10-6-2010

MULTILEVEL FUNCTIONAL PRINCIPAL COMPONENT ANALYSIS FOR HIGH-DIMENSIONAL DATA

Vadim Zipunnikov
Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics, vzipunni@jhsph.edu

Brian Caffo
Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics

Ciprian Crainiceanu
Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics

David M. Yousem
Johns Hopkins Hospital, Department of Radiology

Christos Davatzikos
University of Pennsylvania, Department of Radiology

See next page for additional authors

Authors
Vadim Zipunnikov, Brian Caffo, Ciprian Crainiceanu, David M. Yousem, Christos Davatzikos, and Brian S. Schwartz

Multilevel Functional Principal Component Analysis for High-Dimensional Data

Vadim Zipunnikov, Brian Caffo, Ciprian Crainiceanu,

David M. Yousem, Christos Davatzikos, Brian S. Schwartz*

June 1, 2010

Abstract

We propose fast and scalable statistical methods for the analysis of hundreds or thousands of high dimensional vectors observed at multiple visits. The proposed inferential methods avoid the difficult task of loading the entire data set at once in the computer memory and use sequential access to data. This allows deployment of our methodology on low-resource computers where computations can be done in minutes on extremely large data sets. Our methods are motivated by and applied to a study where hundreds of subjects were scanned using Magnetic Resonance Imaging (MRI) at two visits roughly five years apart. The original data possesses over ten billion measurements. The approach can be applied to any type of study where data can be unfolded into a long vector including densely observed functions and images.

Key Words: Voxel-based morphometry; MRI; brain imaging data.

*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). David M. Yousem is a Professor, Department of Radiology, Johns Hopkins Hospital. Christos Davatzikos is a Professor of Radiology, University of Pennsylvania. Brian S. Schwartz is a Professor of Environmental Health Sciences, Epidemiology, and Medicine, Johns Hopkins Bloomberg School of Public Health

1 Introduction

Massive and complex data sets raise a host of statistical challenges that were unthinkable, even a few years ago. For example, functional and anatomic imaging data on many subjects at one or multiple visits is now routinely being collected in many scientific studies. The volume and dimensionality of these types of data make standard inferential tools computationally infeasible and lead to intensive methodological research (Di et al. (2008), Crainiceanu et al. (2009), Staicu et al. (2010), Greven et al. (2009), Di and Crainiceanu (2010), Zipunnikov et al. (2010), Crainiceanu et al. (2010), Mohamed and Davatzikos (2004), Reiss et al. (2005), Reiss and Ogden (2008), Reiss and Ogden (2010)). Specifically, in this paper we are concerned with modelling data with the following structure: 1) measurements are observed on hundreds or thousands of subjects; 2) each subject is observed at multiple visits; and 3) each individual observation at each visit is very high dimensional. In particular, we focus on MRI data where each image is composed of rougly 3 million measurements at each subject and visit. These types of data are opening an area of research refered to as large n (number of subjects), large J (number of visits), and large p (dimensionality of the observations) problem (Crainiceanu et al., 2010).

To address challenges arising from these types of data we develop Multilevel Functional Principal Component Analysis for High Dimensional (HD-MFPCA) data. HD-MFPCA combines powerful data compression techniques and statistical inference to decompose the observed data in population- and visit-specific means and subject-specific within and between level variability. Our inferential methods are very fast even on personal computers for the analysis of hundreds or thousands of very large images. To be precise, it currently takes roughly 16 minutes on a standard personal computer for HD-MFPCA analysis of images of hundreds of subjects at two visits. This was achieved by 1) designing new methods for obtaining Karhunen-Loeve (K-

L) expansions of covariance operators; 2) obtaining best linear unbiased predictors (BLUPS) of subject-and visit-specific scores as a by-product of a singular value decomposition (SVD); 3) using the parallel between KL expansion and SVD; 4) relying on partitioning of images into blocks that can be loaded into the computer memory; 5) changing all calculations to rely only on block calculations and sequential access to memory; and 6) using SVD for matrices that have at least one dimension smaller than $10,000$. Using these methods our algorithms are linear in the dimensionality of the data and do not require loading the data matrix into memory, an impossible and wasteful task in many applications. HD-MFPCA opens up a road for multilevel modelling of data of size that was previously thought to be prohibitively large for statistically principled analysis.

The motivation for this work came from the analysis of cross-sectional and longitudinal brain volume via magnetic resonance imaging (MRI). The data arise from a population-based study of lead exposure and its relationship to brain structure and function (Schwartz et al., 2007; Caffo et al., 2008; Su et al., 2008; Chen et al., 2009)

1.1 Description of imaging data and its processing

In this manuscript, we consider voxel (three dimensional pixel) based analysis of brain magnetic resonance imaging (MRI) data. The motivating data arises from a study of voxel-based morphometry (VBM) (Ashburner and Friston, 2000) in former lead manufacturing workers. VBM is a whole-brain technique for studying localized changes in brain shape which has become a standard component in the toolbox for neuroscientists. The primary benefits of VBM are the lack of need for a-priori specified regions of interest and its exploratory nature that facilitates identification of spatio-temporally complex and perhaps previously unknown patterns of brain structure and function.

Typically, VBM is performed via the following steps: 1) registration methods warp each subject's brain into a template brain space where the Jacobian of the transformation is retained for analysis, 2) the Jacobian maps, which are aligned in template space, are then smoothed using an aggressive filter, and 3) the maps are analyzed across subjects using voxel-wise regression models; in this paper we do not analyze voxels separately and focus instead on the joint analysis of the whole-brain data. Steps 1) and 2) imply that the image used for analysis in VBM is not the original MRI and that intensities represent relative sizes of the original image in normalized space. That is, the intensity of a given voxel represents how relatively large or small that same space was in the subject's original image. For example, two subjects with different head sizes would have the same VBM image shape; however the smaller subject would have a darker overall image while the larger subject would have a brighter image.

The specific method we use for creating VBM images, step 1), in this study is a generalization of regional analysis of volumes in normalized space (RAVENS) algorithm (Davatzikos et al., 2001). This method is an early and influential standard in VBM analysis. To obtain more accurate registration and analysis, each subject's brain is segmented into gray matter, white matter and cerebrospinal fluid. Hence, analysis is separated by tissue type (gray or white). Moreover, the data for this analysis uses a new 4D longitudinal registration technique that reduces errors by accounting for within-subject correlations. We registered brains to a template that was aligned with the International Consortium for Brain Mapping's ICBM152 template (Mazziotta et al., 1995). For smoothing we used a 10 mm full width at half maximum (FWHM) isotropic Gaussian smoother, where the FWHM is the diameter of the spherical cross-section of the 3D Gaussian kernel at one half of the height at the mode.

The data were derived from an epidemiologic study of the central nervous system effects of organic and inorganic lead in former organolead manufacturing workers (Stewart et al. (1999), Schwartz et al. (2000), Schwartz et al. (2000)). Subject scans were from a GE 1.5 Tesla Signa scanner. RAVENS image processing was performed on the T1-weighted volume acquisitions.

Our analysis is different from VBM by exploring hierarchical models of morphometric variation using a generalization of multilevel functional principal components. That is, we offer the first approach to decompose a population of morphometric (RAVENS) images into cross-sectional and longitudinal directions of variation. While our methods are motivated by and applied to a large longitudinal study of RAVENS images, they are general and can be adapted or scaled up to other longitudinal studies where images or very large vectors are collected at multiple visits.

The remainder of the paper is organized as follows. Section 2 briefly describes the MFPCA approach and its limitation when dealing with high-dimensional data. Section 3 provides a solution to the problem and introduces HD-MFPCA as well as shows an efficient way of calculating Best Linear Unbiased Predictors (BLUPs) for principal scores of the model. The results of two simulation studies are provided in Section 4. Our methods are applied to RAVENS data in Section 5. The paper is concluded with discussion in Section 6.

2 Basic multilevel model for high-dimensional data

In this section we will provide a basic framework for MFPCA developed in (Di et al., 2008). We will discuss the reasons why the methods can not be applied directly to high-dimensional data sets such as the one described above in Section 1.1.

Suppose that we have a sample of images $\mathbf{X}_{ij}$, where $\mathbf{X}_{ij}$ is a recorded brain image of the i^{th} subject, $i = 1, \ldots, I$ at visit $j, j = 1, \ldots, J$. Every image is a 3-dimensional

array structure of dimension $p = p_1 \times p_2 \times p_3$. For example, in the RAVENS data example introduced in Section 1.1 $p = 256 \times 256 \times 198 = 12,976,128$. For the purpose of this paper we represent the data X_{ij} as a $p \times 1$ dimensional vector containing the voxels in a particular order, where the order is preserved across all voxels.

Following (Di et al., 2008) we consider the following functional two-way ANOVA model

$$X_{ij}(\mathbf{v}) = \mu(\mathbf{v}) + \eta_j(\mathbf{v}) + Z_i(\mathbf{v}) + W_{ij}(\mathbf{v}), \tag{1}$$

where $\mu(\mathbf{v})$ is the overall mean image, $\eta_j(\mathbf{v})$ is the visit-specific image shift from the overall mean image, $Z_i(\mathbf{v})$ is a subject-specific image deviation from the visit-specific population mean, and $W_{ij}(\mathbf{v})$ is the visit-specific image deviation from the subject-specific mean. We make the standard assumptions that $\mu(\mathbf{v})$ is smooth and that $Z_i(\mathbf{v})$ and $W_{ij}(\mathbf{v})$ are realizations of a square integrable stochastic process with smooth covariance matrix.

We denote by $\mathbf{K}_B$ and $\mathbf{K}_W$ the covariance operators of the latent (unobserved) processes $Z_i(v)$ and $W_{ij}(v)$ in (1), respectively. Using the standard argument about the Karhunen-Loeve expansions of the random processes (Karhunen, 1947) $Z_i(v) = \sum_{k=1}^{\infty} \xi_{ik}\phi_k^{(1)}(v)$ and $W_{ij}(v) = \sum_{l=1}^{\infty} \zeta_{il}\phi_l^{(2)}(v)$, where $\phi_k^{(1)}$ and $\phi_l^{(2)}$ are the eigenfunctions of the $\mathbf{K}_B$ and $\mathbf{K}_W$ operators, respectively. With these changes the MFPCA model becomes a mixed effect model

$$\begin{cases} X_{ij}(v) = \mu(v) + \eta_j(v) + \sum_{k=1}^{\infty} \xi_{ik}\phi_k^{(1)}(v) + \sum_{l=1}^{\infty} \zeta_{ijl}\phi_l^{(2)}(v) \\ \xi_{ik} \overset{i.i.d.}{\sim} (0, \lambda_k^{(1)}),\ \zeta_{ijl} \overset{i.i.d.}{\sim} (0, \lambda_l^{(2)}), \end{cases} \tag{2}$$

where $\overset{i.i.d}{\sim} (0, \sigma^2)$ indicates that variables are i.i.d. with mean zero and variance σ^2; please note that no distributional assumptions are required.

For practical purposes we consider a model projected on the first N_1 and N_2

components, respectively. Assuming that N_1 and N_2 are known the model becomes

$$\begin{cases} X_{ij}(v) = \mu(v) + \eta_j(v) + \sum_{k=1}^{N_1} \xi_{ik}\phi_k^{(1)}(v) + \sum_{l=1}^{N_2} \zeta_{ijl}\phi_l^{(2)}(v) \\ \xi_{ik} \overset{i.i.d.}{\sim} (0, \lambda_k^{(1)}),\ \zeta_{ijl} \overset{i.i.d.}{\sim} (0, \lambda_l^{(2)}). \end{cases} \tag{3}$$

As the number of principal components, or eigenvectors, N_1 and N_2 is typically small (Di et al. (2008)), this is a particularly nice representation of the family of images over multiple visits. Statistical estimation of this model conditional on the eigenvectors is not very difficult; see, for example, the projection idea (Di et al., 2008). Thus, estimating the eigenvectors is a crucial first step of the methods. As both $\mathbf{K}_B$ and $\mathbf{K}_W$ are $p \times p$ dimensional matrices, diagonalizing them is not possible when p exceeds, say, $10,000$. Below we explain in more details the problems associated with the standard MFPCA analysis in this context. In Section 3 we show how these problems can be avoided.

The main reason why MFPCA is infeasible within the high dimensional settings is that the covariance operators involved can not be calculated, stored or diagonalized. Indeed, standard MFPCA proceeds as follows. The mean image $\hat{\boldsymbol{\mu}}$ is estimated by the sample average $\sum_{ij} \mathbf{X}_{ij}/(IJ)$. All visit effects $\hat{\boldsymbol{\eta}}_j, j \geq 1$ are estimated by $\sum_i \mathbf{X}_{ij}/I - \hat{\boldsymbol{\mu}}$. To obtain the first and second level decompositions of (3) for the unexplained part of the image, $\widetilde{\mathbf{X}}_{ij} = \mathbf{X}_{ij} - \hat{\boldsymbol{\mu}} - \hat{\boldsymbol{\eta}}_j$, an eigenanalysis has to be performed. Denote by $\widetilde{\mathbf{X}} = (\widetilde{\mathbf{X}}_1, \ldots, \widetilde{\mathbf{X}}_J)$ where $\widetilde{\mathbf{X}}_j$ is a centered $p \times I$ matrix where the column i, $i = 1, \ldots, I$, contains the unfolded image for subject i at visit j. The overall covariance operator $\hat{\mathbf{K}}_T$ is estimated as $\hat{\mathbf{K}}_T = \frac{1}{IJ}\sum_{j=1}^{J} \widetilde{\mathbf{X}}_j \widetilde{\mathbf{X}}_j'$, which can be decomposed as a sum of "between" and "within" covariance operators defined as

$$\hat{\mathbf{K}}_B = \frac{1}{I(J^2 - J)} \sum_{j_1 \neq j_2} \widetilde{\mathbf{X}}_{j_1} \widetilde{\mathbf{X}}_{j_2}' \text{ and } \hat{\mathbf{K}}_W = \frac{1}{2I(J^2 - J)} \sum_{j_1 \neq j_2} (\widetilde{\mathbf{X}}_{j_1} - \widetilde{\mathbf{X}}_{j_2})(\widetilde{\mathbf{X}}_{j_1} - \widetilde{\mathbf{X}}_{j_2})', \tag{4}$$

respectively. It is easy to see that $\hat{\mathbf{K}}_T = \hat{\mathbf{K}}_B + \hat{\mathbf{K}}_W$. The size of operators is $p \times p$ and the standard eigenanalysis requires $O(p^3)$ operations, which makes calculations infeasible for $p > 10^4$. The data in our application is 4 to 5 orders of magnitude larger than the infeasibility limit. Even the storage of the covariance operators for the RAVENS data with $p \times p = 1.6 \cdot 10^{14}$ is impossible. Therefore, for high-dimensional problems MFPCA, which performs extremely well when the functional dimensionality is in the thousands, fails in very high dimensional setting. The next section provides a methodology capable of handling multi-level models of very high dimensionality. While some attention is required to understand the technical details, our solution works because the ranks of the matrices involved do not exceed the sample size of the study.

3 HD-MFPCA

We now describe our statistical model and inferential methods, with particular focus on solving computational problems. In particular, we show how to calculate all necessary quantities using sequential access to data that cannot be loaded into the computer memory. We introduce high-dimensional MFPCA (HD-MFPCA), which overcomes the problems discussed in the previous section and provides new and deep insights into the geometry of MFPCA. While our approach was inspired by the RAVENS data, HD-MFPCA has enormous potential for the analysis of populations of high dimensional data.

3.1 Eigenanalysis

Constructing and diagonalizing covariance operators for high dimensional data is impossible, which makes a brute-force implementation of the methods in (Di et al.,

2008) infeasible to the type of problems discussed here. In this section we propose an algorithm for calculating the eigenvalues and eigenvectors of the relevant covariance operators that avoids the problem mentioned above. This crucial result uses a sequential access to the data and allows to perform multilevel principal component analysis on observations of extremely high dimensionality.

Our approach starts with constructing the SVD of the matrix $\widetilde{\mathbf{X}}$, the matrix obtained by column-binding the centered subject-specific data matrices

$$\widetilde{\mathbf{X}} = \mathbf{V}\mathbf{\Sigma}^{1/2}\mathbf{U}'. \tag{5}$$

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

From the SVD (5) the $p \times IJ$ matrix $\mathbf{V}$ can be obtained as $\mathbf{V} = \widetilde{\mathbf{X}}\mathbf{U}\mathbf{\Sigma}^{-1/2}$. The actual calculations can be performed on the slices of the partitioned matrix $\widetilde{\mathbf{X}}$ as $\mathbf{V}^m = \widetilde{\mathbf{X}}^m\mathbf{U}\mathbf{\Sigma}^{-1/2}, m = 1; M$. The concatenated slices $[(\mathbf{V}^1)'|(\mathbf{V}^2)'|\ldots|(\mathbf{V}^M)']'$ form the matrix of the left singular vectors $\mathbf{V}'$. Therefore, the SVD (5) can be constructed with a sequential access to the data $\widetilde{\mathbf{X}}$ with p-linear effort.

After obtaining the SVD decomposition of $\widetilde{\mathbf{X}}$ the visit-specific $p \times I$ dimensional matrix $\widetilde{\mathbf{X}}_j$ can be represented as $\widetilde{\mathbf{X}}_j = \mathbf{V}\mathbf{\Sigma}^{1/2}\mathbf{U}_j$, where $\mathbf{U}_j$ is the jth block of the

matrix $\mathbf{U}' = (\mathbf{U}_1, \ldots, \mathbf{U}_J)$ corresponding to the matrix $\widetilde{\mathbf{X}}_j$. Note that the $p \times I$ matrices $\widetilde{\mathbf{X}}_j$ differ only via the matrix factors $\mathbf{U}_j$ of dimension $IJ \times I$. Thus, the high dimensional covariance operators $\hat{\mathbf{K}}_B$ and $\hat{\mathbf{K}}_W$ can be expressed as

$$\hat{\mathbf{K}}_B = \mathbf{V}\boldsymbol{\Sigma}^{1/2}\hat{\mathbf{K}}_B^{\mathbf{U}}\boldsymbol{\Sigma}^{1/2}\mathbf{V}' \quad \text{and} \quad \hat{\mathbf{K}}_W = \mathbf{V}\boldsymbol{\Sigma}^{1/2}\hat{\mathbf{K}}_W^{\mathbf{U}}\boldsymbol{\Sigma}^{1/2}\mathbf{V}', \tag{6}$$

where the inner matrices $\hat{\mathbf{K}}_W^U$ and $\hat{\mathbf{K}}_B^U$ are

$$\hat{\mathbf{K}}_B^{\mathbf{U}} = \frac{1}{I(J^2-J)} \sum_{j_1 \neq j_2} \mathbf{U}_{j_1}\mathbf{U}_{j_2}' \text{ and } \hat{\mathbf{K}}_W^{\mathbf{U}} = \frac{1}{2I(J^2-J)} \sum_{j_1 \neq j_2} (\mathbf{U}_{j_1} - \mathbf{U}_{j_2})(\mathbf{U}_{j_1} - \mathbf{U}_{j_2})'. \tag{7}$$

Equation (6) establishes the correspondence between high-dimensional operators, $\hat{\mathbf{K}}_B$ and $\hat{\mathbf{K}}_W$, and their low-dimensional counterparts, $\hat{\mathbf{K}}_B^U$ and $\hat{\mathbf{K}}_W^U$, respectively through the "sandwich" operator $\mathbf{V}\boldsymbol{\Sigma}^{1/2}\,[\,\cdot\,]\,\boldsymbol{\Sigma}^{1/2}\mathbf{V}'$. These crucial identities can be used to obtain the spectral decompositions $\hat{\mathbf{K}}_B = \boldsymbol{\Phi}_B\boldsymbol{\Lambda}_B\boldsymbol{\Phi}_B'$ and $\hat{\mathbf{K}}_W = \boldsymbol{\Phi}_W\boldsymbol{\Lambda}_W\boldsymbol{\Phi}_W'$ through the spectral decompositions of the low-dimensional matrices $\boldsymbol{\Sigma}^{1/2}\hat{\mathbf{K}}_B^{\mathbf{U}}\boldsymbol{\Sigma}^{1/2}$ and $\boldsymbol{\Sigma}^{1/2}\hat{\mathbf{K}}_W^{\mathbf{U}}\boldsymbol{\Sigma}^{1/2}$.

Indeed, using the spectral decompositions $\boldsymbol{\Sigma}^{1/2}\hat{\mathbf{K}}_B^{\mathbf{U}}\boldsymbol{\Sigma}^{1/2} = \mathbf{A}_B\boldsymbol{\Sigma}_B\mathbf{A}_B'$ and $\boldsymbol{\Sigma}^{1/2}\hat{\mathbf{K}}_W^{\mathbf{U}}\boldsymbol{\Sigma}^{1/2} = \mathbf{A}_W\boldsymbol{\Sigma}_W\mathbf{A}_W'$, where matrices $\mathbf{A}_B$ and $\mathbf{A}_W$ are orthogonal of size $IJ \times IJ$ and matrices $\boldsymbol{\Sigma}_B$ and $\boldsymbol{\Sigma}_W$ are diagonal, the covariance operators (6) can be written as

$$\hat{\mathbf{K}}_B = (\mathbf{V}\mathbf{A}_B)\boldsymbol{\Sigma}_B(\mathbf{V}\mathbf{A}_B)' \quad \text{and} \quad \hat{\mathbf{K}}_W = (\mathbf{V}\mathbf{A}_W)\boldsymbol{\Sigma}_W(\mathbf{V}\mathbf{A}_W)'. \tag{8}$$

The matrices $\mathbf{A}_B$ and $\mathbf{A}_W$ are orthogonal of size $IJ \times IJ$. This implies that the $p \times IJ$ matrices $\mathbf{V}\mathbf{A}_B$ and $\mathbf{V}\mathbf{A}_W$ have orthonormal columns. From the uniqueness of the spectral decomposition of a symmetric matrix it follows that (8) is an alternative representation of the spectral decomposition of the covariance operators. It implies

that

$$\mathbf{\Phi}_B = \mathbf{V}\mathbf{A}_B,\ \mathbf{\Phi}_W = \mathbf{V}\mathbf{A}_W \quad \text{and} \quad \mathbf{\Lambda}_B = \mathbf{\Sigma}_B,\ \mathbf{\Lambda}_W = \mathbf{\Sigma}_W \tag{9}$$

Therefore, diagonalizing the operators $\hat{\mathbf{K}}_B$ and $\hat{\mathbf{K}}_W$ requires a few simple steps that can be performed in p-linear time. Algorithm 1 summarizes these steps in a series of simple actions that can be implemented using linear number of calculations in the dimensionality of the problem and sequential access to the data. The entire data set cannot and is not loaded into computer memory.

Algorithm 1: Computing eigendecompositions in p-linear time.

Input: Matrix $\widetilde{\mathbf{X}} = (\widetilde{\mathbf{X}}_1, \ldots, \widetilde{\mathbf{X}}_J)$ of size $p \times IJ$.

Output: $(\phi_k^{(1)}, \lambda_k^{(1)}), k = 1, \ldots, IJ$ and $(\phi_l^{(2)}, \lambda_l^{(2)}), l = 1, \ldots, IJ$.

begin

Step 1. Construct the spectral decomposition of the symmetric matrix $\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}' = \mathbf{U}\mathbf{\Sigma}\mathbf{U}'$.

Step 2. Construct the spectral decomposition of the symmetric semi-positive definite matrix $\mathbf{\Sigma}^{1/2}\hat{\mathbf{K}}_W^{\mathbf{U}}\mathbf{\Sigma}^{1/2} = \mathbf{A}_W\mathbf{\Sigma}_W\mathbf{A}_W'$.

Step 3. Construct the spectral decomposition of the symmetric matrix $\mathbf{\Sigma}^{1/2}\hat{\mathbf{K}}_B^{\mathbf{U}}\mathbf{\Sigma}^{1/2} = \mathbf{A}_B\mathbf{\Sigma}_B\mathbf{A}_B'$.

Step 4. Calculate $\mathbf{V}\mathbf{A}_B = \widetilde{\mathbf{X}}\mathbf{U}\mathbf{\Sigma}^{-1/2}\mathbf{A}_B$ and $\mathbf{V}\mathbf{A}_W = \widetilde{\mathbf{X}}\mathbf{U}\mathbf{\Sigma}^{-1/2}\mathbf{A}_W$.

Result: ($\mathbf{V}\mathbf{A}_B$ and $\mathbf{\Sigma}_B$) and ($\mathbf{V}\mathbf{A}_W$ and $\mathbf{\Sigma}_W$) contain first and second level eigenvectors and eigenvalues, respectively.

end

Note that the low-dimensional matrices $\mathbf{A}_B$ and $\mathbf{A}_W$ completely determine the geometry of the high-dimensional "between" and "within" spaces as well as the interaction between the two. In the next section, we provide more insight on the matrices $\mathbf{A}_B$ and $\mathbf{A}_W$. In particular, we show that $\mathbf{A}_B$ and $\mathbf{A}_W$ paired with the right singular vectors $\mathbf{U}_j$ provide all the necessary information for estimating the principal scores

of model (3).

3.2 Principal scores

In this section we will show how the BLUP for scores in the model (3) can be obtained without using high-dimensional matrices or vectors. This is essential as a brute force extension of methods in (Crainiceanu et al., 2009) and (Di and Crainiceanu, 2010) would require inversion of $p \times p$ dimensional matrices. In this section we propose a completely different approach to calculating the BLUPS that is computationally feasible for samples of massive images; new insights into the MFPCA geometry are obtained as a by-product.

We introduce some notation first. The subject level principal scores are $\boldsymbol{\xi}_i = (\xi_{i1}, \ldots, \xi_{iN_1})'$, $\boldsymbol{\zeta}_i = (\boldsymbol{\zeta}_{i1}, \ldots, \boldsymbol{\zeta}_{iJ})'$, where $\boldsymbol{\zeta}_{ij} = (\zeta_{ij1}, \ldots, \zeta_{ijN_2})$. We split the $p \times IJ$ dimensional matrix $\widetilde{\mathbf{X}}$ into the $p \times J$-dimensional blocks $\widetilde{\mathbf{X}}_i$, $i = 1, \ldots, I$, corresponding to subjects. The jth column of the $\widetilde{\mathbf{X}}_i$ matrix contains data for visit j of subject i. We refer to this as the by-subject partition to distinguish it from the by-visit partition of the matrix $\widetilde{\mathbf{X}}$ in Section 3.1. After rearranging the columns we denote $\widetilde{\mathbf{X}} = (\widetilde{\mathbf{X}}_1, \ldots, \widetilde{\mathbf{X}}_I)$. As in Section 3.1, the SVD of the matrix $\widetilde{\mathbf{X}}$ can be written in by-subject blocks as $\widetilde{\mathbf{X}}_i = \mathbf{V}\boldsymbol{\Sigma}^{1/2}\mathbf{U}_i'$, where the $IJ \times J$ matrix $\mathbf{U}_i'$ corresponds to subject grouping. The following theorem contains the main result in this section; in particular it provides a simple and fast recipe for calculating the BLUPs for the MFPCA model in the context of high dimensional data.

Theorem 1: *Under the MFPCA model (3), the estimated best linear unbiased predictor (EBLUP) of $\boldsymbol{\xi}_i$ and $\boldsymbol{\zeta}_i$ is given by*

$$\begin{pmatrix} \hat{\boldsymbol{\xi}}_i \\ \hat{\boldsymbol{\zeta}}_i \end{pmatrix} = \begin{pmatrix} J \cdot \mathbf{I}_{N_1} & \mathbf{1}_J' \otimes \mathbf{C}_{BW} \\ \mathbf{1}_J \otimes \mathbf{C}_{BW}' & \mathbf{I}_J \otimes \mathbf{I}_{N_2} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{A}_B^{N_1} \boldsymbol{\Sigma}^{1/2} \mathbf{U}_i' \mathbf{1}_J \\ vec(\mathbf{A}_W^{N_2} \boldsymbol{\Sigma}^{1/2} \mathbf{U}_i') \end{pmatrix}, \tag{10}$$

where $\mathbf{C}_{BW} = \mathbf{A}_B^{N_1}(\mathbf{A}_W^{N_2})'$ *is of size* $N_1 \times N_2$*, the matrix* $\mathbf{A}_B^{N_1}$ *is of size* $N_1 \times IJ$ *and its rows are the first* N_1 *columns of the matrix* $\mathbf{A}_B$*. The* $N_2 \times IJ$ *matrix* $\mathbf{A}_W^{N_2}$ *is defined in a similar way. Vector* $\mathbf{1}_J$ *is a* $J \times 1$ *vector of ones,* $\otimes$ *is the Kronecker product of matrices, and operation* $vec(\cdot)$ *stacks the columns of a matrix on the top of each other.*

The fundamental property of the EBLUP is that all the matrices involved in (10) are low-dimensional and do not depend on the dimension p. Therefore, the EBLUP calculations are almost instantaneous. Careful inspection of the results in Theorem 1 provides novel insights into single and multilevel functional principal component analysis. Indeed, the matrices $\mathbf{A}_B^{N_1}$ and $\mathbf{A}_W^{N_2}$ are computed using only the right singular vectors in the equation (5). Thus, all the necessary information for estimating the scores is projected onto the low-dimensional space spanned by the vectors orthogonal to the right singular vectors. The BLUPs (10) provide a unique geometrical insight. For illustration, assume that "between" and "within" eigenvectors are orthogonal, that is $\mathbf{C}_{BW} = 0$. Then $\hat{\boldsymbol{\xi}}_i = \mathbf{A}_B^{N_1}\boldsymbol{\Sigma}^{1/2}\mathbf{U}_i'\mathbf{1}_J/J$ and $\hat{\boldsymbol{\zeta}}_i = vec(\mathbf{A}_W^{N_2}\boldsymbol{\Sigma}^{1/2}\mathbf{U}_i')$. The ith block $\mathbf{U}_i'$ of size $IJ \times J$ is composed of the IJ directions each of which is of size J. The product $\boldsymbol{\Sigma}^{1/2}\mathbf{U}_i'\mathbf{1}_J/J$ results in averaging over each of the directions and scaling those averages by the singular values. The matrix $\mathbf{A}_B^{N_1}$ then defines the BLUPs of the "between" scores by calculating N_1 orthonormal linear combinations of the values of $\boldsymbol{\Sigma}^{1/2}\mathbf{U}_i'\mathbf{1}_J/J$. If we set $J = 1$ the MFPCA model (3) reduces to the single-level functional model obtained in (Zipunnikov et al., 2010). In this case, there is no "within" level, $\mathbf{A}_B = \mathbf{I}_I$ and the BLUPs of the "between" scores are $\hat{\boldsymbol{\xi}}_i = (\boldsymbol{\Sigma}^{N_1})^{1/2}(\mathbf{U}_i^{N_1})'$ exactly the ones obtained in (Zipunnikov et al., 2010). Similarly, the matrix $\mathbf{A}_W^{N_2}$ defines the BLUPs of the "within" scores $\hat{\boldsymbol{\zeta}_{ij}} = \mathbf{A}_W^{N_2}\boldsymbol{\Sigma}^{1/2}\mathbf{U}_{ij}'$ as the N_2 orthonormal combinations of the coordinates of the jth column of the block $\mathbf{U}_i'$ weighted by the singular values. For non-orthogonal "between" and "within" bases the interaction

matrix $\mathbf{C}_{BW}$ has to be taken into account. This generalizes the results of (Zipunnikov et al., 2010) and unifies the theory of single and multilevel functional principal component analysis for low, moderate and high dimensional data.

4 Simulations

In this section, we will illustrate our HD-MFPCA method with two simulation studies. The first one replicates the simulation scenario in (Di et al., 2008) and shows the equivalence of the two approaches in low and moderate dimensional applications. The second one tests how well spatial bases are recovered by HD-MFPCA in an application where the approach of (Di et al., 2008) cannot be implemented. Both studies were run on a four core i7-2.67Gz PC and 6Gb of RAM memory.

First, we start with the scenario considered by (Di et al., 2008). Data are generated according to the following model

$$\begin{cases} X_{ij}(v) = \sum_{k=1}^{N_1} \xi_{ik}\phi_k^{(1)}(v) + \sum_{l=1}^{N_2} \zeta_{ijl}\phi_l^{(2)}(v), & v \in \mathcal{V} \\ \xi_{ik} \overset{i.i.d.}{\sim} N(0, \lambda_k^{(1)}) \quad \text{and} \quad \zeta_{ijl} \overset{i.i.d.}{\sim} N(0, \lambda_l^{(2)}), \end{cases}$$

where ξ_{ik}'s and ζ_{ijl}'s are mutually independent. To match the RAVENS data design we slightly deviate from (Di et al., 2008) and choose $I = 350$ and $J = 2$. We set the number of eigenfunctions at level 1, N_1, to 4 and at level 2, N_2, to 4. The true eigenvalues at level 1 and level 2 are the same, $\lambda_k^{(1)} = \lambda_k^{(2)} = 0.5^{k-1}, k = 1, 2, 3, 4$. We follow Case 2 of Section 4 in (Di et al., 2008) and choose the following non-orthogonal bases

Level 1: $\{ \sqrt{2}\sin(2\pi v), \sqrt{2}\cos(2\pi v), \sqrt{2}\sin(4\pi v), \sqrt{2}\cos(4\pi v) \}$,

Level 2: $\{ 1, \sqrt{3}(2v-1), \sqrt{5}(6v^2-6v+1), \sqrt{7}(20v^3-30v^2+12v-1) \}$,

which are measured on a regular grid of p equidistant points of interval $[0, 1]$, $\mathcal{V} =$

$\{1/p, 2/p, 3/p, \ldots, 1\}$ with $p = 50,000$. Note that a brute-force extension of standard MFPCA would be infeasible for such a large p even for one data set. HD-MFPCA allows to easily handle such a large dimension and conduct a simulation study in 27 minutes. The simulation study was implemented in Matlab 2010a and the software is available upon request.

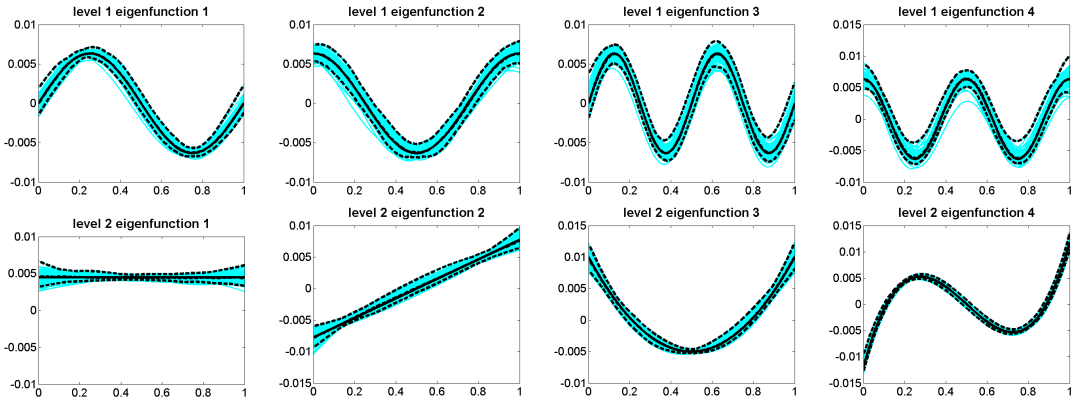

Figure 1: *True and estimated eigenfunctions* $\phi_k^{(1)}$ *and* $\phi_l^{(2)}$ *for scenario* 1 *replicated* 100 *times. Each box shows the estimated functions (cyan solid lines), a true function (solid black line), the pointwise median and the* 5^{th} *and* 95^{th} *pointwise percentile curves (dashed black lines).*

Figure 1 displays the true and estimated eigenfunctions using the HD-MFPCA approach at levels 1 and 2, respectively. The results for $\phi_k^{(1)}$ are displayed in the top panel and $\phi_l^{(2)}$ are shown in the bottom panel. Shown are the true function (solid black line), the estimated functions (cyan lines), the pointwise median of estimated eigenvectors (indistinguishable from the true functions) and the pointwise 5th and 95th percentiles of estimated eigenvectors (black dashed lines). Comparing Figure 1 with Figure 5 in (Di et al., 2008) we conclude that our estimation procedure completely reproduce the eigenfunction results obtained using the standard MFPCA approach.

Figure 2 shows the boxplots of the estimated level 1 and level 2 eigenvalues. We display the centered and standardized eigenvalues, $(\hat{\lambda}_k^{(1)} - \lambda_k^{(1)})/\lambda_k^{(1)}$ and $(\hat{\lambda}_l^{(2)} - \lambda_l^{(2)})/\lambda_l^{(2)}$, respectively. The results confirm the very good performance of the estima-

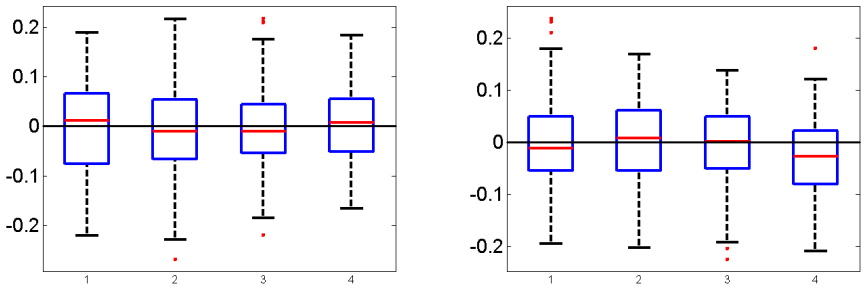

Figure 2: *Boxplots of the normalized estimated level* 1 *eigenvalues,* $(\hat{\lambda}_k^{(1)} - \lambda_k^{(1)})/\lambda_k^{(1)}$*, (left box) and the normalized estimated level* 2 *eigenvalues,* $(\hat{\lambda}_l^{(2)} - \lambda_l^{(2)})/\lambda_l^{(2)}$*, (right box) based on scenario* 1 *with* 100 *replications. The zero is shown by the solid black line.*

tion procedure and are consistent with those reported for the no noise case in Figure 4 of (Di et al., 2008).

To estimate the scores ξ_{ik} and ζ_{ijl}, we use the EBLUP results of Theorem 1. Each replication of the scenario provides 350 estimates of level 1 scores, $\xi_{ik}, k = 1, 2, 3, 4$, and 700 estimates of level 2 scores, $\zeta_{ijl}, l = 1, 2, 3, 4$. The total number of the scores ξ_{ik} estimated in the study is $35,000$ for each k. Similarly, the total number of the estimated scores ζ_{ijl} is $70,000$ for each l. However, the estimated scores within each replication are dependent even if they are a-priori independent. The boxplots of the normalized estimated scores $(\xi_{ik} - \hat{\xi}_{ik})/\sqrt{\lambda_k^{(1)}}$ and $(\zeta_{ijl} - \hat{\zeta}_{ijl})/\sqrt{\lambda_l^{(2)}}$ are reported in the left panels of Figure 3, respectively. The distribution of the normalized estimated scores corresponding to the fourth eigenfunction at level 1, $\phi_4^{(1)}$, has a slightly wider spread around zero; this is likely due to the smaller signal/noise ratio. The right panels of Figure 3 display the medians, 0.5%, 5%, 90% and 99.5% quantiles of the distribution of the normalized estimated scores. These plots confirm the theoretical results of Theorem 1 stating that the EBLUPs given by equation (10) are unbiased. This demonstrates that the estimation procedures are unbiased for both approaches. To summarize, the first simulation study showed that HD-MFPCA approach replicates

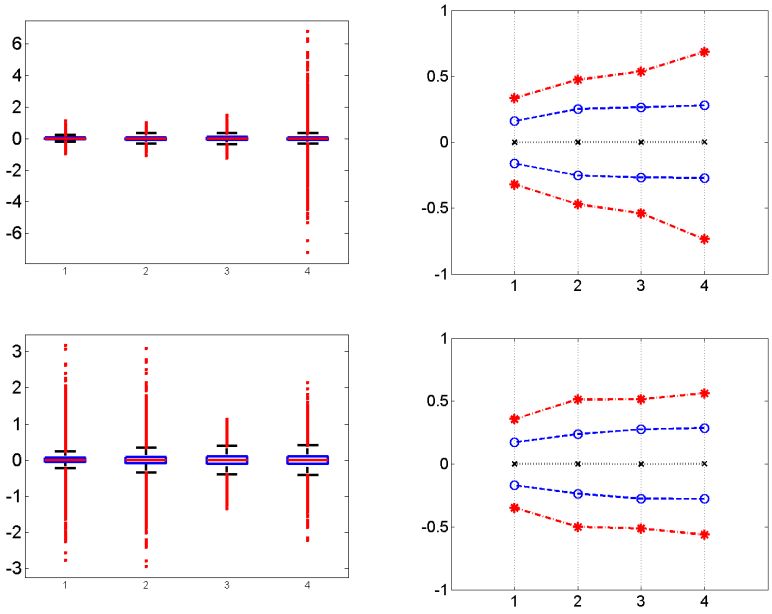

Figure 3: *Top row shows the distribution of the normalized estimated level* 1 *scores,* $(\xi_{ik} - \hat{\xi}_{ik})/\sqrt{\lambda_k^{(1)}}$. *Boxplots are given in left column. The right column shows the medians (black marker),* 5% *and* 95% *quantiles (blue markers), and* 0.5% *and* 99.5% *quantiles (red markers). Similarly, the distribution of the normalized estimated level* 2 *scores,* $(\zeta_{ijl} - \hat{\zeta}_{ijl})/\sqrt{\lambda_l^{(2)}}$ *is provided at the bottom row.*

the results of the standard MFPCA but has the key advantage of being able to efficiently handle arbitrarily high-dimensional multilevel functional data.

In the second scenario, we considered the extension of our methods to 2D bases. We simulated 100 data sets from the model

$$\begin{cases} X_{ij}(v) = \sum_{k=1}^{3} \xi_{ik}\phi_k^{(1)}(v) + \sum_{l=1}^{2} \zeta_{ijl}\phi_l^{(2)}(v), & v \in \mathcal{V} \\ \xi_{ik} \overset{i.i.d.}{\sim} N(0, \lambda_k^{(1)}) \quad \text{and} \quad \zeta_{ijl} \overset{i.i.d.}{\sim} N(0, \lambda_l^{(2)}), \end{cases} \tag{11}$$

where $\mathcal{V} = [1, 300] \times [1, 300]$. Level 1 and 2 eigenbases, or eigenimages, are displayed at the top and bottom of Figure 4, respectively. The images in this scenario can be thought of as 2D grayscale images with pixel intensities on the $[0, 1]$ scale. The black pixels are set to 1 and the white ones are set to zero. Eigenimages are uncorrelated

within the same level but are correlated across levels. We assume that $I = 350$, $J = 2$, and the true eigenvalues at level 1, $\lambda_k^{(1)} = 0.5^{k-1}, k = 1, 2, 3$, and the ones at level 2, $\lambda_l^{(2)} = 0.5^{l-1}, l = 1, 2$. To apply HD-MFPCA we unfold each image X_{ij} and obtain vectors of size $p = 300 \times 300 = 90,000$. The simulation study took 32 minutes.

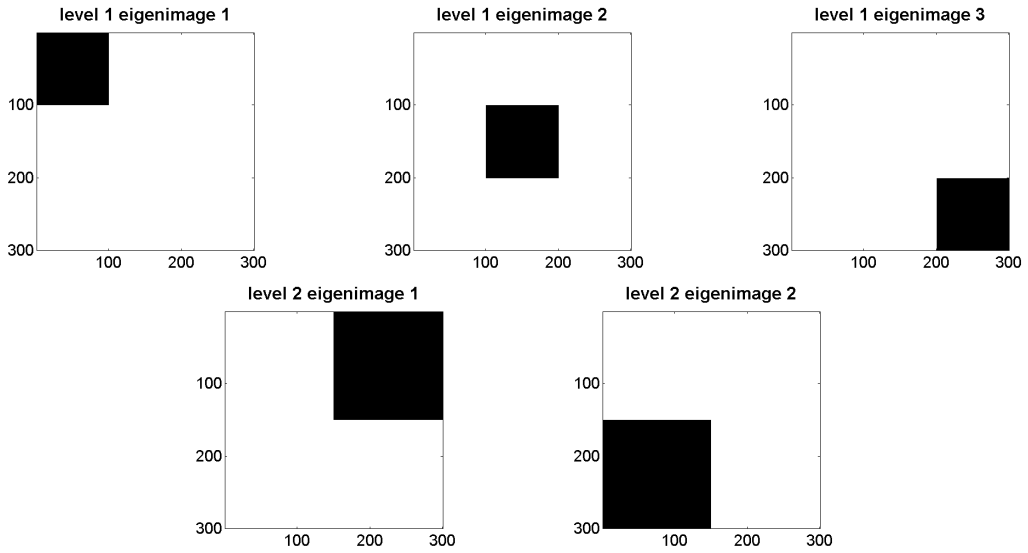

Figure 4: *Level 1 (top row) and Level 2 (bottom row) grayscale eigenimages for the 2nd scenario.*

Figures 5, 6, and 7 display the mean of the estimated eigenimages and the pointwise 5th and 95th percentile images, respectively. To obtain a grayscale image with pixel values in the $[0, 1]$ interval, each estimated eigenvector, $\hat{\boldsymbol{\phi}} = (\hat{\phi}_1, \ldots, \hat{\phi}_p)$, was normalized as $\hat{\boldsymbol{\phi}} \rightarrow (\hat{\boldsymbol{\phi}} - \min_s \hat{\phi}_s)/(\max_s \hat{\phi}_s - \min_s \hat{\phi}_s)$. Figure 5 displays how on average our method recovers the spatial configuration of both bases. The percentile images in Figures 6 and 7 show a similar pattern as the average with small distortions from the true functions (please note the light gray areas). To better understand the quantile patterns consider as an example the behavior of $\hat{\phi}_1^{(2)}$. A closer inspection of $\hat{\mathbf{K}}_W$ and model (11) will reveal that an estimated eigenvector $\hat{\phi}_1^{(2)}$ is a linear combination of the true level 2 eigenvectors $\phi_1^{(2)}$ and $\phi_2^{(2)}$, with the weights being functions of

the level 2 eigenvalues and principal scores. The larger the sample size, the closer the estimated eigenfunction $\hat{\phi}_1^{(2)}$ to $\phi_1^{(2)}$. Large differences between the weights create the pattern observed in the 5% quantile plot of Figure 6. Small differences between the weights result in the pattern shown in the 95% quantile plot of Figure 7. Similarly, it can be shown that the estimates of the first level eigenvectors are linear combinations of both true level 1 and level 2 eigenvectors. This leads to the overlapping patterns shown in the top row panels of Figures 6 and 7. We conclude that the estimation of the 2D eigenimages is very good.

Figure 2 shows the boxplots of the estimated level 1 and level 2 eigenvalues. As before, we display the estimated normalized eigenvalues, $(\hat{\lambda}_k^{(1)} - \lambda_k^{(1)})/\lambda_k^{(1)}$ and $(\hat{\lambda}_l^{(2)} - \lambda_l^{(2)})/\lambda_l^{(2)}$, respectively. The results confirm the very good performance of the estimation procedure and are consistent with those reported for the no noise case in Figure 4 of (Di et al., 2008).

Figure 8 shows the boxplots of the estimated level 1 and level 2 eigenvalues. We plot the estimated normalized eigenvalues, $(\hat{\lambda}_k^{(1)} - \lambda_k^{(1)})/\lambda_k^{(1)}$ and $(\hat{\lambda}_l^{(2)} - \lambda_l^{(2)})/\lambda_l^{(2)}$, indicating that eigenvalues are estimated with essentially no bias. Similarly to Scenario 1, the total number of the estimated scores ξ_{ik} is $35,000$ for each k and there are $70,000$ estimated scores ζ_{ijl} for each l. The boxplots of the normalized estimated scores $(\xi_{ik} - \hat{\xi}_{ik})/\sqrt{\lambda_k^{(1)}}$ and $(\zeta_{ijl} - \hat{\zeta}_{ijl})/\sqrt{\lambda_l^{(2)}}$ are displayed in the left panels of Figure 9, respectively. Results show that the EBLUPs approximate true scores very well. The scores corresponding to the second and third eigenimages at level 1 have a slightly larger spread due to a reduced signal to noise ratio. The right panels of Figure 9 display the medians, 0.5%, 5%, 90% and 99.5% quantiles quantiles of the distribution of the normalized estimated scores.

5 The analysis of the RAVENS data

In this section we apply HD-MFPCA to RAVENS images discussed in Section 1.1. The RAVENS images are $256 \times 256 \times 198$ dimensional for 352 subjects, each with two visits. This is on the order of ten billion (10^{10}) numbers. Assuming 32 bits per number, this represents 34 gigabytes of data for the gray matter images alone. To emphasize processing issues, one must multiply this number by three to represent both tissue types (gray and white) and cerebro-spinal fluid and hence at least an additional one hundred gigabytes is required for every processing step saved. However, even restricting ourselves only to the processed gray matter data, the data matrix is too large (ten billion) to work with in statistical models without further restrictions; therefore, the following strategy was executed. First, after processing, the intersection of non-background voxels across images was collected. Such an intersection greatly reduces the dimension of the data matrix from ten billion numbers to two billion numbers divided as three million relevant voxels per subject per visit, and seven hundred and four subjects/visits. However, even this relevant intersection remains prohibitive for modern systems. Hence the data matrix, of size 704 by 3 million, was divided into 100 submatrices of size 704 by 30 thousand (ten million numbers each). In what follows, we demonstrate methods for performing all calculations in such a way that only one of the manageable submatrices of ten million needs to be stored in memory at any given moment. Note that on lower-resource computers the only change would be to reduce the size of subsets. A parallel computation implementation of our approach would be practically instantaneous. The entire analysis performed in Matlab 2010a took around 16 minutes on a PC with a quad core i7-2.67Gz processor and 6Gb of RAM memory.

A small technical concern was of a few artifactual negative values in the data from the preprocessing. These voxels were removed from the analysis. In addition,

negative eigenvalues occurred when calculating the spectral decomposition of the between covariance operator, $\hat{\mathbf{K}}_B$. Estimating a non-negative definite matrix, it is not necessarily non-negative definite itself (see Section 2.2 of Di et al. (2008) and (Hall et al., 2008)). The negative eigenvalues of $\hat{\mathbf{K}}_B$ represented 2.27 percent of the overall variation. Following Hall et al. (2008) all the negative eigenvalues with corresponding eigenvectors were trimmed to zero for the analysis.

In the analysis, we first estimated the mean and the visit effects by the methods of moments. The visit specific mean images are uniform over the template and simply convey the message that localized changes in morphometry within subgroups get averaged over in the visit-specific mean calclulations. In our eigenimage analysis we de-mean the data by subtracting out these vectors.

Next, the total variation was decomposed into "between" and "within" levels. Most of the total variability, 91.42%, is explained by the level 1 (between-subject) variability while 8.58% is explained by level 2 (within-subject between visit) variability. This is expected, but nonetheless, scientifically interesting. That is, the majority of variation in brain shape in this large sample is cross-sectional across subjects rather than longitudinal. A big contribution of our approach is to actually quantify these differences. It would be of interest to see how the ratio of between to within variation changes with the age of the sample. For example, we would expect the proportion of variability explained longitudinally to be much higher when comparing infants and their first five years of life or very elderly subjects, where there is much more longitudinal change.

Table 1 provides the percentages of explained variability by the first 10 eigenimages for between and within levels, respectively. We also plot the the percentages explained by the first 30 between and within eigenimages on Figure 10. First 30 between eigenimages explain roughly 47.13% of the between variability. For the within level,

Figure 10 indicates that 52.52% of the within variability is explained by the first 30 within eigenimages.

The 3D-renderings of the first three estimated between eigenimages at five different angles are shown on Figure 11. Similar renderings are presented for the within level on Figure 12. Figures 13 and 14 show eleven equidistant axial slices of the first three estimated between and within eigenimages, respectively.

We interpret the between eigenimages as follows. The first eigenimage loads positively on the whole cortex and has a small negative region around the ventricles. This clearly represents overall brain size and shape. A person loading positively on this eigenimage will have a larger brain. This component of variation is expected, simply due to variation in head sizes within the sample. The second loads positively on the majority of the cortex and negatively on the temporal lobe and cerebellum. The third loads negatively on the cerebellum. It is extremely encouraging, that the eigenimages obey regional boundaries fairly strictly, despite the algorithm's complete ignorance of spatial relationships. We believe that this is the first characterization of cross-sectional brain morphometry using VBM data.

The within eigenimages are more difficult to interpret. They appear to load positively and negatively on layers of the cortex and no clear anatomical boundaries are seen. Moreover, they appear to span large areas of the brain. We interpret this as suggesting that the most common form of longitudinal brain aging is a uniform volumetric loss. Such effects could also be seen if there are localized volume losses, yet the locations are random across subjects. It is also important to note that a longer follow up time, or investigation different populations might display more longitudinal variation. Moreover, specific diseases, such as Alzheimer or multiple sclerosis, may differentially affect regional brain loss across time. Our method is designed to identify and quantify such potentially interesting processes.

6 Discussion

Longitudinal observational studies that collect high-dimensional functional measurements at multiple visits appear in an increasing number. These studies create a great demand for methods that allow the separation of the observed functional variability into cross-sectional and longitudinal components. However, high dimensionality and size of the observed data pose enormous computational challenges for existing methods of multilevel analysis. We have developed a powerful and computationally efficient solution responding to these challenges, HD-MFPCA. The conducted simulation studies confirmed the high potency of the methodology to recover the spatial configurations in high-dimensional subject-specific and subject/visit-specific spaces. The approach was applied to analyze a large imaging study of 704 RAVENS images containing over ten billion measurements. The high-dimensional multilevel methodology developed in this paper can be applied equally to any longitudinal study with the assumed design.

A potential limitation of our approach is the need to have the total number of functional observations not prohibitively large. In particular, the number of subjects, I, and the number of observations per subject, J, must be of moderate size to be able to perform a spectral decomposition of a $IJ \times IJ$ matrix. Our methods are designed for IJ of the order of $10,000$ to $15,000$ and arbitrarily large p. Although, to the best of our knowledge, there are currently no longitudinal imaging studies of such a size, it is not difficult to predict a new generation of observational studies where the number of subjects reaches tens of thousands with dozens or hundreds of observations per subject. In such a situation, a possible alternative to our approach would be an adaptive aggregation of the eigen-decompositions calculated on subsamples of the subjects. Efficient ways of aggregation remain an open problem and will need to be addressed in future. Another limitation is that our methods assume

a balanced design. Motivated by RAVENS data which represent preprocessed and smoothed images we have not assumed noise in the model. However, one can easily think of applications where functional observations are measured with non-ignorable noise. Another possible scenario is the sparsity of the high-dimensional functional observations. These situations have been considered in (Di et al., 2008) and (Di and Crainiceanu, 2010), respectively, and efficient solutions have been proposed. However, these solutions require smoothing of the covariance operators which is infeasible for high-dimensional data. Thus, a computationally efficient procedure of covariance smoothing or its equivalent is highly desirable to address these problems in the high dimensional context.

Acknowledgment

The research of Vadim Zipunnikov, Brian Caffo and Ciprian Crainiceanu was supported by grant R01NS060910 from the National Institute of Neurological Disorders and Stroke.

7 References

References

Ashburner, J. and K. Friston (2000). Voxel-based morphometry - the methods. *Neuroimage 11*(6), 805–821.

Caffo, B., S. Chen, W. Stewart, K. Bolla, D. Yousem, C. Davatzikos, and B. S. Schwartz (2008). Are brain volumes based on magnetic resonance imaging mediators of the associations of cumulative lead dose with cognitive function? *American Journal of Epidemiology 167(4)*, 429–437.

Chen, S., C. Wang, L. Eberly, B. Caffo, and B. Schwartz (2009). Adaptive control of the false discovery rate in voxel-based morphometry. *Human Brain Mapping 30*(7).

Crainiceanu, C. M., B. S. Caffo, S. Luo, V. V. Zipunnikov, and N. M. Punjabi (2010). Statistical analysis of population of images. *Submitted*.

Crainiceanu, C. M., A.-M. Staicu, and C.-Z. Di (2009). Generalized multilevel functional regression. *Journal of the American Statistical Association 104, 488*, 1550–1561.

Davatzikos, C., A. Genc, D. Xu, and S. Resnick (2001). Voxel-based morphometry using the ravens maps: methods and validation using simulated longitudinal atrophy. *NeuroImage 14*(6), 1361–1369.

Di, C.-Z. and C. M. Crainiceanu (2010). Multilevel sparse functional principal component analysis. *Technical Report*.

Di, C.-Z., C. M. Crainiceanu, B. S. Caffo, and N. M. Punjabi (2008). Multilevel functional principal component analysis. *Annals of Applied Statistics 3(1)*, 458–488.

Greven, S., C. M. Crainiceanu, B. S. Caffo, and D. Reich (2009). Longitudinal functional principal component analysis. *Submitted*.

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

Karhunen, K. (1947). Uber lineare methoden in der wahrscheinlichkeitsrechnung. *Annales Academie Scientiarum Fennicae 37*, 1–79.

Mazziotta, J., A. Toga, A. Evans, P. Fox, and J. Lancaster (1995). A probabilistic atlas of the human brain: theory and rationale for its development the international consortium for brain mapping (ICBM). *Neuroimage 2*(2PA), 89–101.

Mohamed, A. and C. Davatzikos (2004). Shape representation via best orthogonal basis selection. *Medical Image Computing and Computer-Assisted Intervention 3216*, 225–233.

Reiss, P. and R. T. Ogden (2008). Functional generalized linear models with applications to neuroimaging. *Philip T. Reiss*, 10.

Reiss, P. and R. T. Ogden (2010). Functional generalized linear models with images as predictors. *Biometrics 66.1*, 61–69.

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

Schwartz, B., S. Chen, B. Caffo, W. Stewart, K. Bolla, D. Yousem, and C. Davatzikos (2007). Relations of brain volumes with cognitive function in males 45 years and older with past lead exposure. *Neuroimage 37*(2), 633–641.

Schwartz, B., W. Stewart, K. Bolla, D. Simon, K. Bandeen-Roche, B. Gordon, J. Links, and A. Todd (2000). Past adult lead exposure is associated with longitudinal decline in cognitive function. *Neurology 55*(8), 1144.

Schwartz, B., W. Stewart, K. Kelsey, D. Simon, S. Park, J. Links, and A. Todd (2000). Associations of tibial lead levels with BsmI polymorphisms in the vitamin D receptor in former organolead manufacturing workers. *Environmental Health Perspectives 108*(3), 199.

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

Stewart, W., B. Schwartz, D. Simon, K. Bolla, A. Todd, and J. Links (1999). Neurobehavioral function and tibial and chelatable lead levels in 543 former organolead workers. *Neurology 52*(8), 1610.

Su, S., B. S. Caffo, L. E. Eberly, E. Garrett-Mayer, W. F. Stewart, S. Chen, D. Yousem, C. Davatzikos, and B. Schwartz (2008). On the merits of voxel-based morthometric path-analysis for investigating volumetric mediation of a toxicant's influence on cognitive function. *Johns Hopkins University, Department of Biostatistics Working Papers 160.*

Zipunnikov, V., B. Caffo, C. M. Crainiceanu, D. M. Yousem, C. Davatzikos, and B. S. Schwartz (2010). Functional principal component models for high dimensional brain volumetrics. *Johns Hopkins University, Working paper*.

8 Appendix.

Proof of Theorem 1. We assume $\mu = 0$ and $\eta_j = 0$ for all $j = 1, \ldots, J$. Model (3) can be written in a vector form as $\widetilde{\mathbf{X}}_i = \mathbf{A}\mathbf{u}_i, i = 1, \ldots, I$, where $\widetilde{\mathbf{X}}_i = (\widetilde{\mathbf{X}}_{i1}^{'}, \ldots, \widetilde{\mathbf{X}}_{iJ}^{'})^{'}$ is a pJ-dimensional vector, principal scores $\mathbf{u}_i = (\boldsymbol{\xi}_i^{'}, \boldsymbol{\zeta}_i^{'})^{'}$ and $p \times (N_1 + JN_2)$ partitioned matrix $\mathbf{A} = [\mathbf{A}_B | \mathbf{A}_W]$ with $\mathbf{A}_B = \mathbf{1}_J \otimes \boldsymbol{\Phi}_B$ and $\mathbf{A}_W = \mathbf{I}_J \otimes \boldsymbol{\Phi}_W$. From (3) it follows that $\mathbf{u}_i \overset{i.i.d.}{\sim} (0, \boldsymbol{\Lambda}_u)$ where $\boldsymbol{\Lambda}_u = \begin{pmatrix} \boldsymbol{\Lambda}_B & 0 \\ 0 & \mathbf{I}_J \otimes \boldsymbol{\Lambda}_W \end{pmatrix}$. If $p \leq N_1 + JN_2$ the BLUP of $\mathbf{u}_i$ is given by $\hat{\mathbf{u}}_i = Cov(\mathbf{u}_i, \widetilde{\mathbf{X}}_i) Var(\widetilde{\mathbf{X}}_i)^{-1} \widetilde{\mathbf{X}}_i = \boldsymbol{\Lambda}_u \mathbf{A}^{'} (\mathbf{A}\boldsymbol{\Lambda}_u \mathbf{A}^{'})^{-1} \widetilde{\mathbf{X}}_i$. When $p > N_1 + JN_2$ matrix $\mathbf{A}\boldsymbol{\Lambda}_u\mathbf{A}^{'}$ is not invertible and the generalized inverse of $\mathbf{A}\boldsymbol{\Lambda}_u\mathbf{A}^{'}$ is used. In that case, $\hat{\mathbf{u}}_i = \boldsymbol{\Lambda}_u \mathbf{A}^{'} (\mathbf{A}\boldsymbol{\Lambda}_u \mathbf{A}^{'})^{-} \widetilde{\mathbf{X}}_i = \boldsymbol{\Lambda}_u^{1/2} (\boldsymbol{\Lambda}_u^{1/2} \mathbf{A}^{'} \mathbf{A} \boldsymbol{\Lambda}_u^{1/2})^{-1} \boldsymbol{\Lambda}_u^{1/2} \mathbf{A}^{'} \widetilde{\mathbf{X}}_i =$

$\hat{\mathbf{u}}_i = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\widetilde{\mathbf{X}}_i$. Estimated BLUPs use (5) and (9) which gives

$$\mathbf{A}'\mathbf{A} = \begin{pmatrix} J \cdot \mathbf{I}_{N_1} & \mathbf{1}_J' \otimes \mathbf{C}_{BW} \\ \mathbf{1}_J \otimes \mathbf{C}_{BW}' & \mathbf{I}_J \otimes \mathbf{I}_{N_2} \end{pmatrix} \quad \text{and} \quad \mathbf{A}'\widetilde{\mathbf{X}}_i = \begin{pmatrix} \mathbf{A}_B^{N_1}\boldsymbol{\Sigma}^{1/2}\mathbf{U}_i'\mathbf{1}_J \\ vec(\mathbf{A}_W^{N_2}\boldsymbol{\Sigma}^{1/2}\mathbf{U}_i') \end{pmatrix}.$$

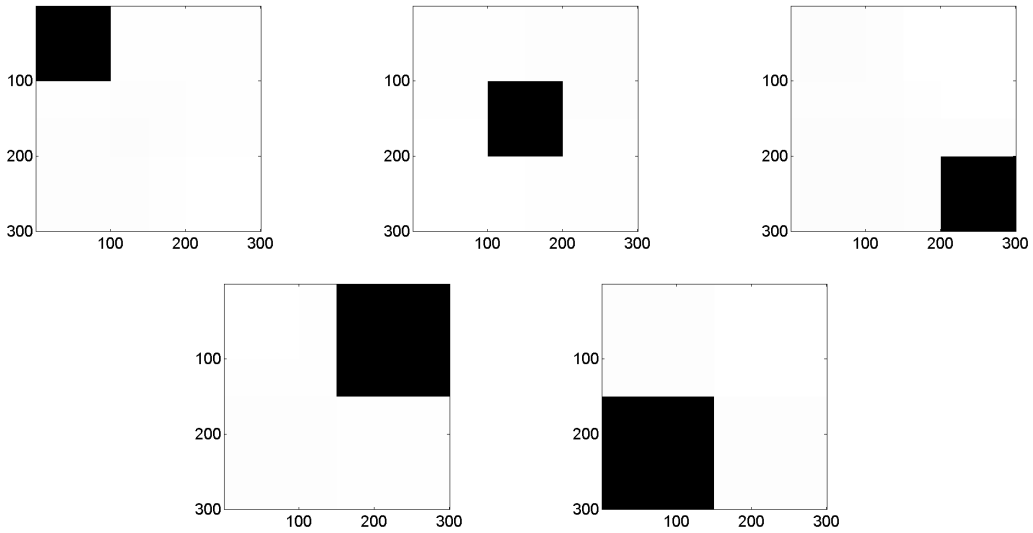

Figure 5: *Grayscale images of the averages for the level 1 (top row) and level 2 (bottom row) estimated eigenimages calculated from* 100 *replications of the 2nd scenario.*

Level 1 eigenvalues										
Component	1	2	3	4	5	6	7	8	9	10
eigenvalue ($\times 10^8$)	8.79	1.97	1.84	1.50	1.22	1.03	0.90	0.89	0.87	0.83
cum % var	13.46	16.48	19.30	21.60	23.47	25.06	26.43	27.79	29.13	30.40
Level 2 eigenvalues										
Component	1	2	3	4	5	6	7	8	9	10
eigenvalue ($\times 10^7$)	7.47	4.71	2.99	1.62	1.44	1.14	1.12	0.92	0.79	0.74
cum % var	11.91	19.44	24.21	26.80	29.09	30.91	32.70	34.16	35.42	36.60

Table 1: *Estimated eigenvalues for Level 1 and Level 2 for RAVENS data using HD-MFPCA. Ten first components are reported for each level. cum % var shows the cumulative percentage of variance explained.*

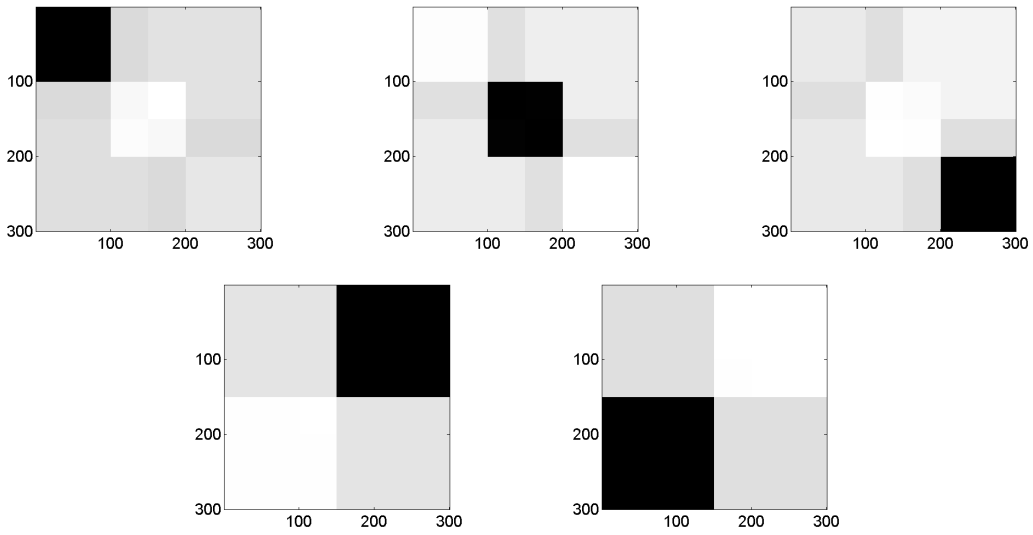

Figure 6: *Grayscale images of the* 5^{th} *pointwise percentiles for the level 1 (top row) and level 2 (bottom row) estimated eigenimages calculated from* 100 *replications of the 2nd scenario.*

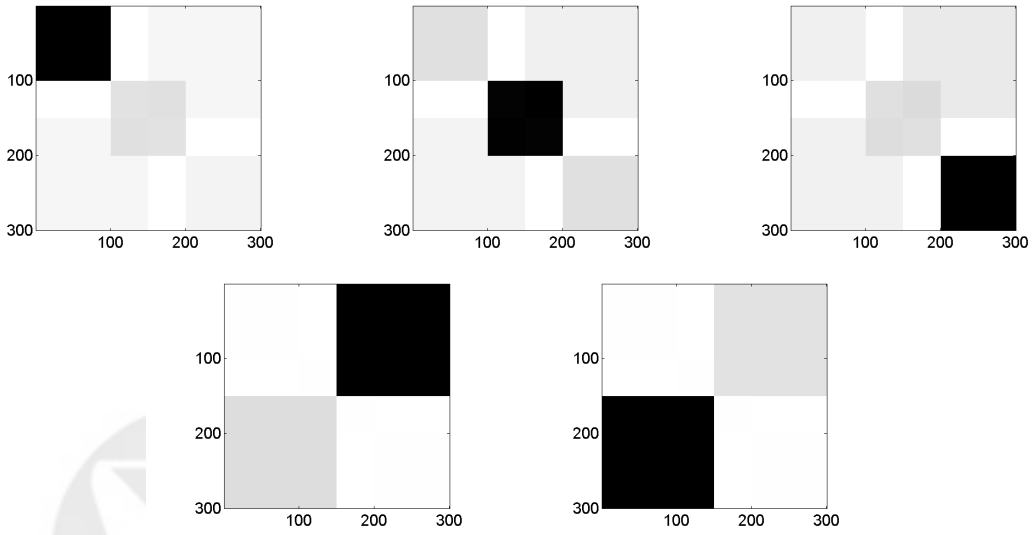

Figure 7: *Grayscale images of the* 95^{th} *pointwise percentiles for the level 1 (top row) and level 2 (bottom row) estimated eigenimages calculated from* 100 *replications of the 2nd scenario.*

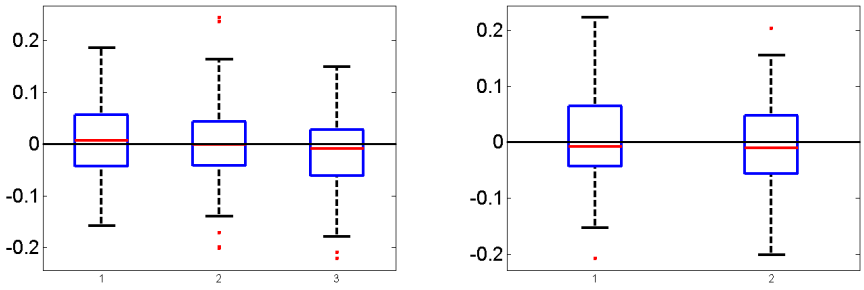

Figure 8: *Boxplots of the normalized estimated level* 1 *eigenvalues,* $(\hat{\lambda}_k^{(1)} - \lambda_k^{(1)})/\lambda_k^{(1)}$*, (left box) and the normalized estimated level* 2 *eigenvalues,* $(\hat{\lambda}_l^{(2)} - \lambda_l^{(2)})/\lambda_l^{(2)}$*, (right box) based on scenario* 2 *with* 100 *replications. The zero is shown by the solid black line.*

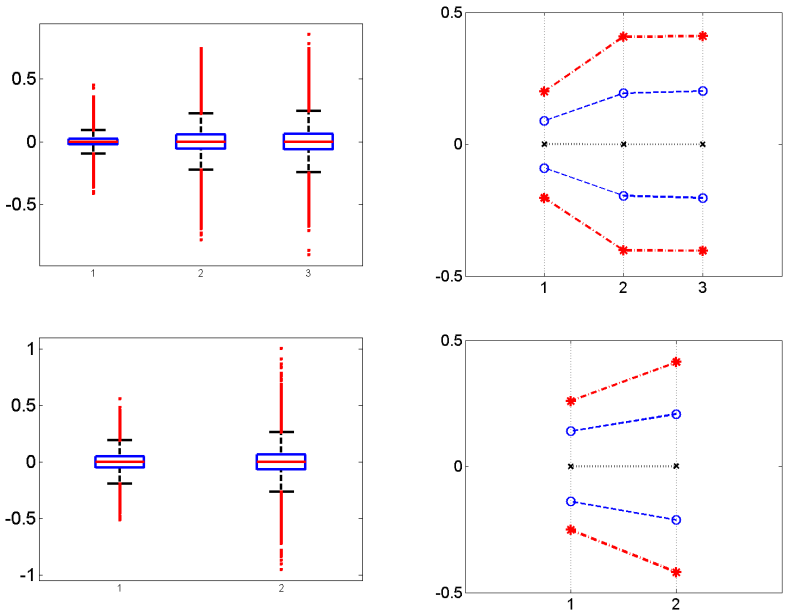

Figure 9: *Top row shows the distribution of the normalized estimated level* 1 *scores,* $(\xi_{ik} - \hat{\xi}_{ik})/\sqrt{\lambda_k^{(1)}}$*. Boxplots are given in left box. The right box shows the medians (black marker),* 5% *and* 95% *quantiles (blue markers), and* 0.5% *and* 99.5% *quantiles (red markers). Similarly, the distribution of the normalized estimated level* 2 *scores,* $(\zeta_{ijl} - \hat{\zeta}_{ijl})/\sqrt{\lambda_l^{(2)}}$ *is provided at the bottom row.*

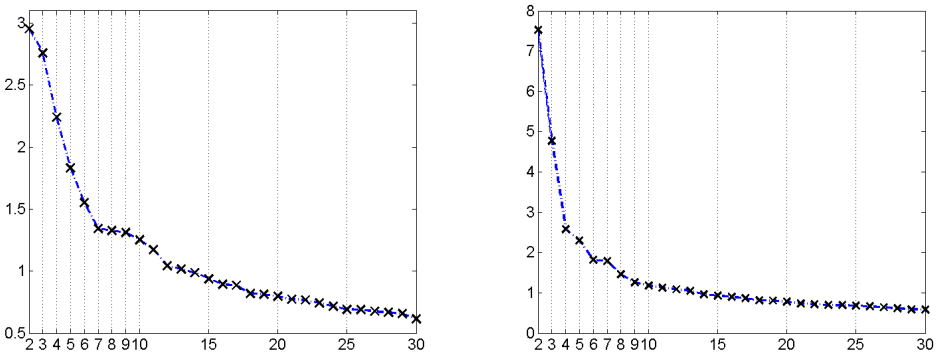

Figure 10: *On the left: proportions explained by the* 2*nd to* 30*th Level 1 eigenimages. First "between" eigenimage explains* 13.15% *of the "between" variation. On the right: proportions explained by the* 2*nd to* 30*th Level 2 eigenimages. First "within" eigenimage explains* 11.91% *of the "within" variation.*

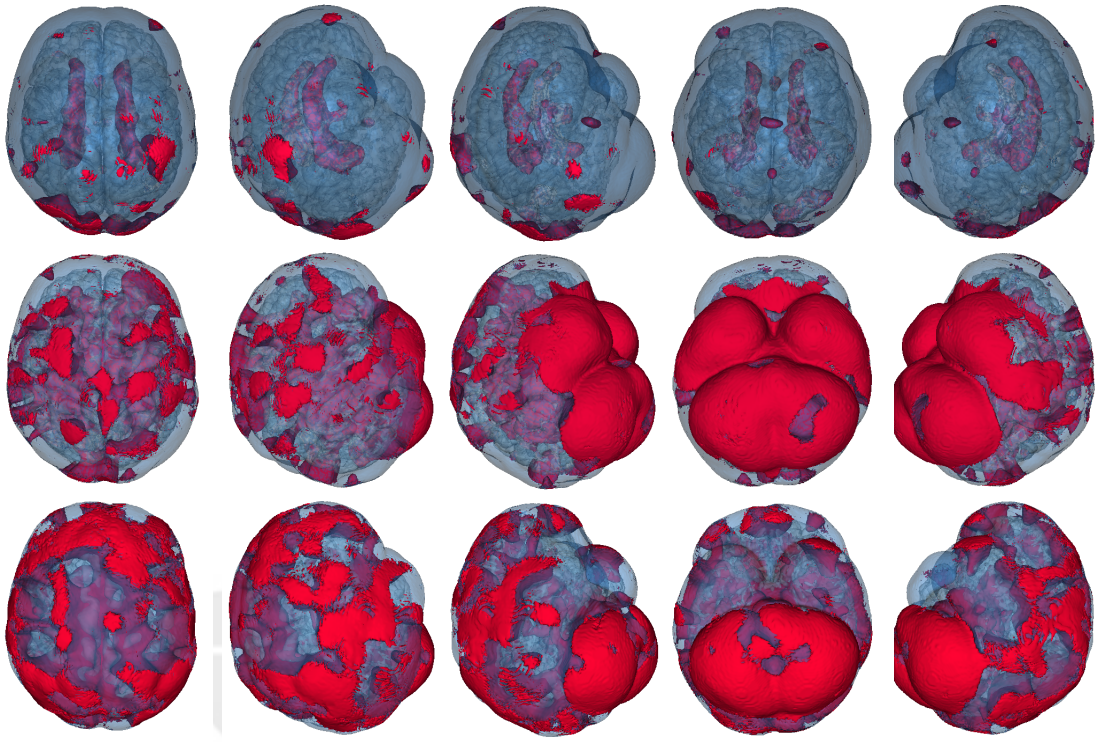

Figure 11: *3-Dimensional rendering of the first three between eigenimages overlaid with a thresholded template. Snapshots taken under five different angles. Negative loadings are depicted in red, positive ones are in blue.*

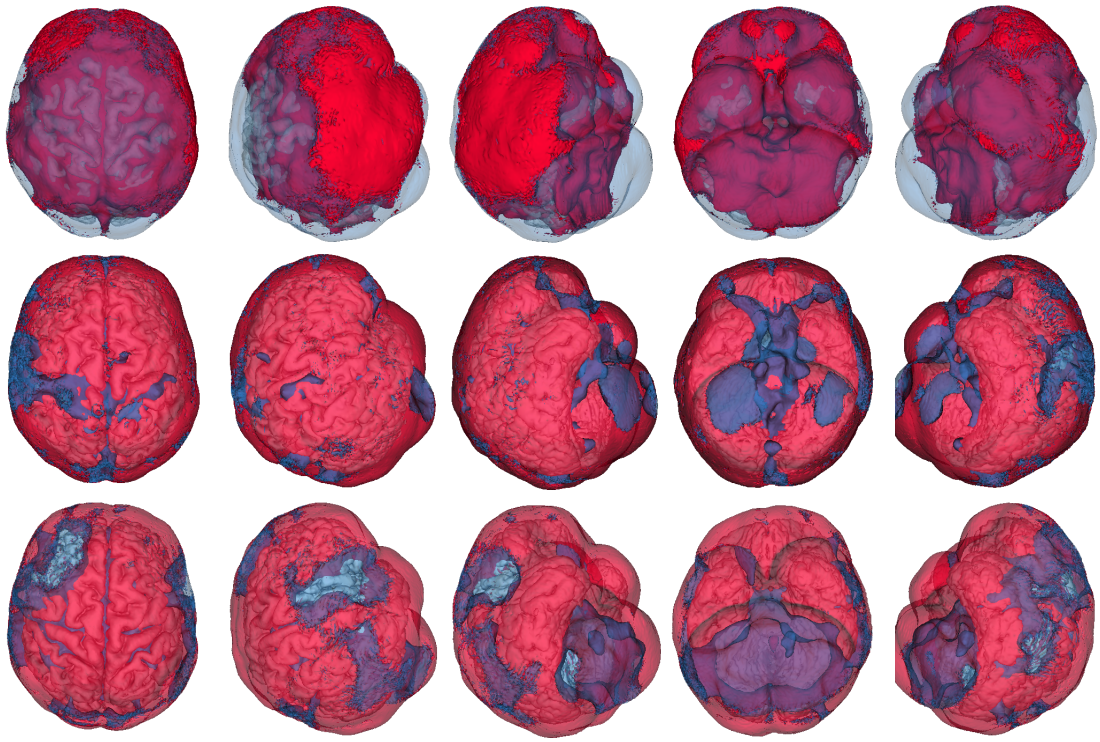

Figure 12: *3-Dimensional rendering of the first three within eigenimages overlaid with a thresholded template. Snapshots taken under five different angles. Negative loadings are depicted in red, positive ones are in blue.*

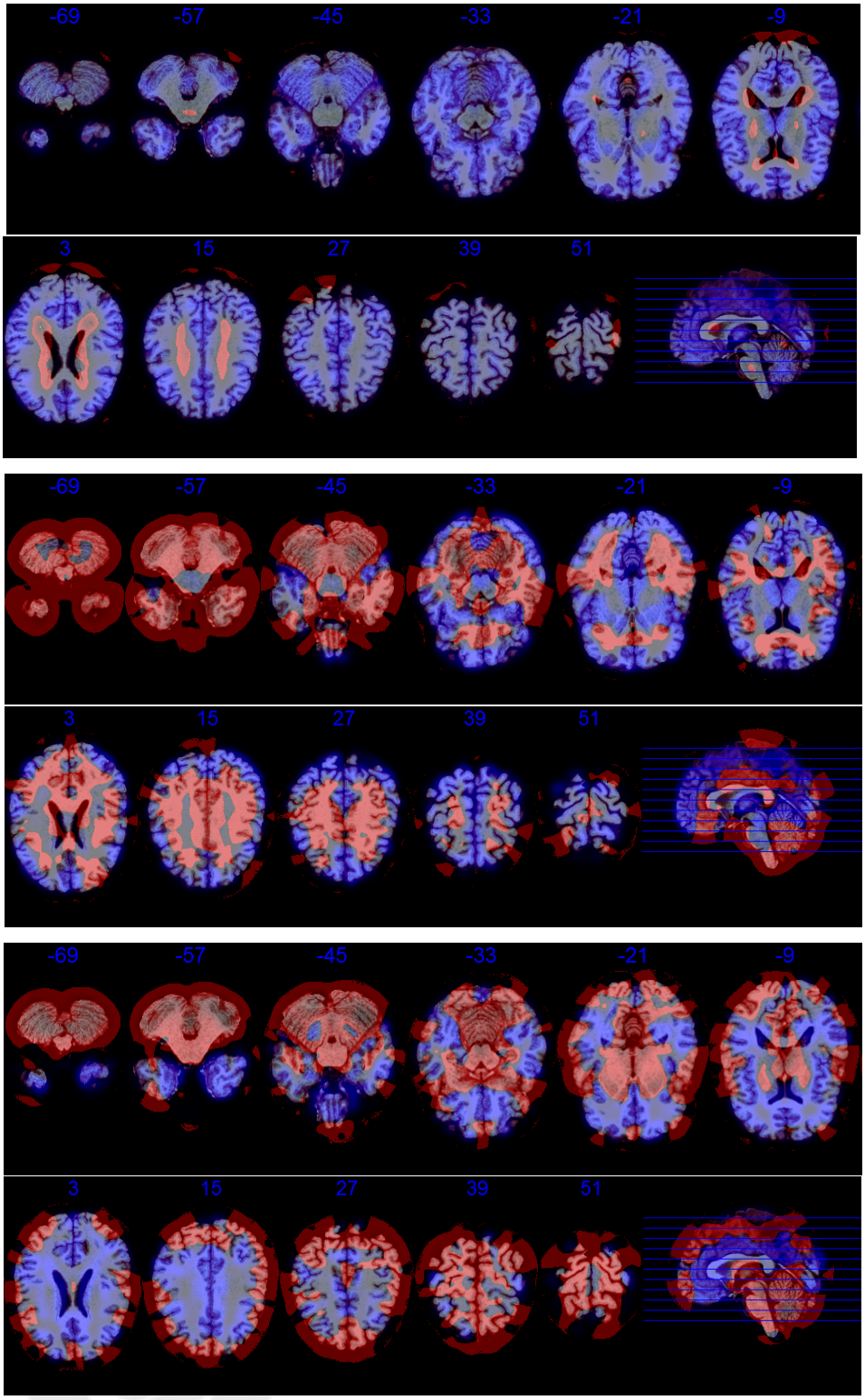

Figure 13: *The first three estimated between eigenimages. Each eigenimage is represented by eleven equidistant axial slices. Negative loadings are depicted in red, positive ones are in blue.*

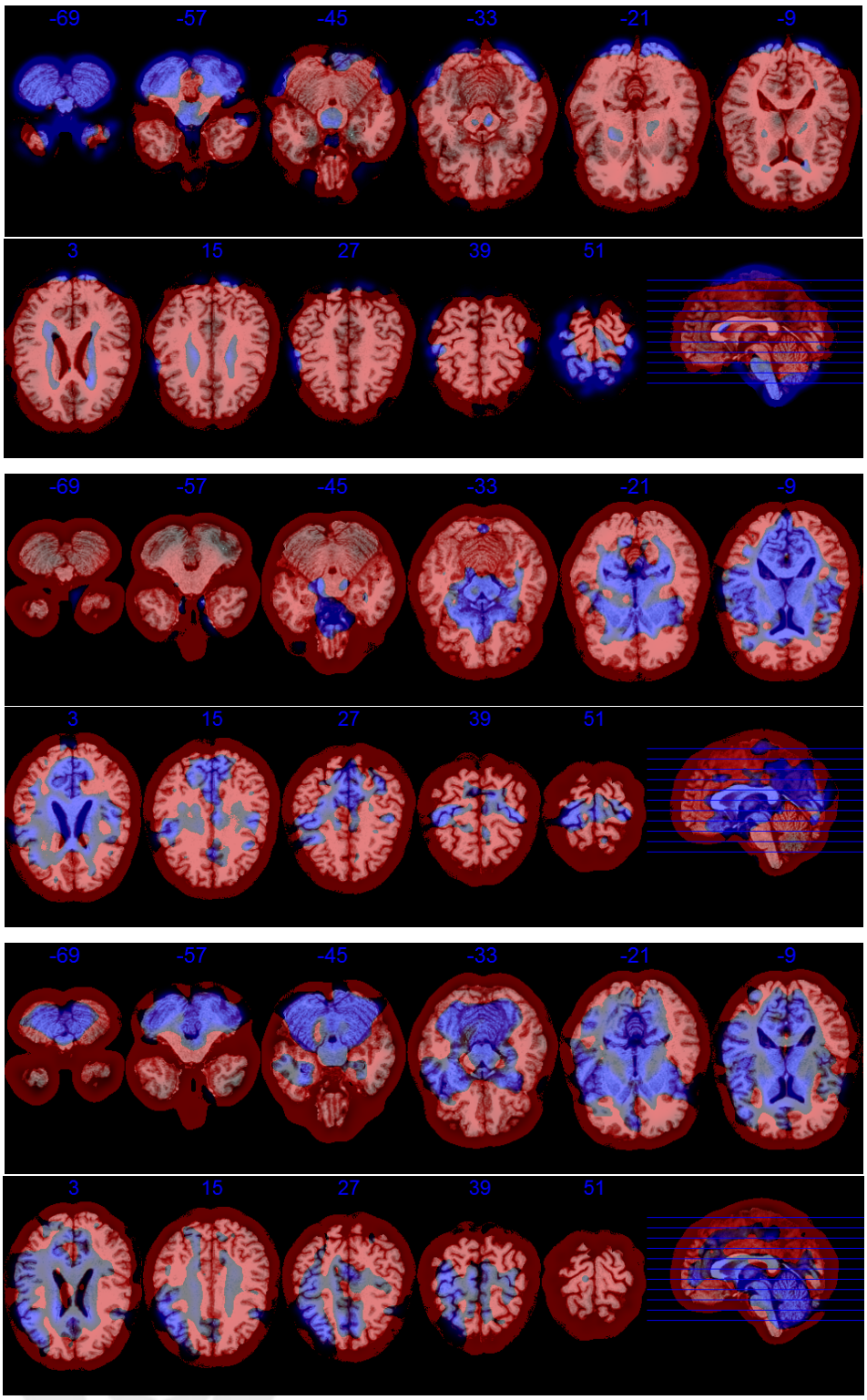

Figure 14: *The first three estimated within eigenimages. Each eigenimage is represented by eleven equidistant axial slices. Negative loadings are depicted in red, positive ones are in blue.*