1-18-2011

FUNCTIONAL PRINCIPAL COMPONENTS MODEL FOR HIGH-DIMENSIONAL BRAIN IMAGING

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

Brian S. Caffo
Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics, bcaffo@jhsph.edu

David M. Yousem
Johns Hopkins Hospital, Department of Radiology

Christos Davatzikos
University of Pennsylvania, Department of Radiology

Brian S. Schwartz
Johns Hopkins Bloomberg School of Public Health, Departments of Environmental Health Sciences and Epidemiology and Medicine

See next page for additional authors

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

Functional principal components model for high-dimensional brain imaging

Vadim Zipunnikov, Brian Caffo, David M. Yousem,

Christos Davatzikos, Brian S. Schwartz, Ciprian Crainiceanu*

January 18, 2011

Abstract

We establish a fundamental equivalence between singular value decomposition (SVD) and functional principal components analysis (FPCA) models. The constructive relationship allows to deploy the numerical efficiency of SVD to fully estimate the components of FPCA, even for extremely high-dimensional functional objects, such as brain images. As an example, a functional mixed effect model is fitted to high-resolution morphometric (RAVENS) images. The main directions of morphometric variation in brain volumes are identified and discussed.

*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

Key Words: Voxel-based morphometry (VBM), MRI, FPCA, SVD, brain imaging data.

1 Introduction

Epidemiological studies of neuroimaging data are becoming increasingly common. Common features of these studies generally include large sample sizes and subtle effects under study. High-resolution three-dimensional brain images exponentially increase the volume of data, making many standard inferential tools computationally infeasible. This and other high dimensional data sets have motivated an intensive effort in the statistical community on methodological research for functional data analysis (Di et al., 2008; Crainiceanu et al., 2009; Staicu et al., 2010; Greven et al., 2010; Di and Crainiceanu, 2010; Crainiceanu et al., 2010; Mohamed and Davatzikos, 2004; Reiss et al., 2005; Reiss and Ogden, 2008, 2010).

We put forward a generalization of principal components to understand major directions of variation in such large-scale neuroimaging studies. However, unlike most eigenimaging approaches, we connect the methods to formal mixed models for imaging data. Therefore, the approach yields a fully specified model and inferential framework. We further give a didactic explanation of easy methods for handling the necessary high dimensional calculations on even modest computing infrastructures.

Our proposed data-driven methods apply generally, though in this manuscript we specifically apply it to morphometric images that would typically be used for

voxel-based morophometry (Ashburner and Friston, 2000). In an imaging setting, the basic data requirement is a sample of spatially registered images, where the study of population variation in the registered intensities is of interest. Since the methods vectorize the imaging array information as a first step, whether the images are one, two, three or four (as in fMRI or PET studies) dimensional is irrelevant; though we stipulate that alternate methods that separate spatial and temporal variation (Beckmann and Smith, 2005; Caffo et al., 2010) are more relevant in the 4D cases. Regardless, the methods are generic and portable to a wide variety of imaging and non-imaging settings.

We also discuss the practical computing for the methods. We specifically demonstrate that model fitting can be performed via a SVD that can be applied iteratively, loading only components of the data at a time. Thereby, we demonstrate that the methods are scalable to large studies and can be executed on modest computing infrastructures.

The manuscript is laid out as follows. Section 2 describes the motivating data, regional tissue volume maps (RAVENS maps) derived from structural brain MRI of former organolead manufacturing workers. Section 3 explains why fitting FPCA model is identical to constructing SVD of the data matrix as well as provides necessary numerical adaptation to high-dimensional data. In Section 4 the method is applied to the RAVENS data. Section 5 concludes with a discussion.

2 Motivating data

The motivating data arises from a study of voxel-based morphometry (VBM) (Ashburner and Friston, 2000) in former organolead manufacturing workers. VBM is a common approach to analysis of structural MRI. The primary benefits of VBM are its lack of need for a-priori specified regions of interest and its exploratory nature. VBM facilitates identification of complex, and perhaps previously unknown, patterns of brain structure via regression models of exposure or disease status on deformation maps.

However, VBM, as its name suggests, is applied at a voxel-wise level, resulting in tens or hundreds of thousands of tests considered independently. In contrast, regional analyses are primarily confirmatory, requiring both specified regional hypotheses as well as an anatomical parcellation. We instead analyze morphometric images to find principal directions of cross-sectional variation of brain image shapes. While this approach is useful for both analyzing deformation fields as an outcome (functional principal components analysis), it is also useful for regression models where morphometric deformation is a predictor (functional principal component *regression*), (Ramsay and Silverman, 2010).

The data were derived from an epidemiologic study of the central nervous system effects of organic and inorganic lead in former organolead manufacturing workers, described in detail elsewhere (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 (described further below) was performed on the T1-weighted volume acquisitions.

RAVENS stands for Regional Analysis of VolumE in Normalized Space, and represents a standard method for discovering localized changes in brain shape related to exposures (Goldszal et al., 1998; Shen and Davatzikos, 2003). It has been shown to be scalable and viable on large epidemiological cohort studies (Davatzikos et al., 2008; Resnick et al., 2009). The method analyzes smoothed deformation maps obtained when registering subjects to a standard template. Processing, and hence analysis, is performed separately for different tissues types (gray/white) and possibly for the analysis of cerebrospinal fluid (CSF), which may be informative for ventricular volume and shape. A complete description of RAVENS processing can be found in Goldszal et al. (1998) and Shen and Davatzikos (2003). In this study, we consider images collected over two visits roughly five years apart that were registered using a novel 4D generalization of RAVENS processing (Xue et al., 2006). Hence we investigate cross-sectional variation, separately at the first and second visits, as well as longitudinal variation as summarized by difference maps between the two time points.

We emphasize that our proposed modeling does not depend on imaging modality and processing. (Though, of course, processing and scientific context will dictate the utility of the models.) The necessary inputs for the procedure are images registered in a standardized space, where voxel-specific intensities are of interest. For example, the methods equally apply to PET images of a tracer or DTI summary (e.g. fractional anisotropy, mean diffusivity) maps.

3 Methods

In this section we discuss FPCA model. The relationship between FPCA and SVD will be highlighted. This link will allow us to address efficiently the computational issues arising for FPCA model in high-dimensional settings. Furthermore, the geometrical interpretation of left and right singular vectors within FPCA framework will be closely examined.

3.1 Single level FPCA

Suppose that we have a sample of images $\mathbf{X}_i$, where $\mathbf{X}_i$ is a vectorized image of the i^{th} subject, $i = 1, \ldots, I$. 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 described in Section 2 $p = 256 \times 256 \times 198 = 12,976,128$. Of course, efficient masking of the data reduces this number drastically (to three million in the case of the RAVENS data). Hence, we represent the data $\mathbf{X}_i$ as a $p \times 1$ dimensional vector containing non-background voxels in a particular order, where the order is preserved across all voxels.

Following Di et al. (2008) we consider a single level functional model: $X_i(v) = \mu(v) + Z_i(v), i = 1, \ldots, I$ and v denotes a voxel coordinate. The image $\mu(v)$ is the overall mean image and $Z_i(v)$ is a subject-specific image deviation from the overall mean. We assume that $\mu(v)$ is fixed and $Z_i(v)$ is a zero-mean second-order stationary stochastic process with continuous covariance function. Using Karhunen-Loeve expansions of the random processes (Karhunen, 1947) $Z_i(v) = \sum_{k=1}^{\infty} \zeta_{ik}\phi_k(v)$ where

ϕ_k are the eigenfunctions of the $\mathbf{K}$ and ζ_{ik} are uncorrelated eigenscores with non-increasing variances σ_k. For practical purposes, we consider a model projected on the first N components. In addition, we assume that ζ_{ik}'s are independent and follow the same distribution with zero mean and variance σ_k. Further, it will be more convenient to introduce random variable $\xi_{ik} \overset{i.i.d}{\sim} (0,1)$ having the same distribution as ζ_{ik} but normalized to have unit variance. With these changes the FPCA model becomes a mixed effect model (McCulloch and Searle, 2001, Ch.6)

$$X_i(v) = \mu(v) + \sum_{k=1}^{N} \sigma_k^{1/2} \xi_{ik} \phi_k(v), \quad \xi_{ik} \overset{i.i.d.}{\sim} (0,1). \tag{1}$$

Typically, a small number of principal components (or eigenimages), N, can explain the most of the variation (Di et al., 2008). Statistical estimation of model (1) includes estimating eigenimages ϕ_k with eigenvalues σ_k and eigenscores ξ_{ik} .

The clear estimate $\boldsymbol{\mu}$, the vectorized version of $\mu(v)$, is the sample point-wise arithmetic average $\hat{\boldsymbol{\mu}} = \sum_{i=1}^{I} \mathbf{X}_i / I$. The unexplained part of the image, $\widetilde{\mathbf{X}}_i = \mathbf{X}_i - \hat{\boldsymbol{\mu}}$, is eigen-analyzed to obtain the eigenvectors ϕ_k and eigenvalues σ_k. Denote $\widetilde{\mathbf{X}} = (\widetilde{\mathbf{X}}_1, \ldots, \widetilde{\mathbf{X}}_I)$ where $\widetilde{\mathbf{X}}_i$ is a centered $p \times 1$ vector containing the unfolded image for subject i. Then covariance operator $\hat{\mathbf{K}}$ is estimated as $\hat{\mathbf{K}} = \frac{1}{I} \sum_{i=1}^{I} \widetilde{\mathbf{X}}_i \widetilde{\mathbf{X}}_i'$. Given $rank(\hat{\mathbf{K}}) = r$ the covariance operator $\hat{\mathbf{K}}$ can be decomposed as $\hat{\mathbf{\Phi}} \hat{\mathbf{\Sigma}} \hat{\mathbf{\Phi}}'$ where $p \times r$ matrix $\hat{\mathbf{\Phi}}$ has orthonormal columns, $\hat{\phi}_k$, and $r \times r$ diagonal matrix $\hat{\mathbf{\Sigma}}$ has non-negative diagonal elements $\hat{\sigma}_1 \geq \hat{\sigma}_2 \geq .. \geq \hat{\sigma}_r > 0$. The number of principal components, N, is typically chosen to make the explained variability $(\hat{\sigma}_1 + \ldots + \hat{\sigma}_N)/(\hat{\sigma}_1 + \ldots + \hat{\sigma}_r)$

large enough. Although, there are more sophisticated methods (Di et al. (2008) and Crainiceanu et al. (2009)).

The size of the covariance operator $\hat{\mathbf{K}}$ is $p \times p$. For high-dimensional p the brute-force eigenanalysis requires $O(p^3)$ operations and as a result is infeasible. Calculating and storing $\hat{\mathbf{K}}$ becomes impossible when p reaches infeasible levels.

Nevertheless, it is still possible to get eigendecomposition of $\hat{\mathbf{K}}$ by using the fact that the number of subjects, I, is typically much smaller than p. Indeed, if $I < p$ then matrix $\widetilde{\mathbf{X}} = (\widetilde{\mathbf{X}}_1, \ldots, \widetilde{\mathbf{X}}_I)$ has at most rank I and the SVD of $\widetilde{\mathbf{X}}$

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

can be obtained with $O(pI^2 + I^3)$ computational effort (Golub and Loan, 1996). Here, the matrix $\mathbf{V}$ is $p \times I$ with I orthonormal columns, $\mathbf{S}$ is a diagonal $I \times I$ diagonal matrix and $\mathbf{U}$ is a $I \times I$ orthogonal matrix. Full details on efficient SVD calculation for ultra high-dimensional p will be provided in the next section. Now we will show the relation between FPCA (1) and SVD (2).

Assume for a moment that we calculated (2). Then $\hat{\mathbf{K}} = \hat{\mathbf{\Phi}}\hat{\mathbf{\Sigma}}\hat{\mathbf{\Phi}}' = (1/I)\mathbf{V}\mathbf{S}\mathbf{V}'$. Given all eigenvalues are different, the eigendecomposition of $\hat{\mathbf{K}}$ is unique. Thus,

$$\hat{\mathbf{\Phi}} = \mathbf{V} \quad \text{and} \quad \hat{\mathbf{\Sigma}} = (1/I)\mathbf{S}. \tag{3}$$

which determines the estimates of eigenimages $\hat{\phi}_k$ and eigenvalues $\hat{\sigma}_k$. Estimated

eigenfunctions $\hat{\phi}_k$ and eigenvalues $\hat{\sigma}_k$ are used to calculate the estimated best linear unbiased predictors (EBLUPs) (McCulloch and Searle, 2001, Ch.9) of the scores ξ_{ik}. Again, brute-force calculation of EBLUPs requires the inversion of high-dimensional matrices (see Di and Crainiceanu, 2010; Crainiceanu et al., 2009) and becomes prohibitive. Next we will show that the EBLUPs for model (1) is nothing but vectors orthonormal to the right singular vectors of the SVD.

Denote $\boldsymbol{\xi}_i = (\boldsymbol{\xi}_{i1}, \ldots, \boldsymbol{\xi}_{iN})$. Conditional expectation of $\boldsymbol{\xi}_i$ given $\widetilde{\mathbf{X}}_i$ provides the BLUP of $\boldsymbol{\xi}_i$ (McCulloch and Searle, 2001, Ch.9). If $p \leq I$ it can be written as:

$$\hat{\boldsymbol{\xi}}_i = E(\boldsymbol{\xi}_i|\widetilde{\mathbf{X}}_i) = E(\boldsymbol{\xi}_i\widetilde{\mathbf{X}}_i')Var(\widetilde{\mathbf{X}}_i)^{-1}\widetilde{\mathbf{X}}_i = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\Phi}'(\boldsymbol{\Phi}\boldsymbol{\Sigma}\boldsymbol{\Phi}')^{-1}\widetilde{\mathbf{X}}_i. \tag{4}$$

If matrices $\boldsymbol{\Phi}$ and $\boldsymbol{\Sigma}$ are known then $\hat{\boldsymbol{\xi}}_i = \boldsymbol{\xi}_i$; in other words, with known variances, we can exactly recover the eigenscores $\boldsymbol{\xi}_i$. However, in practice both the eigenvectors and the variances are estimated and these estimators are plugged into (4) to get estimated BLUPs. Combining (3) and (2) in (4) leads to

$$\hat{\boldsymbol{\xi}}_i = \mathbf{U}(i, 1:N) \tag{5}$$

where $\mathbf{U}(i, 1:N)$ denotes the first N coordinates of the ith row of the matrix $\mathbf{U}$. Note that the independence of images $\mathbf{X}_i$'s translates geometrically into orthogonality of the rows of $\mathbf{U}$. The independence of eigenscores ξ_{ik} is equivalent to orthogonality of the columns of $\mathbf{U}$.

We are interested in a situation when p is larger than I and (4) can not be applied directly. In this case, the BLUP is expressed via pseudo-inverse matrices (Harville, 1976): $\hat{\boldsymbol{\xi}}_i = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\Phi}'(\boldsymbol{\Phi}\Sigma\boldsymbol{\Phi}')^{-}\widetilde{\mathbf{X}}_i$ where $(\boldsymbol{\Phi}\boldsymbol{\Sigma}\boldsymbol{\Phi}')^{-}$ is the unique generalized inverse of the matrix $\boldsymbol{\Phi}\boldsymbol{\Sigma}\boldsymbol{\Phi}'$ which equals to $\boldsymbol{\Phi}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Phi}'$ (see Demidenko, 2004, Appendix). Combining (2) and (3) in (4) and using the form of the generalized inverse we obtain that (5) is true for case $p \geq I$.

To summarize, we demonstrated that: i) the eigenvectors ϕ_k are given by the left singular vectors $\mathbf{v}_k$; ii) the normalized principal scores ξ_{ik} are given by vectors orthonormal to the right singular vectors $\mathbf{u}_k$; and iii) the variances σ_k are estimated by the scaled singular values s_k/I.

3.2 Implementation

Now we give details of a fast and efficient algorithm for calculating SVD with $O(pI^2 + I^3)$ computational effort and sequential access to the memory. It was easily implemented on a regular PC and completed in minutes for the Former Lead Worker's RAVENS data. First step is to use $I \times I$ symmetric matrix $\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}$ and its spectral decomposition $\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}} = \mathbf{U}\mathbf{S}\mathbf{U}'$ to get $\mathbf{U}$ and $\mathbf{S}^{1/2}$. For high-dimensional p the matrix $\widetilde{\mathbf{X}}$ can not be loaded into the memory. The solution we suggest 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 $p/M \times I$ 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 $\sum_{m=1}^{M}(\widetilde{\mathbf{X}}^m)'\widetilde{\mathbf{X}}^m$ and requires $O(pI^2)$ operations. Spectral decomposition for $\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}$ requires $O(I^3)$ opera-

tions and calculates matrices $\mathbf{U}$ and $\mathbf{S}$. The $p \times I$ matrix $\mathbf{V}$ can now be obtained as $\mathbf{V} = \widetilde{\mathbf{X}}\mathbf{U}\mathbf{S}^{-1/2}$. Actual calculations can be performed on the slices of the partitioned matrix $\widetilde{\mathbf{X}}$ as $\mathbf{V}^m = \widetilde{\mathbf{X}}^m\mathbf{U}\mathbf{S}^{-1/2}, m = 1; M$ and can be done with $O(pI^2)$ operations. The concatenated slices $[(\mathbf{V}^1)'|(\mathbf{V}^2)'|\ldots|(\mathbf{V}^M)']$ form the matrix of the left singular vectors $\mathbf{V}'$. Hence, all components of the SVD can be calculated without loading the entire data matrix into memory. The analysis scales to nearly arbitrary large parameter spaces on very modest computing infrastructures.

4 Application to RAVENS images

In this section we apply our method to the RAVENS images described in Section 2. The RAVENS images are $256 \times 256 \times 198$ dimensional for 352 subjects, each with two visits roughly five years apart. We analyze visit 1 and visit 2 separately. In addition, to identify the principal directions of the longitudinal change we consider a difference between images taken at visit 1 and visit 2. Although the data contains both white and gray matter as well as CSF, for illustration, the analysis is restricted only to the processed gray matter data. A small technical concern was of a few artifactual negative values in the data from the preprocessing. These voxels were removed from the analysis. After processing, the intersection of non-background voxels across images was collected. Such an intersection greatly reduced the dimension of the data matrix from ten billion numbers to two billion numbers divided as three million relevant voxels per subject per visit with seven hundred and four subject-visits.

Following Section 3.2 all calculations were performed in such a way that only one of the manageable submatrices $\widetilde{\mathbf{X}}^m$ needs to be stored in memory at any given moment. The data matrix, of size 704 by 3 million, was divided into 100 submatrices of size 704 by 30 thousand (ten million numbers each). Note that on lower-resource computers the only change would be to reduce the size of submatrices. All calculations repeated for each of the three data sets were performed in Matlab 2010a and took around 15 minutes for each set on a PC with a quad core i7-2.67Gz processor and 6Gb of RAM memory.

In the analysis, we first estimated the mean by the empirical voxel-specific arithmetic average. 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. The same is true for the mean of the longitudinal differences. In our eigenimage analysis we de-mean the data by subtracting out these vectors and work with de-meaned matrix $\widetilde{\mathbf{X}}$.

Figure 1 shows the proportions of morphometric variation explained by the first thirty eigenimages for visit 1, visit 2, and the longitudinal difference. Cumulatively, the first thirty eigenimages explain 46.6%, 45.7%, and 52.5% of variation in data for visit 1, visit 2, and the longitudinal difference, respectively. The way eigenvalues decay on the most right graph of Figure 1 is a clear indication that the longitudinal changes can be accurately described by the first thirty principal components explaining more than half of the longitudinal variation. Although the number of principal components, N, is usually chosen to explain enough variation (Di et al., 2008), our primary interest

is the first few which identify the regions of brain exhibiting the most morphometric variation. The pattern of the percentage decrease on all three graphs of Figure 1 flattens out after approximately the first ten principal components. Therefore, we concentrate our analysis on the first ten principal components.

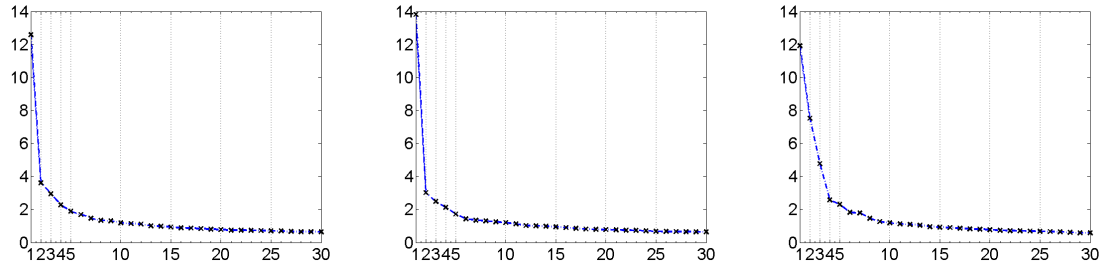

Figure 1: *Proportions of morphometric variation explained by the first thirty eigenimages (from left to right: visit* 1*, visit* 2*, and the longitudinal difference).*

Table 1 provides the cumulative percentages of variability explained by the first ten eigenimages. For visit 1 (top row) and visit 2 (middle row), they explain roughly the same amount of observed variation, 30%. For the longitudinal difference (bottom row), they explain 36.5% of the observed variability.

visit 1										
Component	1	2	3	4	5	6	7	8	9	10
cum % var	12.58	16.20	19.15	21.42	23.31	25.00	26.47	27.81	29.11	30.29
visit 2										
Component	1	2	3	4	5	6	7	8	9	10
cum % var	13.81	16.82	19.30	21.43	23.15	24.57	25.92	27.22	28.48	29.68
longitudinal difference										
Component	1	2	3	4	5	6	7	8	9	10
cum % var	11.91	19.44	24.21	26.80	29.09	30.91	32.70	34.16	35.42	36.60

Table 1: *Cumulative percentage of variation explained by first ten eigenimages for RAVENS data (visit* 1 *(top row), visit 2 (middle row), and the longitudinal difference (bottom row)).*

Top panel of Figure 2 provides the estimated actual eigenvalues for the eigenim-

ages. Notice, however, that we are more interested in the relative size of the eigenvalues representing quantitative measure of variability of the related eigenscores. Bottom panel of Figure 2 plots the distributions of the eigenscores corresponding to the first ten eigenimages. In Section 3.1 we showed that the estimates of the normalized eigenscores are given by the right singular vectors of matrix $\widetilde{\mathbf{X}}$. Therefore, the estimates of unnormalized eigenscores can be obtained once we multiply them by the square root of the corresponding eigenvalues provided in the top panel of Figure 2. The estimated eigenscores serve as (signed) quantifiers relating eigenimages to subjects and their RAVENs maps. As we can see, the distribution of eigenscores in visit 1 and visit 2 are close to each other.

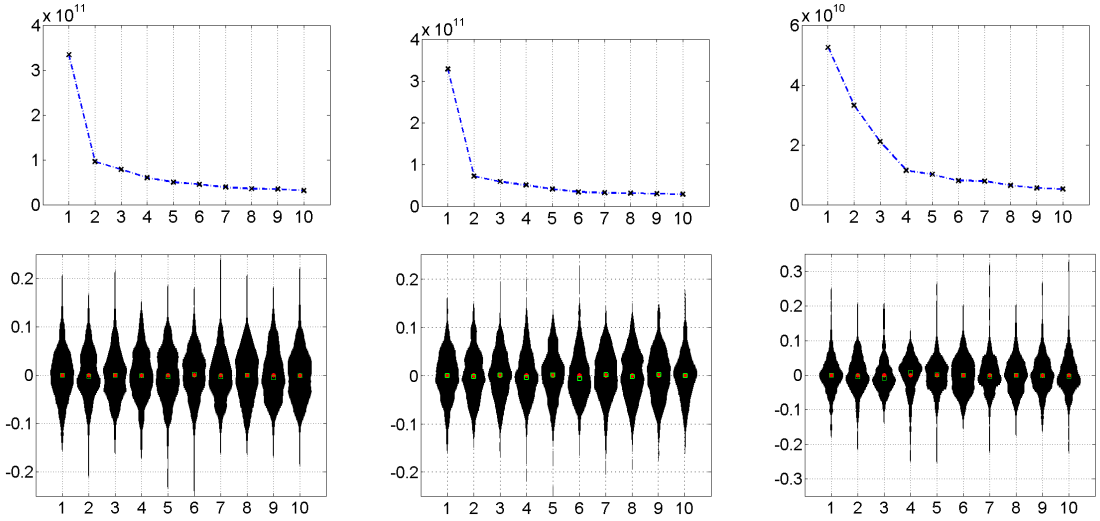

Figure 2: *normalized distributions of the eigenscores corresponding to the first ten eigenimages (from left to right: visit* 1*, visit* 2*, and the longitudinal difference).*

We now discuss overlap of the eigenimages with anatomical regions. Due to space limitations we discuss and depict only the first three eigenimages. The k^{th} eigenimage explains $\sigma_k = \sigma_k \phi_k^{'} \phi_k$ amount of variation. Recall, each coordinate of ϕ_k corresponds

to a voxel in template space. Therefore, if the template is parcellated into regions, then we can calculate the proportion of the variance explained by this particular region within eigenimage ϕ_k - on a scale from 0 to 1. In our study, the template has been divided into $R = 91$ regions displayed in Table 5. However, the approach is general and applicable to any parcellation. Therefore, the variance explained by the k^{th} eigenimage can be further decomposed as $\sigma_k = \sigma_k \sum_{r=1}^{91} w_{kr}$,where non-negative weights w_{kr} sum over the 91 regions to one and represent the proportion of variance σ_k explained by region r. In Table 2 we provide the variance explained by the labeled regions of the template for Visit 1. The twenty five regions with the highest loadings for each of the first three eigenimages are provided. Note that because of sign invariance of the decomposition, the separation between positive and negative loading is comparable only within an eigenimage. Tables 1, 2, and 5 give now a way to determine a (signed) quantitative contribution of each particular region. For instance, the right middle temporal gyrus (130) explains 4.5% of the variance within eigenimage 1, which in turn explains 12.58% of the total variation. Hence, the right middle temporal gyrus explains $4.5\% * 12.58\% = 0.57\%$ of the total variation and has a mostly positive loading within eigenimage 1. Similarly, Tables 3 and 4 provide the regional quantifications of explained variation for Visit 2 and the longitudinal difference, respectively.

As we showed in Section 3.1 the estimated principal components (eigenimages) are left singular vectors of matrix $\widetilde{\mathbf{X}}$. Each left singular vector is of size $p \approx 3 \cdot 10^6$ unfolded voxels. Therefore, each voxel is represented by a small value between

negative and positive one and squares of the voxel values are summed to one. Negative and positive voxel values correspond to the opposite directions (loadings) of variation. The distribution of the negative and positive voxel loadings are presented in Figure 3 in red and blue, respectively. The voxel values of the estimated eigenimage $\hat{\boldsymbol{\phi}} = (\hat{\phi}_1, \ldots, \hat{\phi}_p)$ were transformed as $\hat{\boldsymbol{\phi}} \rightarrow 256 \cdot (\hat{\boldsymbol{\phi}} - \min_s \hat{\phi}_s)/(\max_s \hat{\phi}_s - \min_s \hat{\phi}_s)$ separately for voxels with positive and negative loadings. The transformed negative and positive loadings overlaid with the template are presented in Figures 4, 5, and 6.

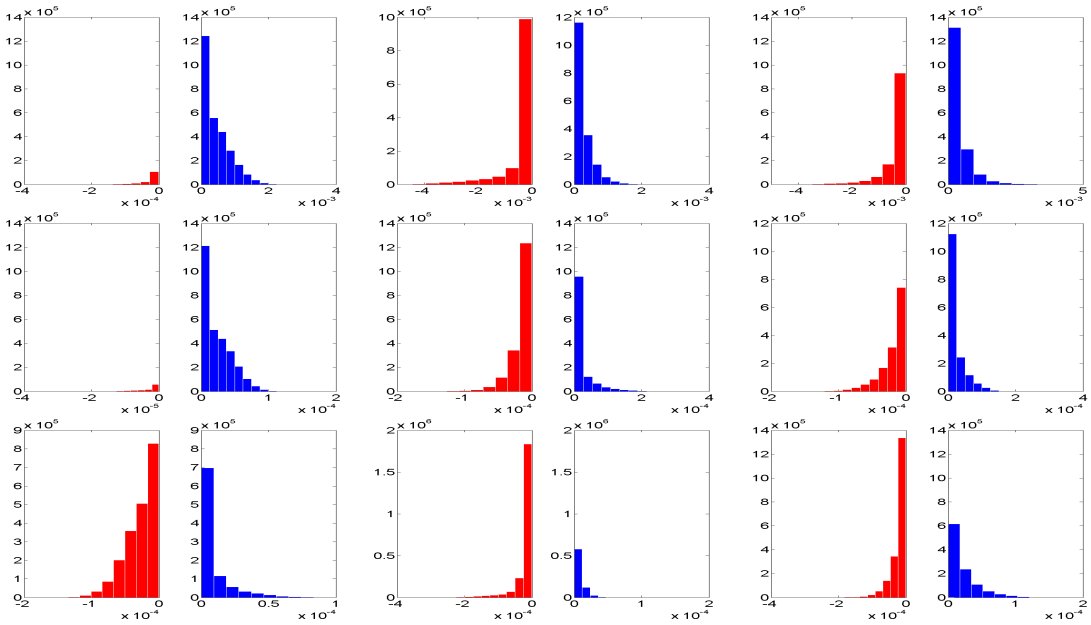

Figure 3: *Distributions of the intensities of the first three eigenimages (visit* 1 *(top row), visit* 2 *(middle row), and the longitudinal difference (bottom row)).*

5 Discussion

In this paper we proved a connection between SVD and functional mixed effect models. This coupling allowed us to develop efficient model-based computing techniques.

The developed approach was applied to a novel morphometric data set with 704 RAVENS images. Principal components of morphometric variation were identified and studied. An alternative to our analysis would be a more formal separation of cross-sectional and longitudinal morphometric variation within multilevel functional principal component analysis framework suggested in (Di et al., 2008).

There are a few important limitations in the presented methodology. First, we have not assumed noise in the model. RAVENS data represent preprocessed and smoothed images. However, there are considerable number of studies collecting functional observations measured with non-ignorable noise. In addition, the model we considered does not allow any sparsity of the high-dimensional functional observations. This issue was addressed in (Di et al., 2008) and (Di and Crainiceanu, 2010) for multilevel models. The proposed efficient solutions were based on smoothing of the covariance operator which is infeasible for high-dimensional data. Therefore, there is a great demand in computationally efficient procedures of covariance smoothing 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.

References

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

Beckmann, C. and S. Smith (2005). Tensorial extensions of independent component analysis for multisubject fMRI analysis. *Neuroimage 25*(1), 294–311.

Caffo, B., C. Crainiceanu, G. Verduzco, S. Joel, S. H. Mostofsky, S. S. Bassett, and J. J. Pekar (2010). Two-stage decompositions for the analysis of functional connectivity for fmri with application to alzheimer's disease risk. *NeuroImage 51*, 1140–1149.

Crainiceanu, C. M., B. S. Caffo, S. Luo, V. V. Zipunnikov, and N. M. Punjabi (2010). Population value decomposition, a framework for the analysis of image populations. *Johns Hopkins University, Dept. of Biostatistics Working Paper 220*.

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., Y. Fan, X. Wu, D. Shen, and S. M. Resnick (2008). Detection of prodromal alzheimers disease via pattern classification of magnetic resonance imaging. *Neurobiology of Aging 29*, 514–523.

Demidenko (2004). *Mixed Models: Theory and Applications*. John Wiley and Sons.

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.

Goldszal, A., C. Davatzikos, D. Pham, M. Yan, R. Bryan, and S. R. SM (1998). An image processing protocol for the analysis of mr images from an elderly population. *Journal of Computer Assisted Tomography 22*, 827–837.

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

Greven, S., C. M. Crainiceanu, B. S. Caffo, and D. Reich (2010). Longitudinal functional principal component analysis. *Electronic Journal of Statistics 4*, 1022–1054.

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

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

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

Mohamed, A. and C. Davatzikos (2004). Shape representation via best orthogonal ba-

sis selection. *Medical Image Computing and Computer-Assisted Intervention 3216*, 225–233.

Ramsay, J. and B. Silverman (2010). *Functional Data Analysis* (second ed.). Springer.

Reiss, P. and R. T. Ogden (2008). Functional generalized linear models with applications to neuroimaging. *Poster presentation Workshop on Contemporary Frontiers in High-Dimensional Statistical Data Analysis*.

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.

Resnick, S., M. A. Espeland, S. A. Jaramillo, C. Hirsch, M. L. Stefanick, A. M. Murray, J. Ockene, and C. Davatzikos (2009). Postmenopausal hormone therapy and regional brain volumes: the whims-mri study. *Neurology 72*, 135–142.

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.

Shen, D. G. and C. Davatzikos (2003). Very high resolution morphometry using mass-preserving deformations and hammer elastic registration. *NeuroImage 18*, 28–41.

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.

Xue, Z., D. Shen, and C. Davatzikos (2006). CLASSIC: consistent longitudinal alignment and segmentation for serial image computing. *Neuroimage 30*(2), 388–399.

sectionAppendix

Eigenimage 1				Eigenimage 2				Eigenimage 3			
255	0.0508	0.0508	0.0000	27	0.0222	0.0222	0.0000	255	0.0719	0.0540	0.0179
130	0.0450	0.0450	0.0000	30	0.0179	0.0170	0.0009	83	0.0295	0.0009	0.0287
17	0.0410	0.0410	0.0000	255	0.0175	0.0124	0.0051	165	0.0285	0.0240	0.0045
30	0.0399	0.0399	0.0000	17	0.0143	0.0134	0.0008	64	0.0275	0.0011	0.0265
59	0.0381	0.0380	0.0001	7	0.0129	0.0129	0.0000	102	0.0255	0.0255	0.0000
145	0.0331	0.0331	0.0000	83	0.0124	0.0114	0.0010	95	0.0254	0.0000	0.0254
83	0.0298	0.0297	0.0000	59	0.0097	0.0085	0.0012	203	0.0235	0.0235	0.0000
61	0.0287	0.0287	0.0000	203	0.0073	0.0041	0.0032	30	0.0212	0.0180	0.0032
64	0.0268	0.0268	0.0000	6	0.0072	0.0072	0.0000	99	0.0202	0.0026	0.0177
27	0.0237	0.0237	0.0000	196	0.0066	0.0022	0.0044	108	0.0200	0.0199	0.0001
99	0.0221	0.0221	0.0000	105	0.0059	0.0055	0.0004	17	0.0185	0.0160	0.0025
2	0.0205	0.0205	0.0000	102	0.0059	0.0004	0.0054	94	0.0181	0.0025	0.0156
7	0.0201	0.0201	0.0000	3	0.0058	0.0058	0.0000	92	0.0176	0.0000	0.0176
75	0.0197	0.0197	0.0000	57	0.0057	0.0052	0.0005	21	0.0176	0.0000	0.0176
196	0.0187	0.0187	0.0000	90	0.0052	0.0052	0.0000	119	0.0172	0.0106	0.0066
119	0.0166	0.0166	0.0000	64	0.0051	0.0049	0.0001	196	0.0169	0.0071	0.0097
15	0.0158	0.0158	0.0000	119	0.0050	0.0017	0.0033	4	0.0158	0.0157	0.0001
105	0.0155	0.0154	0.0001	8	0.0050	0.0050	0.0000	59	0.0157	0.0048	0.0109
57	0.0150	0.0150	0.0000	75	0.0049	0.0047	0.0002	61	0.0118	0.0016	0.0101
165	0.0148	0.0148	0.0000	133	0.0047	0.0046	0.0001	88	0.0114	0.0103	0.0012
50	0.0147	0.0147	0.0000	61	0.0045	0.0031	0.0015	75	0.0114	0.0113	0.0001
4	0.0146	0.0146	0.0000	52	0.0045	0.0042	0.0003	114	0.0111	0.0107	0.0004
5	0.0144	0.0144	0.0000	99	0.0039	0.0025	0.0015	5	0.0108	0.0104	0.0004
108	0.0141	0.0141	0.0000	20	0.0039	0.0000	0.0039	145	0.0105	0.0100	0.0005
74	0.0116	0.0116	0.0000	32	0.0036	0.0036	0.0000	9	0.0100	0.0096	0.0004

Table 2: *Visit 1: Proportion of the variance explained by the regions of the template (see Table 5 for the template parcellation). The twenty five regions with the highest loadings are provided. For each eigenimage: first column shows the label, second shows the proportions of variance explained within this eigenimage (in decreasing order), third quantifies the positive loading (blue), fourth quantifies the negative loading(red).*

Eigenimage 1				Eigenimage 2				Eigenimage 3			
255	0.0542	0.0542	0.0000	255	0.0425	0.0120	0.0306	255	0.0612	0.0500	0.0113
30	0.0438	0.0437	0.0001	99	0.0274	0.0258	0.0016	30	0.0342	0.0132	0.0211
17	0.0420	0.0419	0.0001	30	0.0165	0.0004	0.0161	17	0.0251	0.0061	0.0190
130	0.0390	0.0390	0.0000	165	0.0165	0.0066	0.0099	27	0.0222	0.0006	0.0215
59	0.0338	0.0336	0.0002	196	0.0153	0.0120	0.0033	83	0.0220	0.0015	0.0205
145	0.0307	0.0307	0.0000	17	0.0144	0.0005	0.0139	88	0.0218	0.0174	0.0043
61	0.0288	0.0288	0.0000	108	0.0137	0.0000	0.0137	105	0.0206	0.0059	0.0147
83	0.0275	0.0273	0.0002	119	0.0136	0.0107	0.0030	108	0.0189	0.0141	0.0047
64	0.0268	0.0268	0.0000	4	0.0129	0.0000	0.0129	64	0.0182	0.0005	0.0177
27	0.0214	0.0214	0.0000	203	0.0123	0.0000	0.0123	7	0.0178	0.0002	0.0176
2	0.0207	0.0207	0.0000	102	0.0115	0.0000	0.0115	61	0.0156	0.0061	0.0095
75	0.0187	0.0187	0.0000	15	0.0107	0.0000	0.0107	59	0.0141	0.0018	0.0123
7	0.0180	0.0180	0.0000	83	0.0099	0.0088	0.0011	165	0.0137	0.0126	0.0011
99	0.0174	0.0174	0.0000	75	0.0098	0.0000	0.0098	52	0.0125	0.0095	0.0030
50	0.0173	0.0173	0.0000	114	0.0084	0.0000	0.0084	57	0.0125	0.0025	0.0099
15	0.0170	0.0170	0.0000	59	0.0083	0.0049	0.0035	203	0.0119	0.0119	0.0000
105	0.0167	0.0165	0.0002	64	0.0082	0.0053	0.0029	102	0.0111	0.0111	0.0000
196	0.0166	0.0166	0.0000	95	0.0078	0.0078	0.0000	19	0.0110	0.0059	0.0050
57	0.0166	0.0165	0.0001	145	0.0076	0.0003	0.0073	130	0.0105	0.0023	0.0082
5	0.0148	0.0148	0.0000	9	0.0066	0.0000	0.0065	196	0.0091	0.0060	0.0031
165	0.0143	0.0143	0.0000	88	0.0065	0.0002	0.0063	4	0.0090	0.0051	0.0039
119	0.0133	0.0133	0.0000	94	0.0064	0.0050	0.0014	14	0.0089	0.0089	0.0000
74	0.0132	0.0132	0.0000	92	0.0062	0.0062	0.0000	9	0.0085	0.0067	0.0019
4	0.0130	0.0130	0.0000	130	0.0057	0.0020	0.0037	95	0.0082	0.0000	0.0082
108	0.0128	0.0128	0.0000	5	0.0054	0.0002	0.0052	92	0.0082	0.0000	0.0082

Table 3: *Visit 2: Proportion of the variance explained by the regions of the template (see Table 5 for the template parcellation). The twenty five regions with the highest loadings are provided. For each eigenimage: first column shows the label, second shows the ordered proportions of variance explained within this eigenimage (in decreasing order), third quantifies the positive loading (blue), fourth quantifies the negative loading(red).*

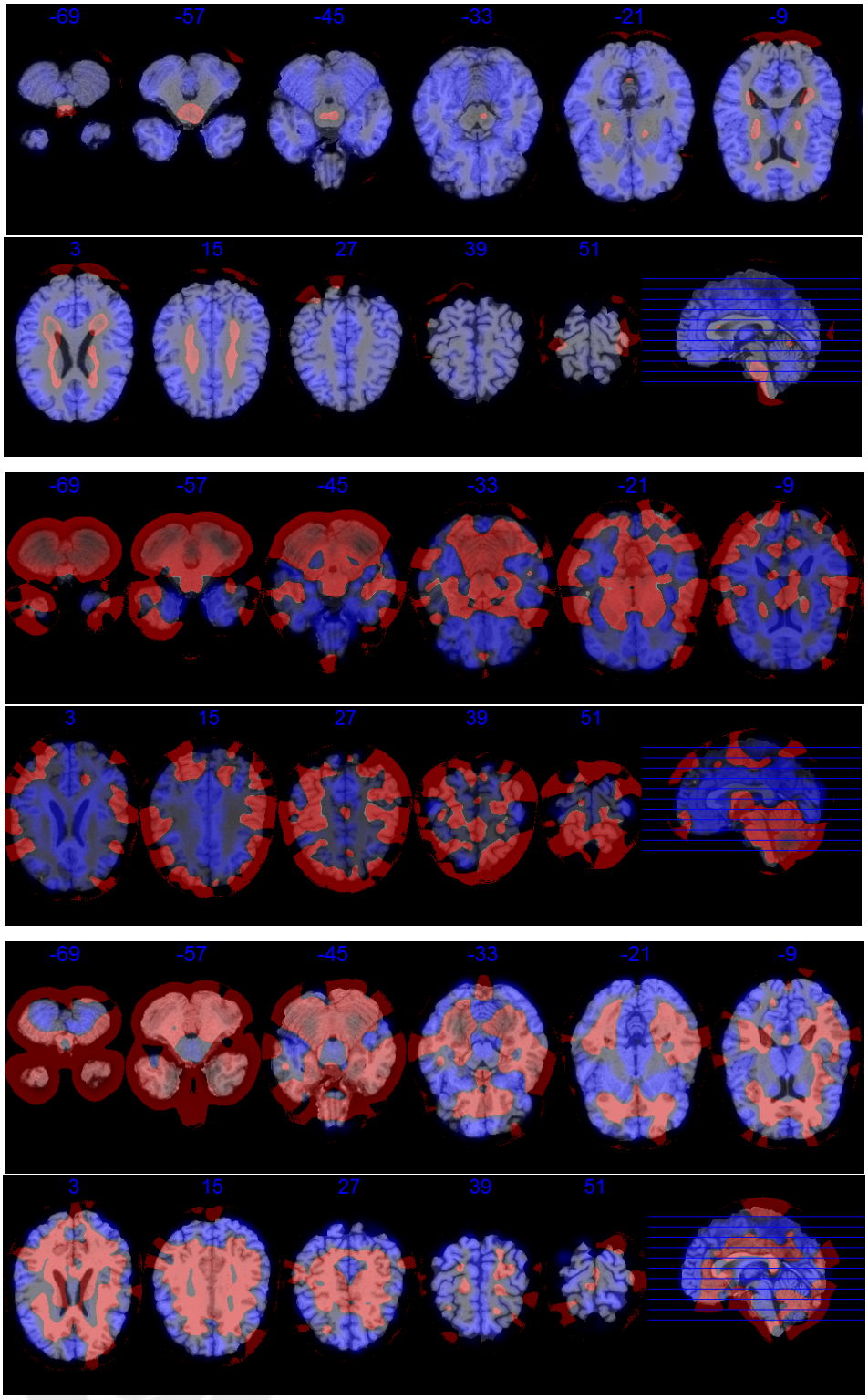

Figure 4: *The first three estimated eigenimages for visit* 1*. Each eigenimage is represented by eleven equidistant axial slices. Negative loadings are depicted in red, positive ones are in blue.*

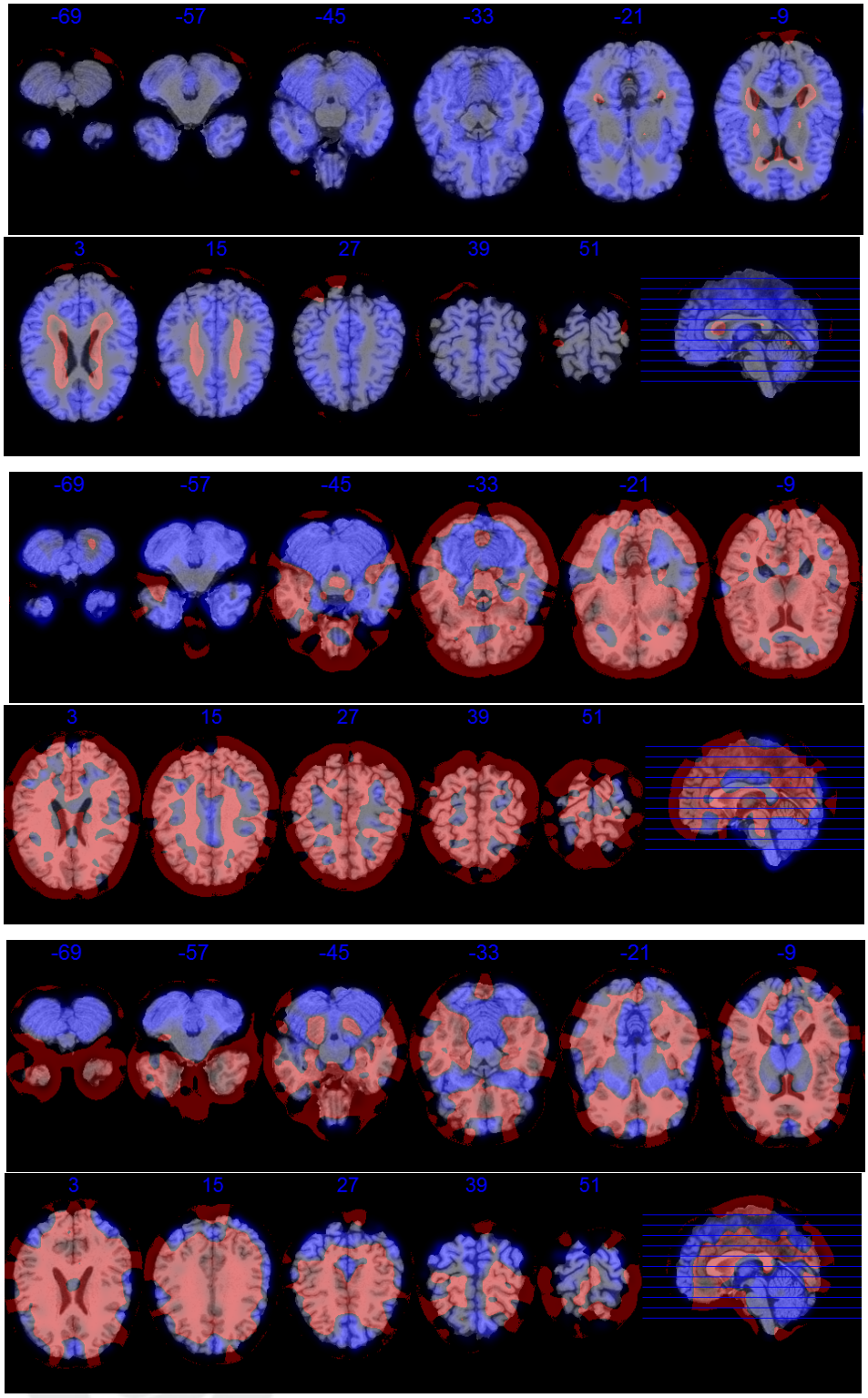

Figure 5: *The first three estimated eigenimages for visit* 2*. Each eigenimage is represented by eleven equidistant axial slices. Negative loadings are depicted in red, positive ones are in blue.*

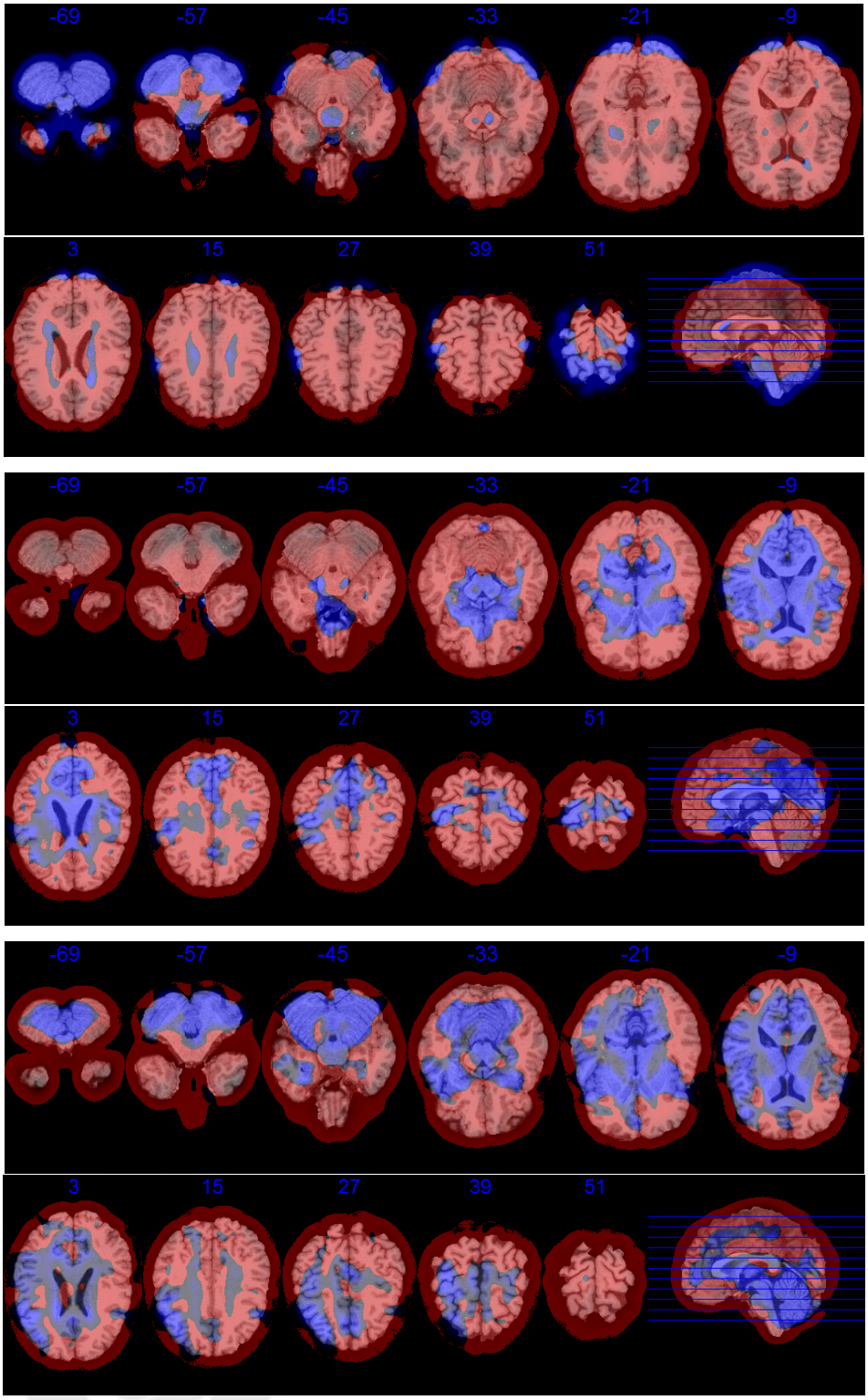

Figure 6: *The first three estimated eigenimages for the longitudinal difference. Each eigenimage is represented by eleven equidistant axial slices. Negative loadings are depicted in red, positive ones are in blue.*

Eigenimage 1				Eigenimage 2				Eigenimage 3			
255	0.0620	0.0029	0.0591	64	0.0188	0.0000	0.0188	255	0.0687	0.0087	0.0599
30	0.0439	0.0003	0.0436	255	0.0179	0.0014	0.0165	64	0.0636	0.0000	0.0636
17	0.0419	0.0003	0.0416	130	0.0142	0.0000	0.0142	94	0.0344	0.0000	0.0344
27	0.0341	0.0000	0.0341	94	0.0133	0.0000	0.0133	83	0.0288	0.0023	0.0265
59	0.0320	0.0000	0.0320	83	0.0077	0.0008	0.0069	17	0.0259	0.0192	0.0067
145	0.0291	0.0000	0.0291	196	0.0070	0.0000	0.0070	30	0.0239	0.0058	0.0180
61	0.0283	0.0000	0.0283	102	0.0066	0.0066	0.0000	21	0.0227	0.0000	0.0227
83	0.0257	0.0000	0.0257	21	0.0060	0.0000	0.0060	130	0.0218	0.0021	0.0196
130	0.0235	0.0000	0.0235	30	0.0048	0.0008	0.0040	90	0.0190	0.0002	0.0189
7	0.0231	0.0000	0.0231	140	0.0047	0.0000	0.0047	95	0.0178	0.0000	0.0178
75	0.0211	0.0000	0.0211	59	0.0046	0.0011	0.0035	61	0.0171	0.0002	0.0169
4	0.0196	0.0000	0.0196	61	0.0044	0.0001	0.0043	140	0.0151	0.0004	0.0148
108	0.0179	0.0000	0.0179	50	0.0040	0.0000	0.0040	59	0.0145	0.0083	0.0062
64	0.0174	0.0000	0.0174	37	0.0040	0.0000	0.0040	4	0.0144	0.0144	0.0000
6	0.0169	0.0000	0.0169	17	0.0035	0.0012	0.0023	5	0.0118	0.0066	0.0052
99	0.0167	0.0000	0.0167	95	0.0031	0.0000	0.0031	6	0.0115	0.0004	0.0111
105	0.0160	0.0001	0.0159	52	0.0027	0.0002	0.0025	16	0.0098	0.0098	0.0000
57	0.0153	0.0001	0.0152	251	0.0025	0.0000	0.0025	15	0.0097	0.0009	0.0088
88	0.0149	0.0001	0.0148	145	0.0023	0.0003	0.0020	102	0.0097	0.0097	0.0000
90	0.0149	0.0000	0.0149	203	0.0023	0.0023	0.0000	75	0.0090	0.0076	0.0014
2	0.0145	0.0000	0.0145	90	0.0022	0.0003	0.0019	50	0.0088	0.0001	0.0088
52	0.0143	0.0000	0.0143	99	0.0021	0.0000	0.0021	154	0.0083	0.0000	0.0083
114	0.0141	0.0004	0.0137	15	0.0020	0.0000	0.0019	99	0.0080	0.0053	0.0027
196	0.0141	0.0001	0.0139	70	0.0019	0.0000	0.0019	145	0.0080	0.0073	0.0006
15	0.0137	0.0000	0.0137	6	0.0019	0.0005	0.0013	196	0.0079	0.0011	0.0069

Table 4: *Longitudinal difference: Proportion of the variance explained by the regions of the template (see Table 5 for the template parcellation). The twenty five regions with the highest loadings are provided. For each eigenimage: first column shows the label, second shows the ordered proportions of variance explained within this eigenimage (in decreasing order), third quantifies the positive loading (blue), fourth quantifies the negative loading(red).*

1	medial front-orbital gyrus right
2	middle frontal gyrus right
3	lateral ventricle left
4	insula right
5	precentral gyrus right
6	lateral front-orbital gyrus right
7	cingulate region right
8	lateral ventricle right
9	medial frontal gyrus left
10	superior frontal gyrus right
11	globus palladus right
12	globus palladus left
14	putamen left
15	inferior frontal gyrus left
16	putamen right
17	frontal lobe WM right
19	angular gyrus right
23	subthalamic nucleus right
25	nucleus accumbens right
26	uncus right
27	cingulate region left
29	fornix left
30	frontal lobe WM left
32	precuneus right
33	subthalamic nucleus left
34	PLICICPL*
35	PLICICPR*
36	hippocampal formation right
37	inferior occipital gyrus left
38	superior occipital gyrus right
39	caudate nucleus left
41	supramarginal gyrus left
43	anterior limb of internal capsule left
45	occipital lobe WM right
50	middle frontal gyrus left
52	superior parietal lobule left
53	caudate nucleus right
54	cuneus left
56	precuneus left
57	parietal lobe WM left
59	temporal lobe WM right
60	supramarginal gyrus right
61	superior temporal gyrus left
62	uncus left
63	middle occipital gyrus right
64	middle temporal gyrus left

69	lingual gyrus left
70	superior frontal gyrus left
72	nucleus accumbens left
73	occipital lobe WM left
74	postcentral gyrus left
75	inferior frontal gyrus right
80	precentral gyrus left
83	temporal lobe WM left
85	medial front-orbital gyrus left
86	perirhinal cortex right
88	superior parietal lobule right
90	lateral front-orbital gyrus left
92	perirhinal cortex left
94	inferior temporal gyrus left
95	temporal pole left
96	entorhinal cortex left
97	inferior occipital gyrus right
98	superior occipital gyrus left
99	lateral occipitotemporal gyrus right
100	entorhinal cortex right
101	hippocampal formation left
102	thalamus left
105	parietal lobe WM right
108	insula left
110	postcentral gyrus right
112	lingual gyrus right
114	medial frontal gyrus right
118	amygdala left
119	medial occipitotemporal gyrus left
128	anterior limb of internal capsule right
130	middle temporal gyrus right
132	occipital pole right
133	corpus callosum
139	amygdala right
140	inferior temporal gyrus right
145	superior temporal gyrus right
154	middle occipital gyrus left
159	angular gyrus left
165	medial occipitotemporal gyrus right
175	cuneus right
196	lateral occipitotemporal gyrus left
203	thalamus right
243	background
251	occipital pole left
254	fornix right

Table 5: *Labeled regions of the brain template. Abbreviations: PLICICPL = posterior limb of internal capsule including cerebral peduncle left, PLICICPR = posterior limb of internal capsule including cerebral peduncle right.*