, where the $n \times J_i$ matrix $\mathbf{U}_i^{'}$ corresponds to the subject i. Model (8) can be rewritten as

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

where,

$\mathbf{B}_i=[\mathbf{B}_i^X:\mathbf{B}_i^W], \mathbf{B}_i^X=\boldsymbol{Z}_{i,0}\otimes\boldsymbol{\Phi}^{X,0}+\boldsymbol{Z}_{i,1}\otimes\boldsymbol{\Phi}^{X,1}+\cdots+\boldsymbol{Z}_{i,q}\otimes\boldsymbol{\Phi}^{X,q}, \mathbf{B}_i^W=\mathbf{I}_{J_i}\otimes\boldsymbol{\Phi}^W, \boldsymbol{Z}_{i,k}=(Z_{i1,k},\ldots,Z_{iJ_i,k})', \boldsymbol{\omega}_i=(\boldsymbol{\xi}_i'$

the subject level principal scores $\boldsymbol{\zeta}_i=(\boldsymbol{\zeta}_{i1}',\ldots,\boldsymbol{\zeta}_{iJ_i}')'$, $\otimes$ is the Kronecker product of matrices, and operation *vec*(·) stacks the columns of a matrix on top of each other. The following lemma contains the main result of this section; it shows how the estimated BLUPs can be calculated for the LFPCA model.

Lemma 5: *Under the general LFPCA model (8), the estimated best linear unbiased predictor (EBL UP) of* $\boldsymbol{\xi}_i$ *and* $\boldsymbol{\zeta}_i$ *is given by*

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

where all matrix factors on the right side can be written in terms of the low-dimensional right singular vectors.

The proof of the lemma is 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*. Section 6 in Appendix briefly describes how the framework can be adapted to settings with tens or hundreds of thousands images.

3.4 HF-LFPCA model with white noise

The original LFPCA model in Greven et al. (2010) was developed for functional observations and contained an additional white noise term. In this section, we show how the HD-LFPCA framework can be extended to accommodate such a term and how the extended model can be estimated.

We now seek to fit the following model

$$\boldsymbol{Y}_{ij}=\eta+Z_{ij,0}\sum_{k=1}^{N_X}\xi_{ik}\phi_k^{X,0}+Z_{ij,1}\sum_{k=1}^{N_X}\xi_{ik}\phi_k^{X,1}+\ldots+Z_{ij,q}\sum_{k=1}^{N_X}\xi_{ik}\phi_k^{X,q}+\sum_{l=1}^{N_W}\zeta_{ijl}\phi_l^W+\varepsilon_{ij}, \quad (12)$$

where ε_{ij} is a *p*-dimensional white noise variable, that is $E(\varepsilon_{ij}) = 0_p$ for any *i*, *j* and $E(\varepsilon_{i_1j_1}\varepsilon_{i_2j_2}) = \sigma^2\delta_{i_1i_2}\delta_{j_1j_2}\mathbf{I}_p$. The white noise process $\varepsilon_{ij}(\upsilon)$ is assumed to be uncorrelated with processes $X_i(\upsilon)$ and $W_{ij}(\upsilon)$.

Lemma 3 applied to (12) shows that $\hat{K}_{\sigma^2}^W=\sum_{i,j_1,j_2}\tilde{\mathbf{Y}}_{ij_1}\tilde{\mathbf{Y}}_{ij_2}' h_{ij_1j_2}^{(q+1)^2+1}$ is an unbiased estimator of $K^W + \sigma^2\mathbf{I}_p$. To estimate σ^2 in a functional case, we can follow the method in Greven et al. (2010): i) drop the diagonal elements of $\hat{K}_{\sigma^2}^W$ and use a bivariate smoother to get $\tilde{K}_{\sigma^2}^W$, ii) calculate an estimator $\hat{\sigma}^2=\max\{(tr(\hat{K}_{\sigma^2}^W)-tr(\tilde{K}_{\sigma^2}^W)/p,0\}$. To make this approach feasible in very high-dimensional settings ($p \sim 100,000$), we can use the fast covariance estimation (FACE) developed in Xiao et al. (2013), a bivariate smoother that scales up linearly with respect to *p* and preserves the low dimensionality of the estimated covariance operator. Thus HD-LFPCA remains feasible after smoothing by FACE.

When the observations $\mathbf{Y}_{ij}$'s are non-functional, the off-diagonal smoothing approach cannot be used. In this case, if one assumes that model (12) is low-rank then σ^2 can be estimated as $(tr(\hat{K}^W_{\sigma^2}) - \sum_{k=1}^{N_W} \hat{\lambda}^W_k)/(p - N_W)$. Bayesian model selection approaches that estimate both the rank of PCA models and variance σ^2 are discussed in Everson and Roberts (2000) and Minka (2000).

4 Simulations

In this section, three simulation studies are used to explore the properties of our proposed methods. In the first study, we replicate several simulation scenarios in Greven et al. (2010) for functional curves, but we focus on using a number of parameters up to two orders of magnitude larger than the ones in the original scenarios. This increase in dimensionality could not be handled by the original LFPCA approach. In the second study, we explore how methods recover 3D spatial bases when the approach of Greven et al. (2010) cannot be implemented. In the third study, we replicate the unbalanced design in and use time variable T_{ij} from our DTI application and generate data using principal components estimated in Section 5. For each scenario, we simulated 100 data sets. All three 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.

First scenario (ID, functional curves)

We follow Greven et al. (2010) and generate data as follows

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

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 the scores ξ_{ik} are simulated 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, a similar notation holds for ζ_{ijl}. The scores ξ_{ik}'s and ζ_{ijl}'s are mutually independent. We set $I = 100$, $J_i = 4$, $i = 1, \ldots, I$, and 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 orthogonal but not mutually orthogonal bases were

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

which are measured on a regular grid of p equidistant points in the interval [0, 1]. To explore scalability, we consider several grids with increasing number of sampling points, p, equal to 750, 3000, 12000, 24000, 48000, and 96000. Note that a brute-force extension of the

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 $1 \le j \le 3$. The times $T_{i1}, \ldots, T_{i4}$ are normalized to have sample mean zero and variance one. Although no measurement noise is assumed in model (3), we simulate data that also contains white noise, $\varepsilon_{ij}(\upsilon)$. The purpose of this is two fold. First, it is of interest to explore how the presence of white noise affects the performance of methods which do not model it explicitly. Second, the choice of the eigenfunctions in the original simulation scenario of Greven et al. (2010) makes the estimation problem ill-posed if data does not contain white noise. The white noise $\varepsilon_{ij}(\upsilon)$ are assumed to be i.i.d. $N(0, \sigma^2)$ for each i, j, υ and independent of all other latent processes. To evaluate different signal-to-noise ratios we consider values of σ^2 equal to 0.0001, 0.0005, 0.001, 0.005, 0.01. Note that we normalized each of the data generating eigenvectors to have norm one. Thus, the signal-to-noise ratio, $(\sum_{k=1}^{4}\lambda_k^X + \sum_{k=1}^{4}\lambda_k^W)/(p\sigma^2)$, ranges from 50 (for $p = 750$ and $\sigma^2 = 0.0001$) to 0.004 (for $p = 96,000$ and $\sigma^2 = 0.01$).

Table 1 and Tables 4 and 5 in the web-appendix report the average L_2 distances between the estimated and true eigenvectors for $X_{i,0}(\upsilon)$, $X_{i,1}(\upsilon)$, and $W_{ij}(\upsilon)$, respectively. The averages are calculated based on 100 simulated data sets for each (p, σ^2) combination. Standard deviations are shown in brackets. Three trends are obvious: i) eigenvectors with larger eigenvalues are estimated with higher accuracy, ii) larger white noise corresponds to a decreasing accuracy, iii) for identical levels of white noise, accuracy goes down when the dimension p goes up. Similar trends are observed for average distances between estimated and true eigenvalues reported in Tables 6 and 7. These trends follow from the fact that for any fixed σ^2, the signal-to-noise ratio decreases with increasing p and the performance of the approach quickly deteriorates once the signal-to-noise ratio becomes smaller than one.

Figure 8 (web appendix) displays the true and estimated eigenfunctions for the case when $p = 12,000$ and $\sigma^2 = 0.01^2$ and shows the complete agreement with Figure 2 in Greven et al. (2010). The boxplots of the estimated eigenvalues are displayed in Figure 4. In Figure 5, panels one and three report the boxplots of and panels two and four display the medians and quantiles of the distribution 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. This indicates that the estimation procedures provides unbiased estimates.

Second scenario (3D)

Data sets in this study replicate the 3D ROI blocks from the DTI MS data set. We simulated 100 data sets from the model

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

where $\mathscr{V} = [1, 38] \times [1, 72] \times [1, 11]$. Eigenimages $(\phi_k^{X,0}, \phi_k^{X,1})$, and ϕ_l^W are displayed in Figure 3. 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 are set to 0. There are four blue and red sub-blocks corresponding to $\phi_k^{X,0}$ and $\phi_k^{X,1}$, respectively. The eigenfunctions closest to the anterior side of the brain (labeled A in Figure 3) are $\phi_1^{X,0}$ and $\phi_1^{X,1}$, which have the strongest signal proportional to the largest eigenvalue (variance), λ_1^X. The eigenvectors that are progressively closer to the posterior part of the brain (labeled P) correspond to smaller eigenvalues represented as lighter shades of blue and red, respectively. The sub-blocks closest to the P have the smallest signal, which is proportional to λ_4^X. The eigenimages ϕ_k^W shown in green are ordered 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 $X_{i,1}(\upsilon)$. 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 as in the first simulation scenario. 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 entire simulation study took 20 minutes or approximately 12 seconds per data set. Figures 13, 14, and 15 (web-appendix) display the medians of the estimated eigenimages and the voxelwise 5th and 95th percentile images, respectively. All axial slices, or z slices in a x-y-z coordinate system, are the same. 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, $\hat{\varphi} = (\hat{\varphi}_1, \ldots, \hat{\varphi}_p)$, was normalized as $\hat{\varphi} \to (\hat{\varphi} - \min_s \hat{\varphi}_s)/(\max_s \hat{\varphi}_s - \min_s \hat{\varphi}_s)$. Figure 13 displays the voxel-wise medians of the estimator, indicating that the method recovers the spatial configuration of both bases. The 5-percentile and 95-percentile images are displayed in Figures 14 and 15, respectively. Overall, the original pattern is recovered with some small distortions most likely due to the correlation between bases (please note the light gray patches).

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 11. The eigenvalues are estimated consistently. However, in 6 out of 100 cases (extreme values shown in red), the estimation procedure did not distinguish well between ϕ_3^W and ϕ_4^W. This is probably due the relatively low signal.

The boxplots of the estimated eigenscores are displayed in Figure 12. 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 $(\xi_{ik} - \hat{\xi}_{ik})/\sqrt{\lambda_k^X}$ and $(\zeta_{ijl} - \hat{\zeta}_{ijl})/\sqrt{\lambda_l^W}$ are displayed in the first and third panels of Figure 12, respectively. The spread of the distributions increases as the signal-to-noise ratio decreases. The second and fourth panels of Figure 12 display the medians, 0.5%, 5%, 95% and 99.5% quantiles of the distribution of the normalized estimated scores.

Third scenario (3D, empirical basis)

We generate data using the first ten principal components estimated in Section 5. We replicated the unbalanced design of the MS study and used the same time variable T_{ij}'s. The principal scores ξ_{ik} and ζ_{ijk} were simulated as in Scenario 1 with $\lambda_k^X = \lambda_k^W = 0.5^{k-1}$, $k = 1, \ldots,$

10. The white noise variance σ^2 was set to 10^{-4}. Thus, SNR is equal to 1.32. The results are reported in Table 8 of the web-appendix. The average distances between estimated and true eigenvectors for $X_i(\upsilon)$ and $W_{ij}(\upsilon)$ are calculated based on 100 simulated data sets. Principal components and principal scores become less accurate as the signal-to-noise gets smaller.

5 Longitudinal analysis of brain fractional anisotropy in MS patients

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. 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. Cohort characteristics are reported in Table 10 of 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 × 72 × 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. In addition to high dimensionality, another difficulty of analyzing this study was the unbalanced distribution of scans across subjects (see Table 9 in the web-appendix); this is a typical problem in natural history studies. After forming the data matrix $\mathbf{Y}$, we estimated the overall mean as $\hat{\eta}=\frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{J_i}\boldsymbol{Y}_{ij}$ and de-meaned the data. The estimated mean is shown at Figure 16. The mean image across subjects and visits indicates a shape characterized by our scientific collaborators as a “standard corpus callosum template”.

Model 1: First, we start by fitting a random intercept and random slope model (1). To enable comparison of the variability explained by processes $X_i(\upsilon)$ and $W_{ij}(\upsilon)$, we followed the normalization procedure in Section 3.4 in Greven et al. (2010): T_{ij}’s were normalized to have sample mean zero and sample variance one. The estimated covariance matrices are not necessarily non-negative definite. Indeed, we have obtained small negative eigenvalues of the covariance operators $\mathbf{K}^{\hat{X}}$ and $\mathbf{K}^{\hat{W}}$. Following Hall et al. (2008) all the negative eigenvalues were set to zero. The total variation was decomposed into the “subject-specific” part modeled by process X_i and the “exchangeable visit-to-visit” part modeled by the 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.

Table 2 reports the percentage explained by first 10 eigenimages. The first 10 random intercept eigenimages explain roughly 55% of the total variability, while the effect of the

random slope is accounting for only 0.80% of the total variability. The exchangeable variability captured by $W_{ij}(\upsilon)$ accounts for 17.5% of the total variation.

The first three estimated random intercept and slope eigenimages are shown in pairs in Figures 6, 7, and in 17, 18, 19, 20 in the web-appendix, respectively. Figures 21, 22, and 23 in the web-appendix display the first three eigenimages of the exchangeable measurement error process $W_{ij}(\upsilon)$.

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. 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 indicating if a voxel is 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 $X_i(\upsilon)$ can be compared as they share the same principal score. This allows us to contrast the time invariant random intercept with the longitudinal 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 6 and 7 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 random slope part, $\phi_1^{X,1}$, has both positively and negatively loaded areas in the corpus callosum. The areas colored in blue shades share the sign of the random intercept $\phi_1^{X,0}$ whereas the red shades 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 in panels 0 through 5 of Figure 7. 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}. The other two eigenimages of $X_i(\upsilon)$ and eigenimages of $W_{ij}(\upsilon)$ are discussed in the web-appendix.

Next, we explored whether the deviation process $W_{ij}(\upsilon)$ depends on MS severity by analyzing the corresponding eigenscores. To do this, we divided subjects according to their MS type into three sub-groups: relapsing-remitting (RR, 102 subjects), secondary progressive (SP, 40 subjects), and primary progressive (PP, 25 subjects). For each of the first ten eigenimages, we formally tested whether there are differences between the distributions of the scores of the three groups using the t-test and the Mann-Whitney-

Wilcoxon-rank test for equality of means and the Kolmogorov-Smirnov test for equality of distributions. For the first eigenimage, the scores in the SP group have been significantly different from both those in RR and PP groups (p-values < 0.005 for all three tests). For the second eigenimage, scores in the RR group were significantly different from both SP and PP (p-values < 0.01 for all three tests). The two left images of Figure 8 display the group bean plots of the scores for the first eigenimage and the second eigenimages of $W_{ij}(\upsilon)$, respectively.

In addition to MS type, the EDSS scores were recorded at each visit. We divided subjects into two groups according to their EDSS score: i) smaller than 5 and ii) larger than or equal 5. As with MS type, we have conducted tests for the equality of distributions of the eigenscores of these two groups for all ten eigenimages. For eigenimages one and two, the distributions of eigenscores have been found to be significantly different (p-values <0.001 for all three tests). The two right images on Figure 8 display group bean plots of the scores for the first eigenimage and the second eigenimage of $W_{ij}(\upsilon)$, respectively.

We have also conducted a standard analysis based on the scalar mean FA over the CC for each subject/visit and fitted a scalar random intercept/random slope model. In this model, the random intercept explains roughly 94% of the total variation of the mean FAs. Figure 24 displays bean plots of the estimated random intercepts stratified by EDDS score and MS type. For both cases there was a statistically significant difference between the distributions of the random intercepts (EDSS: p-values < 0.001; MS-type, SP vs RR and PP, p-values < 0.002, for all three tests). Similar tests for the distributions of the random slopes did not identify statistically significant differences between groups. We conclude that this simple model agrees with the full HD-LFPCA mode, though the multivariate model provides a detailed decomposition of the total FA variation together with localization variability in the original 3D-space.

Model 2: Second, we fit model (8) using $Z_{ij,1}$ equal to a visit-specific EDSS score. Again, $Z_{ij,1}$'s were normalized to have sample mean 0 and sample variance 1. Table 3 reports percentages explained by the first 10 eigenimages in Model 2. Interestingly, the total variation explained by the random intercept and random slope in both models are approximately the same with 56.0% in Model 1 vs. 54.2% for Model 2. However, the random slope in Model 2 explains a much higher proportion of the total variation: 13.2% in Model 2 using EDSS versus Model 1 using time. The second component of the random slope explains almost 8.5% of the total variation. We have also explored whether the scores of $W_{ij}(\upsilon)$ depend on MS type and EDSS score using the t-test, the Mann-Whitney-Wilcoxon-rank test and the Kolmogorov-Smirnov test. For the first eigenimage, the SP type was significantly different from the RR (p-values < 0.01 for all three tests), though, it was not significantly different from the PP group. For the second eigenimage, the distribution of eigenscores for the SP type was significantly different from that of the scores for the RR (p-values < 0.05 for all three tests), and not significantly different from the distribution of the scores of the PP type. For grouping according to EDSS score, the distributions of the eigenscores of the first two eigenimages have been found to be statistically different (p-values <0.01 for all three tests). Figure 9 displays bean plots similar to Figure 8 for the distributions of the scores in the groups defined by MS types and EDSS. This indicates that

the deviation process $W_{ij}(\upsilon)$ in Models 1 and 2 carries not only useful but also almost identical remaining information regarding severity of MS.

6 Discussion

The methods developed in this paper increase the scope and general applicability of LFPCA 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 that arise with high-dimensional data using a powerful approach referred to as HD-LFPCA. We have developed a procedure designed to identify a low-dimensional space that contains all the information for estimating 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.

There are several outstanding issues for HD-LFPCA that need to be addressed. First, a key assumption of our methods is that they require a moderate sample size that does not exceeds ten thousands, or so, images. This limitation can be circumvented by adopting the methods discussed in Appendix. Second, we have not formally included white noise in our model. Simulation studies in Section 4 demonstrated that a moderate amount of white noise does not have a serious effect on the estimation procedure. However, a more systematic treatment of the related issues is required.

In summary, HD-LFPCA provides a powerful conceptual and practical step towards developing estimation methods for structured ultra high dimensional data.

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. http://www.slicer.org/. web-site. 2011

Aston J, Chiou J-M, Evans JP. Linguistic Pitch Analysis using Functional Principal Component Mixed Effect Models. Journal of the Royal Statistical Society, Series C. 2010; 59:297–317.

Bigelow JL, Dunson DB. Bayesian Semiparametric Joint Models for Functional Predictors. Journal of the American Statistical Association. 2009; 104:26–36.

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

Crainiceanu CM, Caffo BS, Luo S, Zipunnikov VV, Punjabi NM. Population Value Decomposition, a Framework for the Analysis of Image Populations. JASA. 2011; 106(495):775–790.

Crainiceanu CM, Staicu A-M, Di C-Z. Generalized Multilevel Functional Regression. Journal of the American Statistical Association. 2009; 104(488):1550–1561. [PubMed: 20625442]

Demmel, JW. Applied Numerical Linear Algebra. SIAM; 1997.

Di C-Z, Crainiceanu CM, Caffo BS, Punjabi NM. Multilevel Functional Principal Component Analysis. Annals of Applied Statistics. 2008; 3(1):458–488. [PubMed: 20221415]

Di C-Z, Crainiceanu CM, Jank W. Multilevel Sparse Functional Principal Component Analysis. Johns Hopkins University Working Paper. 2010

Everson R, Roberts S. Inferring the Eigenvalues of Covariance Matrices from Limited, Noisy Data. IEEE Transactions on Signal Processing. 2000; 48(7):2083–2091.

Goldsmith J, Crainiceanu CM, Caffo BS, Reich DS. Penalized Functional Regression Analysis of White-Matter Tract Profiles in Multiple Sclerosis. Neurolmage. 2011; 57(2):431–439.

Golub, GH.; Loan, CV. Matrix Computations. The Johns Hopkins University Press; 1996.

Greven S, Crainiceanu CM, Caffo BS, Reich D. Longitudinal Functional Principal Component Analysis. Electronic Journal of Statistics. 2010; 4:1022–1054. [PubMed: 21743825]

Guo W. Functional Mixed Effects Models. Biometrics. 2002; 58:121–128. [PubMed: 11890306]

Hall P, Muller H-G, Yao F. Modelling Sparse Generalized Longitudinal Observations with Latent Gaussian Processes. Journal of the Royal Statistical Society: Series B. 2008; 70(4):703–723.

Harville D. Extension of the Gauss-Markov Theorem to Include the Estimation of Random Effects. The Annals of Statistics. 1976; 4(2):384–395.

Hua ZW, Dunson DB, Gilmore JH, Styner M, Zhu HT. Semiparametric Bayesian Local Functional Models for Diffusion Tensor Tract Statistics. Neurolmage. 2012; 63:460–474.

Karhunen K. Uber Lineare Methoden in der Wahrscheinlichkeitsrechnung. Annales Academie Scientiarum Fennicae. 1947; 37:1–79.

Li Y, Zhu H, Shen D, Lin W, Gilmore JH, Ibrahim JG. Multiscale Adaptive Regression Models for Neuroimaging Data. Journal of the Royal Statistical Society: B. 2011; 73:559–578.

Loeve, M. Probability Theory. Springer-Verlag; 1978.

McCulloch, C.; Searle, S. Generalized, Linear, and Mixed Models. Wiley Interscience; 2001.

Minka TP. Automatic Choice of Dimensionality for PCA. Advances in Neural Information Processing Systems. 2000; 13:598–604.

MIPAV. http://mipav.cit.nih.gov. web-site. 2011.

Mohamed, A.; Davatzikos, C. Medical Image Computing and Computer-Assisted Intervention. Springer: chap. Shape Representation via Best Orthogonal Basis Selection; 2004.

Mori, S. Introduction to Diffusion Tensor Imaging. Elsevier; 2007.

Morris JS, Baladandayuthapani V, Herrick RC, Sanna P, Gulstein H. Automated Analysis of Quantitative Image Data Using Isomorphic Functional Mixed Models, with Application to Proteomics Data. Annals of Applied Statistics. 2011; 5(2A):894–923. [PubMed: 22408711]

Morris JS, Carroll RJ. Wavelet-Based Functional Mixed Models. Journal of the Royal Statistical Society: B. 2006; 68:179–199.

Pujol, S. National Alliance for Medical Image Computing (NA-MIC); 2010. 3D-Slicer (tutorial).

Raine, CS.; McFarland, H.; Hohlfeld, R. Multiple Sclerosis: A Comprehensive Text. Saunders Ltd.; 2008.

Reich DS, Ozturk A, Calabresi PA, Mori S. Automated vs Conventional Tractography in Multiple Sclerosis: Variablity and Correlation with Disability. Neurolmage. 2010; 49:3047–3056.

Reiss PT, Ogden RT. Functional Generalized Linear Models with Applications to Neuroimaging. Poster presentation Workshop on Contemporary Frontiers in High-Dimensional Statistical Data Analysis. 2008

Reiss PT, Ogden RT. Functional Generalized Linear Models with Images as Predictors. Biometrics. 2010; 66.1:61–69. [PubMed: 19432766]

Reiss PT, Ogden RT, Mann J, Parsey RV. Functional Logistic Regression with PET Imaging Data: A Voxel-Level Clinical Diagnostic Tool. Journal of Cerebral Blood Flow & Metabolism. 2005; 25:s635.

Rodriguez A, Dunson DB, Gelfand AE. Nonparametric Functional Data Analysis through Bayesian Density Estimation. Biometrika. 2009; 96:149–162. [PubMed: 19262739]

Roweis S. EM Algorithms for PCA and SPCA. Advances in Neural Information Processing Systems. 1997; 10:626–632.

Shinohara R, Crainiceanu C, Caffo B, Gaita MI, Reich DS. Population Wide Model-Free Quantification of Blood-Brain-Barrier Dynamics in Multiple Sclerosis. Neurolmage. 2011; 57(4): 1430–1446.

Shou H, Zipunnikov V, Crainiceanu C, Greven S. Structured Functional Principal Component Analysis. arXiv:1304.6783. 2013

Staicu A-M, Crainiceanu CM, Carroll RJ. Fast Analysis of Spatially Correlated Multilevel Functional Data. Bio statistics. 2010; 11(2):177–194.

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

Xiao L, Ruppert D, Zipunnikov V, Crainiceanu C. Fast Covariance Estimation for High-Dimensional Functional Data. 2013

Yuan Y, Gilmore JH, Geng X, Styner M, Chen K, Wang JL, Zhu H. Fmem: Functional Mixed Effects Modeling for the Analysis of Longitudinal White Matter Tract Data. Neurolmage. 2014; 84:753–764.

Zhao H, Yuen PC, Kwok JT. A Novel Incremental Principal Component Analysis and Its Application for Face Recognition. IEEE Transactions on Systems, Man, and, Cybernatics, Part B: Cybernatics. 2006; 36(4):873–886.

Zhu H, Brown PJ, Morris JS. Robust Adaptive Functional Regression in Functional Mixed Model Framework. Journal of the American Statistical Association. 2011; 106(495):1167–1179. [PubMed: 22308015]

Zipunnikov V, Caffo B, Yousem DM, Davatzikos C, Schwartz BS, Crainiceanu CM. Functional Principal Component Models for High Dimensional Brain Volumetrics. Neurolmage. 2011a; 58:772–784.

Zipunnikov V, Caffo B, Yousem DM, Davatzikos C, Schwartz BS, Crainiceanu CM. Multilevel Functional Principal Component Analysis for High-Dimensional Data. Journal of Computational and Graphical Statistics. 2011b; 20(4):852–873.

Appendix

AI. Large sample size

The main assumption which has been made in the paper 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 $\tilde{\mathbf{Y}}_{ij}$ as a linear combination of columns of matrices $\mathbf{\Phi}^{X,0}$, $\mathbf{\Phi}^{X,1}$, $\mathbf{\Phi}^{W}$. Assuming that $2N_X + N_W < n$, each $\tilde{\mathbf{Y}}_{ij}$ belongs to an at most $(2N_X + N_W)$-dimensional linear space $\mathscr{L}(\mathbf{\Phi}^{X,0}, \mathbf{\Phi}^{X,1}, \mathbf{\Phi}^{W})$ spanned by those columns. Thus, if model (4) holds exactly the rank of the matrix $\tilde{\mathbf{Y}}$ 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 (7) 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 (7) 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 Y. Of course, in practice, model (4) never holds exactly. Hence, the number of columns of matrix V should be chosen to be large enough to either reasonably exceed $(2N_X + N_w)$ or to capture the most of the 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 $\tilde{\mathbf{Y}}$.

There are 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 $\tilde{\mathbf{Y}}$ during which information about the first k singular vectors is accumulated sequentially. Their complexity is of order $O(k^3p)$. 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 $\tilde{\mathbf{Y}}$. Further development in this direction is beyond the scope of this paper.

A2. Proofs

Proof of Lemma 1: Using the independence of $\mathbf{Y}_i$ and $\mathbf{Y}_k$, the expectation of pairwise quadratics is

$$E(\boldsymbol{Y}_{ij_1}\boldsymbol{Y}'_{kj_2}) = \begin{cases} \eta\eta' & , \text{if} \quad k \neq i \\ \eta\eta' + \mathbf{K}_X^{00} + T_{ij_2}\mathbf{K}_X^{01} + T_{ij_1}\mathbf{K}_X^{10} + T_{ij_1}T_{ij_2}\mathbf{K}_X^{11} + \delta_{j_1j_2}\mathbf{K}^W & , \text{if} \quad i=k, \end{cases} \tag{13}$$

where δ_{j1j2} is 1 if $j_1 = j_2$ and 0 otherwise. From the top equality we get the MM estimator of the mean, $\hat{\boldsymbol{\eta}} = n^{-1}\sum_{i,j}\mathbf{Y}_{ij}$ The covariances $\mathbf{K}^x$ and $\mathbf{K}^w$ can be estimated by de-meaning $\mathbf{Y}_{ij}$ as $\tilde{\mathbf{Y}}_{ij} = \mathbf{Y}_{ij} - \hat{\boldsymbol{\eta}}$ and regressing $\tilde{\mathbf{Y}}_{ij_1}\tilde{\mathbf{Y}}'_{ij_2}$ on 1, T_{ij2}, T_{ij1}, $T_{ij1}T_{ij2}$ and δ_{j1j2}. The bottom equality can be written as $E(\tilde{\mathbf{Y}}^{\upsilon}_{ij_1j_2}) = \mathbf{K}^{\upsilon}\mathbf{f}_{ij_1j_2}$, where $\tilde{\mathbf{Y}}^{\upsilon}_{ij_1j_2} = \tilde{\mathbf{Y}}_{ij_2} \otimes \tilde{\mathbf{Y}}_{ij_1}$ is a $p^2 \times 1$ dimensional vector, the parameter of interest is the $p^2 \times 5$ matrix $\mathbf{K}^{\upsilon} = [\mathrm{vec}(\mathbf{K}_X^{00}), \mathrm{vec}(\mathbf{K}_X^{01}), \mathrm{vec}(\mathbf{K}_X^{10}), \mathrm{vec}(\mathbf{K}_X^{11}), \mathrm{vec}(\mathbf{K}^W)]$, and the covariates are entries m the 5×1 vector $\mathbf{f}_{ij_1J_2} = (\mathrm{l}, T_{ij_2}, T_{ij_1}, T_{ij_1}T_{ij_2}, \delta_{j1j2})'$. With these notations $E\mathbf{Y}^{\upsilon} = \mathbf{K}^{\upsilon}\mathbf{F}$, where $\tilde{\mathbf{Y}}^{\upsilon}$ is $p^2 \times m$ dimensional with $m = \sum_{i=1}^{I} J_i^2$ $\mathbf{F}$ 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 MM estimator of $\mathbf{K}^{\upsilon}$ is thus $\hat{\mathbf{K}^{\upsilon}} = \tilde{\mathbf{Y}}^{\upsilon}\mathbf{F}'(\mathbf{F}\mathbf{F}')^{-1}$ which provides unbiased estimators of the covariances $\mathbf{K}^x$ and $\mathbf{K}^w$. If we denote $\mathbf{H} = \mathbf{F}'(\mathbf{F}\mathbf{F}')^{-1}$, we get the result of the lemma.

Proof of Lemma 2: Let us denote by $\hat{\mathbf{K}}^X_{\mathrm{U}}$ and $\hat{\mathbf{K}}^W_{\mathrm{U}}$ the matrices denned by equations (5) with $\mathbf{S}^{1/2}\mathbf{U}_{ij_1}\mathbf{U}'_{ij_2}\mathbf{S}^{1/2}$ substituted for $\tilde{\mathbf{Y}}_{ij_1}\tilde{\mathbf{Y}}'_{ij_2}$. The $2n \times 2n$ dimensional matrix $\hat{\mathbf{K}}^X_{\mathrm{U}}$ and the $n \times n$ dimensional matrix $\hat{\mathbf{K}}^W_{\mathrm{U}}$ are low-dimensional counterparts of $\mathbf{K}^{\hat{x}}$ and $\mathbf{K}^{\hat{W}}$, respectively. Using the SVD representation $\tilde{\mathbf{Y}}_{ij} = \mathbf{V}\mathbf{S}^{1/2}\mathbf{U}_{ij}$, the estimated high dimensional

covariance matrices can be represented as $\hat{\mathbf{K}}^X = \mathbf{D}\hat{\mathbf{K}}^X_{\mathrm{U}}\mathbf{D}'$ and $\hat{\mathbf{K}}^W = \mathbf{V}\hat{\mathbf{K}}^W_{\mathrm{U}}\mathbf{V}'$, where the matrix **D** is $2p \times 2n$ dimensional with orthonormal columns denned as

$$\mathbf{D} = \begin{pmatrix} \mathbf{V} & \mathbf{0}_{p\times n} \\ \mathbf{0}_{p\times n} & \mathbf{V} \end{pmatrix}. \quad (14)$$

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

Proof of Lemma 3: With notational changes, the proof is identical to the proof of Lemma

1. **Proof of Lemma 4**: With notational changes, the proof is identical to the proof of Lemma

2. **Proof of Lemma 5**: The main idea of the proof is similar to that of Zipunnikov et al. (2011b). We assume that function $\eta(\upsilon, T_{ij}) = 0$. From the model 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 $p \le N_x + J_i N_W$ the BLUP of $\boldsymbol{\omega}_i$ is given by

   $\hat{\omega}_i = Cov(\boldsymbol{\omega}_i, vec(\tilde{\mathbf{Y}}_i))\, Var(vec(\tilde{\mathbf{Y}}_i))^{-1} vec(\tilde{\mathbf{Y}}_i) = \mathbf{\Lambda}_{\boldsymbol{\omega}}\mathbf{B}_i'(\mathbf{B}_i\mathbf{\Lambda}_{\boldsymbol{\omega}}\mathbf{B}_i')^{-1} vec(\tilde{\mathbf{Y}}_i)$ (see McCulloch and Searle, 2001, Section 9). The BLUP is essentially a projection and thus it does not require any distributional assumptions. It may be denned in terms of a projection matrix. If $\boldsymbol{\xi}_i$ and $\boldsymbol{\zeta}_{ij}$ are normal then the BLUP is the best predictor. When $p > N_x + J_i N_w$ the 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,

   $\hat{\omega}_i = \mathbf{\Lambda}_{\boldsymbol{\omega}}\mathbf{B}_i'(\mathbf{B}_i\mathbf{\Lambda}_{\boldsymbol{\omega}}\mathbf{B}_i')^{-} vec(\tilde{\mathbf{Y}}_i) = \mathbf{\Lambda}_{\boldsymbol{\omega}}^{1/2}(\mathbf{\Lambda}_{\boldsymbol{\omega}}^{1/2}\mathbf{B}_i'\mathbf{B}_i\mathbf{\Lambda}_{\boldsymbol{\omega}}^{1/2})^{-1}\mathbf{\Lambda}_{\boldsymbol{\omega}}^{1/2}\mathbf{B}_i' vec(\tilde{\mathbf{Y}}_i) = (\mathbf{B}_i'\mathbf{B}_i)^{-1}\mathbf{B}_i' vec(\tilde{\mathbf{Y}}_i)$
   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 $\hat{\omega}_i = (\hat{\mathbf{B}}_i'\hat{\mathbf{B}}_i)^{-1}\hat{\mathbf{B}}_i' vec(\tilde{\mathbf{Y}}_i)$.

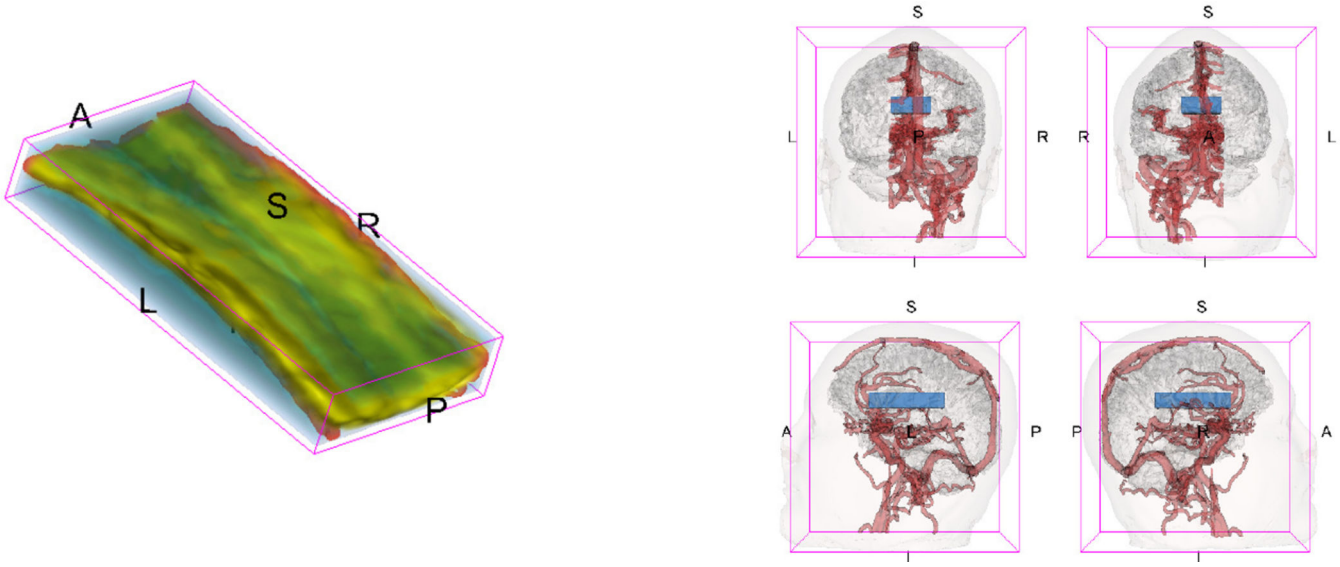

Figure 1.

The 3D-rendering of the region of interest (left), a blue block containing corpus callosum, and the template brain (right). 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).

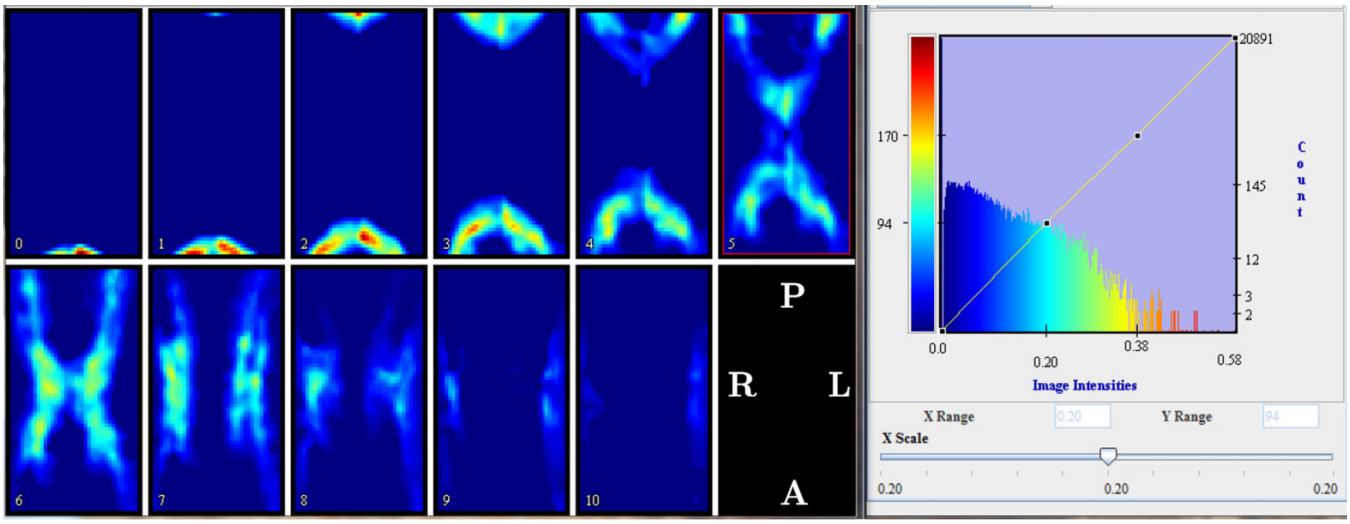

Figure 2.

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).

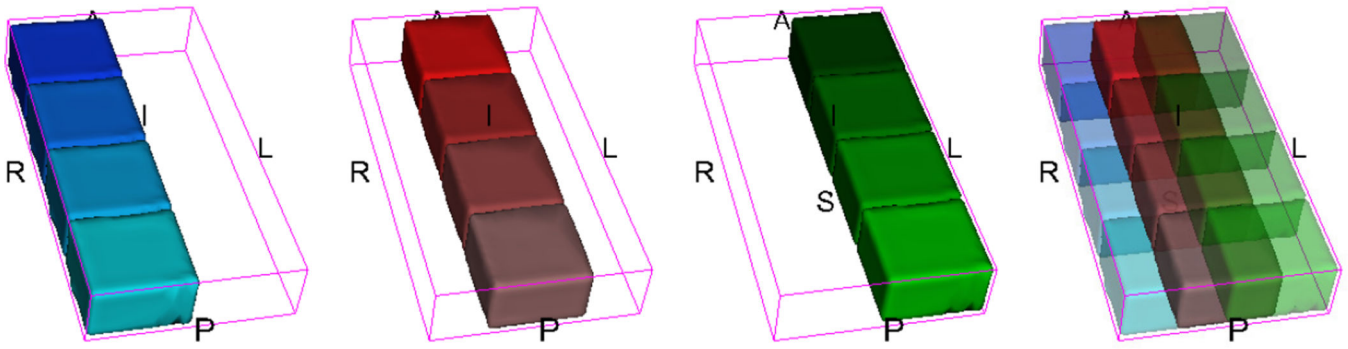

Figure 3.

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).

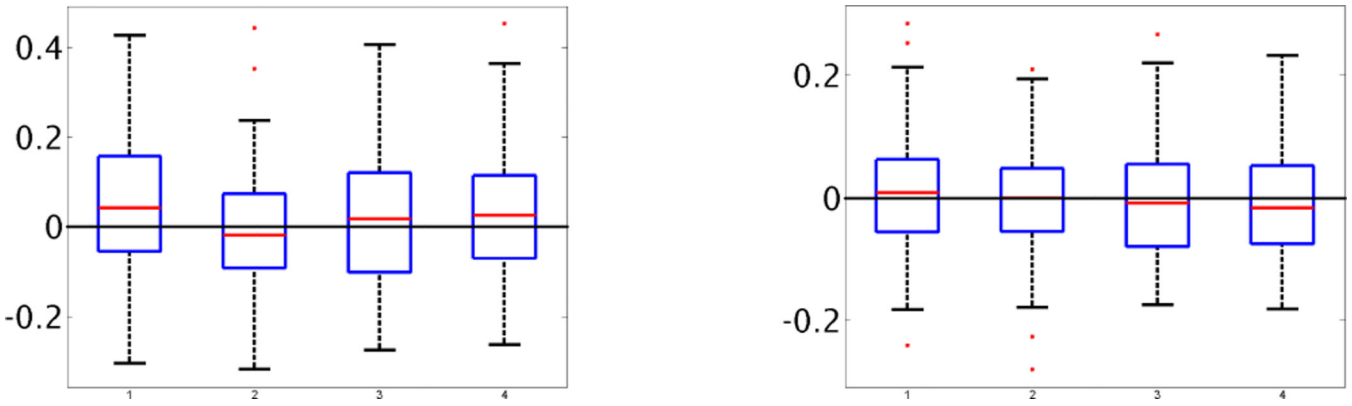

Figure 4.

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

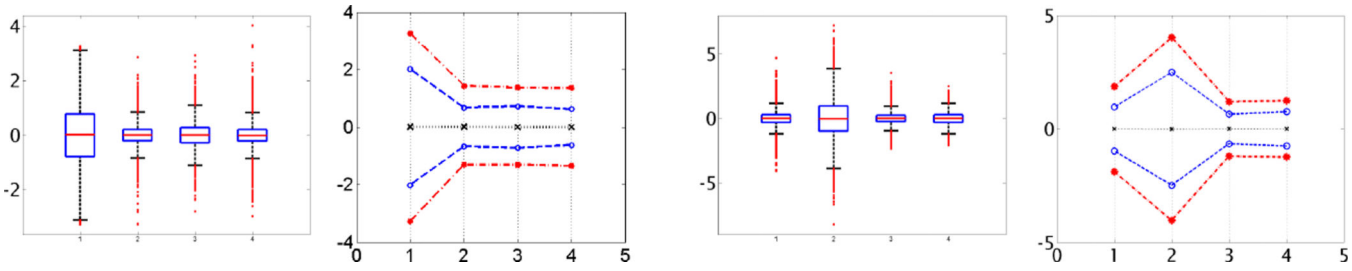

Figure 5.

The left two panels show the distribution of the normalized estimated scores of process $X_i(\upsilon), (\xi_{ik} - \hat{\xi}_{ik})/\sqrt{\lambda_k^X}$. 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 $W_{ij}(\upsilon), (\zeta_{ijl} - \hat{\zeta}_{ijl})/\sqrt{\lambda_l^X}$ is provided at the right two panels.

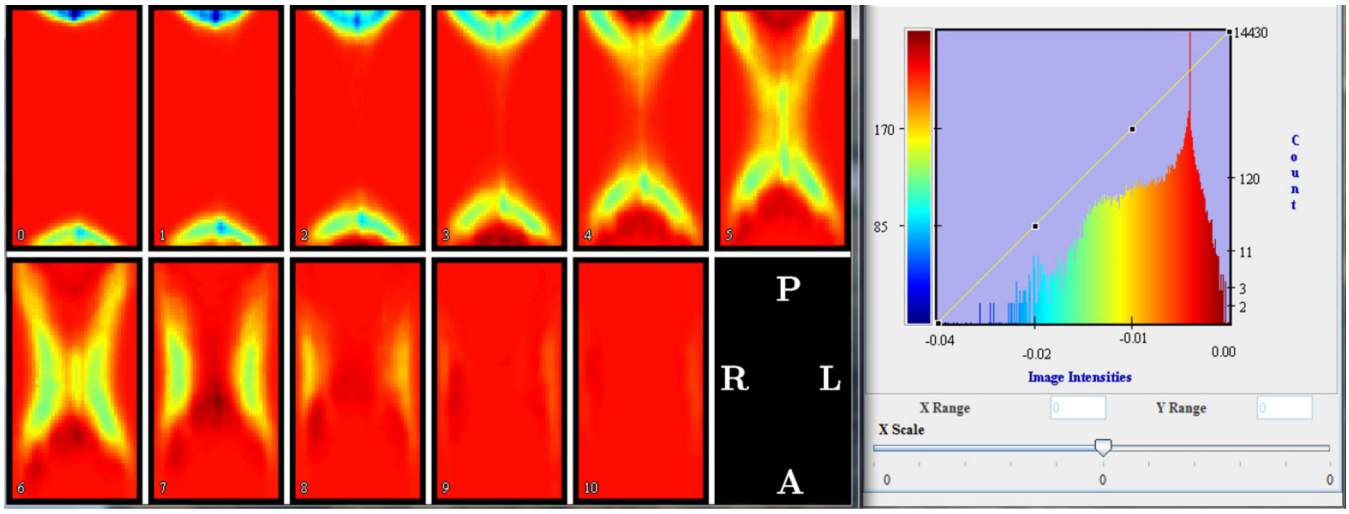

Figure 6.

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).

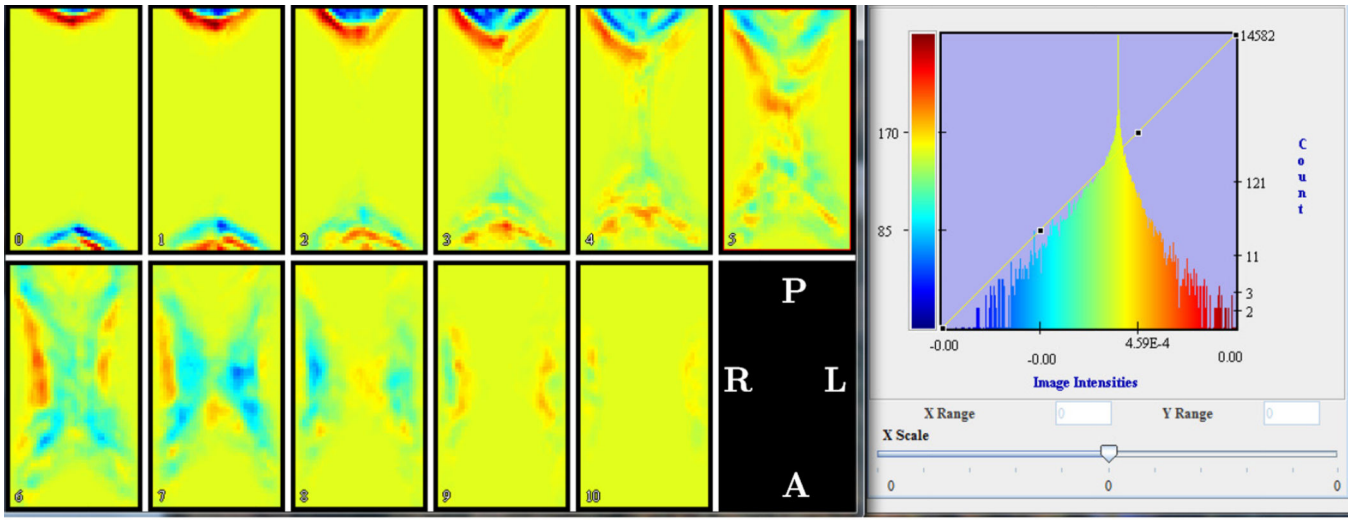

Figure 7.

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).

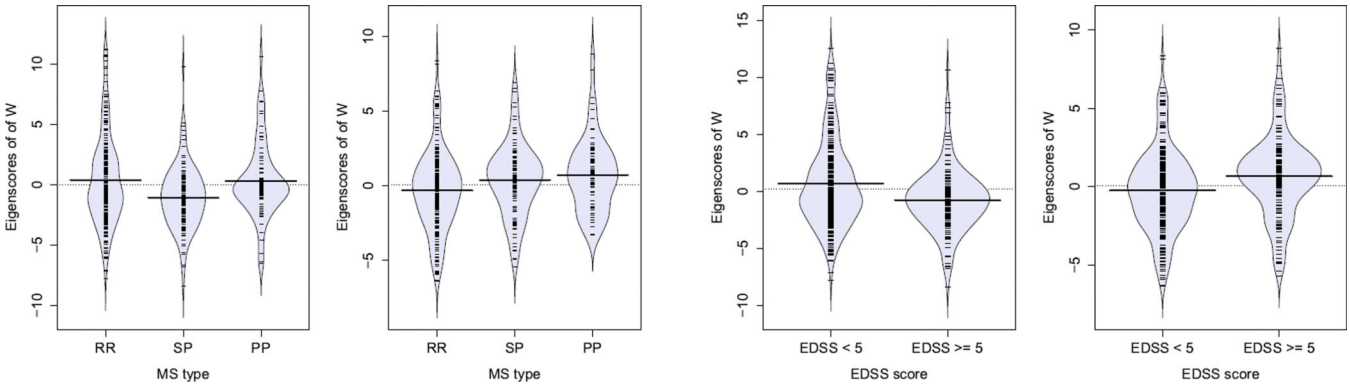

Figure 8.
Model 1: Group bean plots according to MS type (left) and according to EDSS score (right).

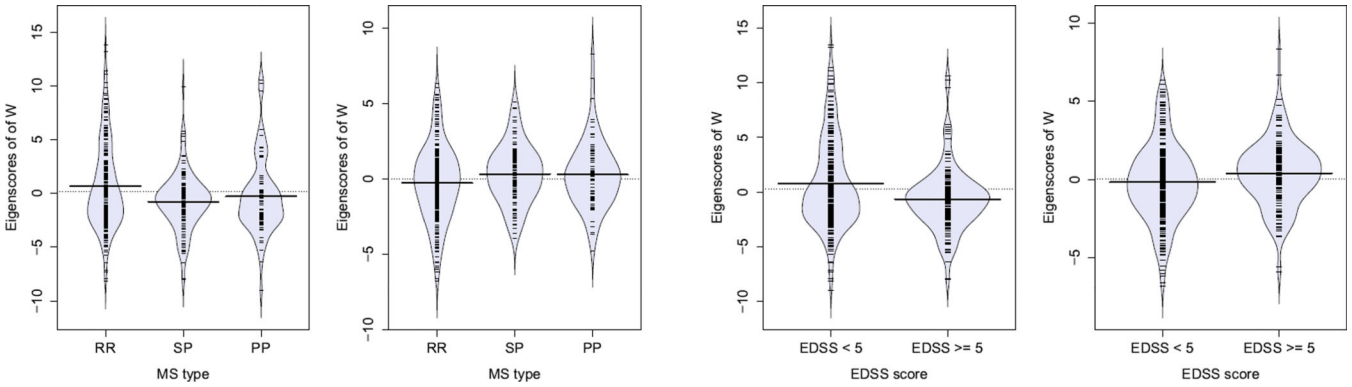

Figure 9.
Model 2: Group bean plots according to MS type (left) and according to EDSS score (right).

Table 1

Based on 100 simulated datasets, average distances between estimated and true eigenvectors of $X_{i,0}(\upsilon)$, standard deviations are given in parentheses.

(p, σ^2)	$\|\phi_1^{X,0} - \hat{\phi}_1^{X,0}\|^2$	$\|\phi_2^{X,0} - \hat{\phi}_2^{X,0}\|^2$	$\|\phi_3^{X,0} - \hat{\phi}_3^{X,0}\|^2$	$\|\phi_4^{X,0} - \hat{\phi}_4^{X,0}\|^2$
(750, 1e-04)	0.034 (0.048)	0.07 (0.069)	0.074 (0.053)	0.081 (0.07)
(750, 5e-04)	0.031 (0.031)	0.055 (0.051)	0.084 (0.097)	0.112 (0.151)
(750, 0.001)	0.035 (0.039)	0.062 (0.054)	0.078 (0.059)	0.139 (0.206)
(750, 0.005)	0.035 (0.039)	0.072 (0.062)	0.096 (0.063)	0.159 (0.084)
(750, 0.01)	0.045 (0.036)	0.079 (0.054)	0.129 (0.102)	0.234 (0.103)
(3000, 1e-04)	0.031 (0.028)	0.064 (0.118)	0.09 (0.13)	0.109 (0.126)
(3000, 5e-04)	0.037 (0.032)	0.065 (0.048)	0.077 (0.06)	0.14 (0.136)
(3000, 0.001)	0.031 (0.027)	0.06 (0.044)	0.087 (0.062)	0.131 (0.07)
(3000, 0.005)	0.058 (0.035)	0.106 (0.058)	0.171 (0.09)	0.324 (0.096)
(3000, 0.01)	0.073 (0.028)	0.142 (0.048)	0.236 (0.074)	0.508 (0.072)
(12000, 1e-04)	0.031 (0.028)	0.062 (0.048)	0.077 (0.056)	0.134 (0.165)
(12000, 5e-04)	0.041 (0.036)	0.078 (0.05)	0.121 (0.069)	0.201 (0.081)
(12000, 0.001)	0.047 (0.04)	0.083 (0.054)	0.164 (0.114)	0.295 (0.118)
(12000, 0.005)	0.112 (0.032)	0.217 (0.064)	0.44 (0.216)	0.758 (0.153)
(12000, 0.01)	0.175 (0.031)	0.338 (0.093)	0.554 (0.132)	0.987 (0.071)
(24000, 1e-04)	0.035 (0.032)	0.066 (0.049)	0.09 (0.141)	0.146 (0.173)
(24000, 5e-04)	0.055 (0.045)	0.097 (0.061)	0.146 (0.09)	0.266 (0.098)
(24000, 0.001)	0.07 (0.038)	0.125 (0.047)	0.23 (0.167)	0.43 (0.15)
(24000, 0.005)	0.183 (0.049)	0.348 (0.097)	0.622 (0.208)	0.998 (0.11)
(24000, 0.01)	0.295 (0.043)	0.518 (0.117)	0.742 (0.102)	1.184 (0.07)
(48000, 1e-04)	0.046 (0.068)	0.076 (0.067)	0.103 (0.059)	0.175 (0.122)
(48000, 5e-04)	0.073 (0.035)	0.13 (0.056)	0.234 (0.1)	0.437 (0.099)
(48000, 0.001)	0.105 (0.051)	0.183 (0.065)	0.407 (0.23)	0.695 (0.192)
(48000, 0.005)	0.307 (0.08)	0.532 (0.151)	0.824 (0.208)	1.19 (0.086)
(48000, 0.01)	0.458 (0.084)	0.712 (0.1)	0.938 (0.074)	1.186 (0.126)
(96000, 1e-04)	0.045 (0.033)	0.087 (0.059)	0.146 (0.103)	0.246 (0.107)
(96000, 5e-04)	0.116 (0.081)	0.194 (0.094)	0.431 (0.268)	0.721 (0.218)
(96000, 0.001)	0.188 (0.089)	0.32 (0.121)	0.787 (0.339)	1.062 (0.216)
(96000, 0.005)	0.457 (0.065)	0.707 (0.107)	0.954 (0.125)	1.298 (0.074)
(96000, 0.01)	0.662 (0.105)	0.926 (0.103)	1.116 (0.075)	1.143 (0.153)

Table 2

Model 1 (T_{ij} change): Cumulative variability explained by the first 10 eigenimages.

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

Model 2 (Z_{ij} change): Cumulative variability explained by the first 10 eigenimages.

k	$\phi_k^{X,0}$	$\phi_k^{X,1}$	ϕ_k^W	**cumulative**
1	17.79	0.42	5.59	23.80
2	0.53	8.46	1.99	34.78
3	6.92	0.39	1.55	43.64
4	4.68	0.76	1.05	50.13
5	3.02	0.52	0.80	54.46
6	2.44	0.29	0.69	57.88
7	1.63	0.77	0.54	60.82
8	1.48	0.67	0.39	63.36
9	1.41	0.51	0.35	65.64
10	1.19	0.38	0.33	67.54
	41.09	13.17	13.28	67.54